PEMERTAHANAN BAHASA MELALUI PERSPEKTIF DAFTAR KATA

I Made Sena Darmasetiyawan
Fakultas Ilmu Budaya, Universitas Udayana; CIRHSS
sena_darmasetyawan@unud.ac.id*

Muthiah Khansa Nurdina
Prodi Sastra Inggris, Fakultas Ilmu Budaya, Universitas Udayana

Janiez Alziran Rivana
Prodi Sastra Inggris, Fakultas Ilmu Budaya, Universitas Udayana

Nathania Putri Dinanti
Prodi Sastra Inggris, Fakultas Ilmu Budaya, Universitas Udayana

Franklyn Benyamin Otto
Prodi Sastra Inggris, Fakultas Ilmu Budaya, Universitas Udayana

Emanuela Patabang
Prodi Sastra Inggris, Fakultas Ilmu Budaya, Universitas Udayana

Abstrak

Berangkat dari penelitian terhadap bahasa Bali dalam analisa mendalam terhadap daftar kata yang diperoleh sebagai bagian dari hibah OCSEAN, penelitian ini berusaha melihat lebih jauh pemertahanan bahasa-bahasa di Indonesia melalui perspektif daftar kata, dengan kembali membandingkan persentase jumlah kata dan aspek kata-kata pinjaman berdasarkan 24 ruang lingkup semantik yang tersedia pada Basis Data Kata Pinjaman Dunia (WOLD). Terdapat 6 bahasa yang dibandingkan, yaitu bahasa Bali, Jawa, Sunda, Muna, Batak Toba, dan Toraja. Melalui perhitungan persentase kata pinjaman, duplikasi, dan replikasi yang terpenuhi pada beberapa ruang lingkup semantik tertentu, pemetaan terhadap pemertahanan bahasa di Indonesia dapat mulai dikembangkan untuk model penelitian lebih lanjut.
Kata Kunci: *Daftar Kata, Ruang Lingkup Semantik, Pemertahanan Bahasa*

Abstract

Following the recent study of Balinese language in the perspective of wordlist as part of the OCSEAN grant, this study attempt to investigate further the language maintenance of languages in Indonesia through using the same wordlist perspective of comparing 24 semantic fields provided from World Loanword Database (WOLD). Languages being compared are Balinese, Javanese, Sundanese, Muna, Toba Batak, and Toraja language. By calculating the percentage of borrowing, duplication, and replication available in specific semantic fields of the wordlist, mapping of the language maintenance in Indonesia can be initiated further into the future.
Keywords: *Wordlist, Semantic Fields, Language Maintenance*

1. Pendahuluan

Dilihat dari keragaman bahasa di Indonesia, penelitian – penelitian bahasa kerap menggunakan bahasa sebagai subjek yang sangat spesifik pada suatu daerah dan cenderung merupakan bagian dari revitalisasi bahasa dan budaya komunitas tersebut. Berangkat dari penelitian serupa sebelumnya (Darmasetiyawan, 2025) yang menggunakan konsep daftar kata dari hibah OCSEAN, sebagai upaya lebih lanjut, penelitian ini berusaha menerapkan metode serupa dari beberapa bahasa yang ada di Indonesia, seperti bahasa Jawa, Sunda, Muna, Batak Toba, dan Toraja. Secara umum, lokasi persebaran dan penggunaan bahasa-bahasa ini dapat dilihat pada gambar berikut

Gambar 1. Peta Indonesia dengan lokasi bahasa-bahasa yang menjadi sumber data
Sumber: Data diolah, 2025

Secara geografis, letak bahasa-bahasa tersebut cukup tersebar dari wilayah Indonesia bagian barat ke Indonesia bagian tengah. Walaupun dengan keterbatasan jumlah bahasa, letak perbandingan beberapa bahasa dari letak geografis dapat menjadi pertimbangan untuk tinjauan menyeluruh terhadap pemetaan bahasa-bahasa di seluruh wilayah Indonesia kedepannya. Selain itu, penelitian menggunakan beberapa bahasa dibandingkan dengan fokus terhadap hanya satu bahasa menjadi penting karena adanya identifikasi dari perbandingan fitur-fitur pada lebih dari satu bahasa (Pye, 2017).

Pemahaman secara makro terhadap linguistik dapat dilihat dari persepsi, representasi, dan rekognisi dalam pemerolehan bahasa tersebut (Darmasetiyawan dkk, 2024). Adanya perubahan terhadap suatu bahasa dipengaruhi oleh beberapa faktor yang bermuara dari kondisi penutur bahasa itu sendiri; apabila persepsi bahasa (atau pemerolehannya secara umum) dari penutur berubah, maka perubahan bahasa itu sendiri akan terjadi pula. Pemertahanan bahasa dalam teori sosiolinguistik dapat ditinjau dari segi kebijakan, penggunaan, dan preferensi (sikap) penutur bahasanya. Salah satu klaim dalam sosiolinguistik menyatakan bahwa keberadaan

dwibahasa (dalam hal ini, bahasa-bahasa daerah di Indonesia dengan bahasa nasional yaitu bahasa Indonesia), tidak serta merta menyebabkan kepunahan bahasa, selama terdapat adanya pembagian ranah penggunaan yang berbeda dari kedua bahasa tersebut (Nababan, 1993); contohnya bahasa Indonesia digunakan pada ranah formal (kerja dan pendidikan), sedangkan bahasa daerah digunakan pada ranah informal (keluarga).

Penggunaan daftar kata dalam studi bahasa berawal dari pemahaman bahwa terdapat kemungkinan dimana beberapa fitur dalam bahasa-bahasa tertentu memiliki karakteristik yang serupa dari garis keturunan, kontak langsung, maupun fitur yang diadopsi (Klamer dan Moro, 2023). Sebagaimana studi leksikostatistik dengan menggunakan daftar kata Swadesh yang sudah dilakukan di Indonesia (contohnya Sulistyono dan Fernandez, 2015; Ntelu dan Djou, 2017; Lailiyah dan Wijayanti, 2022; Rahman dan Agus, 2022), adanya daftar kata dari WOLD ini akan sangat mendukung upaya analisa hubungan komparasi kata atau studi linguistik komparasi secara umum dari sudut pandang artinya (Heggarty, 2010; Kaplan, 2017). Secara konsep, replikasi mengacu pada bentuk kata pinjaman yang secara pola gramatikal dapat dikatakan sama (Wiemer dkk, 2012) atau dapat dipahami sebagai fenomena kata yang sama tapi berbeda makna, sedangkan duplikasi mengarah pada persamaan kata secara parsial untuk makna yang berbeda. Tujuan penelitian yang berpusat pada fenomena kata pinjaman, duplikasi, dan replikasi ini berangkat dari banyaknya bahasa-bahasa daerah di Indonesia yang mengalami penurunan jumlah penutur – sehingga mengubah status bahasa menjadi terancam (*endangered*), bahkan mengarah pada kepunahan bahasa. Salah satu muara dari dampak ini adalah kondisi diglosia terhadap bahasa nasional, bahasa Indonesia; maka dari itu, fenomena kata pinjaman, duplikasi, dan replikasi terpusat pada perbandingan bahasa-bahasa daerah ini dengan bahasa Indonesia.

Penggunaan kata pinjaman, duplikasi, dan replikasi pada daftar kata ini bersumber dari penelitian linguistik komparatif dan linguistik historis dengan fokus pada keseragaman (*cognate*). Penelitian dengan dasar keseragaman sebelumnya lebih didasarkan pada metode analisa komputasi (contohnya Kanojia, 2021; Uban dkk, 2020), dimana dasar penelitian *Natural Language Processing* (NLP) pada penelitian tersebut menggunakan pendekatan representasi vektor kata dengan penyatuan makna kata dari beberapa bahasa. Lebih lanjutnya, keseragaman sebaiknya dilihat sebagai variabel yang tidak sekedar kategori biner (ya atau tidak – berhubungan atau tidak), namun lebih pada analisa gradien atau berlapis (Meelen dan Fellner, 2022). Pada bahasa-bahasa di Asia Tenggara pun, analisa linguistik historik pada keseragaman memerlukan perhatian khusus pada analisa morfologi kompleks di proses anotasinya, dengan

konversi dari keseragaman secara parsial (sebagian) atau keseluruhan kata (Wu dan List, 2023). Maka dari itu, sebagaimana konsep kata pinjaman berkaitan erat dengan dwibahasa dan pemertahanan bahasa, hasil yang diperoleh pun sebaiknya dianalisa secara bertahap dengan mempertimbangkan aspek-aspek penutur dan budayanya.

2. Metodologi

Penelitian ini mengambil inti metode yang pada penelitian daftar kata sebelumnya di bahasa Bali (Darmasetiyawan, 2025), namun kembali menggunakan daftar asli dari Basis Data Dunia Kata Pinjaman (WOLD) yang merujuk pada 24 ruang lingkup semantik yang memiliki total 1386 kata untuk setiap bahasa. 24 ruang lingkup semantik tersebut (dapat dilihat pada https://wold.clld.org/) adalah dunia fisik, kekeluargaan, hewan, bagian tubuh, makanan dan minuman, pakaian dan busana, tempat tinggal, agrikultur dan tumbuh-tumbuhan, aksi dasar dan teknologi, gerakan, benda milik, relasi ruang, angka, waktu, persepsi rasa, emosi dan nilai, kognisi, bahasa dan ucapan, hubungan sosial politik, perang dan perburuan, hukum, agama dan kepercayaan, dunia modern, serta kata fungsi tertentu. Dengan menggabungkan data dari bahasa Bali, total 6 bahasa bersama bahasa Jawa, Sunda, Muna, Batak Toba, dan Toraja membuat daftar kata keseluruhan data menjadi 8316 kata. Menyusul pemerolehan data dari masing-masing informan bahasa, dari 6 bahasa tersebut, terdapat beberapa kata yang tidak terisi pada daftar katanya, namun secara persentase, perhitungan tidak terisinya masih dalam batas wajar (>80%) dengan rincian bahasa Bali 92.71%, bahasa Jawa 95.24%, bahasa Sunda 83.84%, bahasa Muna 90.19%, bahasa Batak Toba 100%, dan bahasa Toraja 99.93%. Seluruh data yang telah digabungkan dapat dilihat melalui laman berikut https://osf.io/y6jgw/overview?view_only=a7994eab3ffd465d83065a7f227f7774. Seperti bagaimana identifikasi kata pinjaman pada sebuah ruang lingkup dapat memberikan batasan yang jelas antara kata pinjaman dengan fenomena homonim, ruang lingkup batasan pada kata pinjaman leksikal ini juga dapat memfokuskan atensi penelitian pada aspek ortografi dan identifikasi kata dasar (Haspelmath dan Tadmor, 2009; Tadmor et al, 2012; Klamer dan Moro, 2023).

Proses pemerolehan data yang dilakukan melalui informan sekilas terlihat subyektif, namun di lain sisi dapat menghasilkan data yang asli dan mendalam, selama diiringi dengan kemampuan untuk menyaring dan memilah data yang diperlukan. Selain itu, dalam proses pemerolehan rekaman pengisian daftar kata, informan cenderung akan berhadapan dengan tekanan yang terlihat maupun tidak, sehingga dapat mempengaruhi input data (Klamer dkk,

2021). Ruang lingkup data yang dipakai adalah bentuk ortografi (tidak termasuk segi fonologis atau simbol fonetiknya) dari kata-kata tersebut. Metode ini dilakukan karena pada penelitian sebelumnya (Darmasetiyawan, 2025) hasil analisa yang dilakukan tidak mencakup segi simbol fonetik (IPA), karena terdapat unsur keseragaman (*cognate*) yang membutuhkan studi komparatif yang lebih luas. Selain itu, model analisa berdasarkan kata pinjaman, duplikasi, dan replikasi umumnya lebih jelas terlihat dari bentuk ortografinya.

3. Hasil

Berdasarkan klasifikasi yang tersedia di Basis Data Kata Pinjaman Dunia (WOLD), data yang tersebar dari 24 ruang lingkup semantik menghasilkan total 1386 kata yang persebarannya dapat dilihat pada tabel berikut

Tabel 1. Persebaran kata pada setiap bahasa

No	Ruang lingkup Semantik	Jumlah Kata
1	Dunia fisik (*The physical world*)	75
2	Kekeluargaan (*Kinship*)	85
3	Hewan (*Animal*)	100
4	Bagian tubuh (*The body*)	100
5	Makanan dan minuman (*Food and drink*)	81
6	Pakaian dan busana (*Clothing and grooming*)	59
7	Tempat tinggal (*The house*)	47
8	Agrikultur dan tumbuh-tumbuhan (*Agriculture and vegetation*)	74
9	Aksi dasar dan teknologi (*Basic actions and technology*)	78
10	Gerakan (*Motion*)	82
11	Benda milik (*Possession*)	46
12	Relasi ruang (*Spatial relations*)	75
13	Angka (*Quantity*)	39
14	Waktu (*Time*)	57
15	Persepsi rasa (*Sense perception*)	49
16	Emosi dan nilai (*Emotions and values*)	48
17	Kognisi (*Cognition*)	51
18	Bahasa dan ucapan (*Speech and language*)	41
19	Hubungan sosial politik (*Social and political relations*)	36
20	Perang dan perburuan (*Warfare and hunting*)	40
21	Hukum (*Law*)	26
22	Agama dan kepercayaan (*Religion and belief*)	26
23	Dunia modern (*Modern world*)	57
24	Kata fungsi tertentu (*Miscellaneous function words*)	14

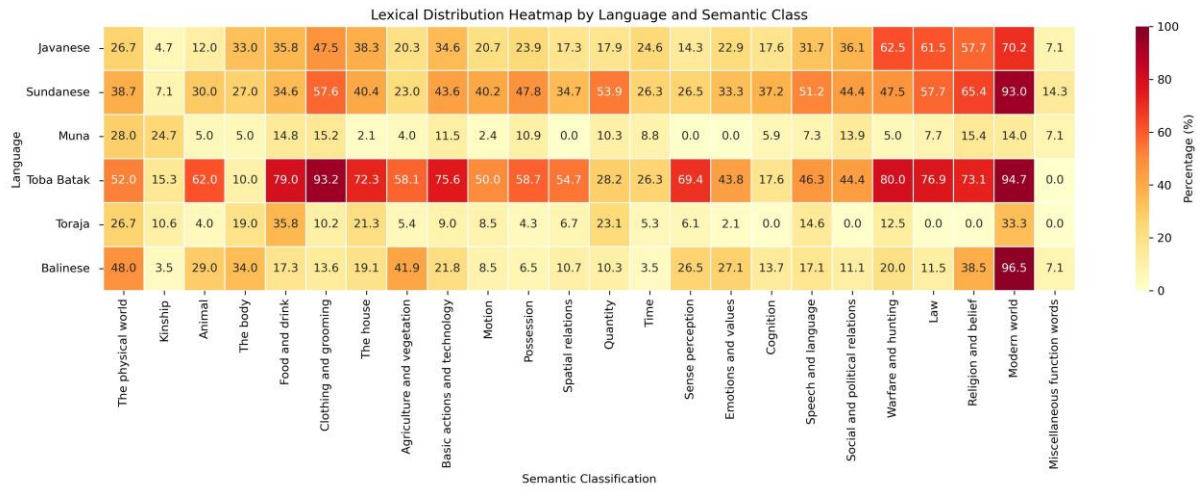
Sumber: https://wold.clld.org/meaning, 2025

Persebaran jumlah kata ini secara tidak langsung menunjukkan bagaimana penutur bahasa secara umum melihat kekayaan kosakata dalam bahasanya sesuai dengan hal-hal yang penting dalam kehidupannya, seperti hal-hal yang berkaitan dengan kekeluargaan, hewan, bagian tubuh, makanan dan minuman, serta gerakan. Di lain sisi, hal-hal seperti hukum, agama dan kepercayaan, serta kata fungsi tertentu termasuk ruang lingkup semantik yang cenderung jarang dimiliki oleh suatu komunitas penutur bahasa sebelum mengikuti pola kenegaraan yang

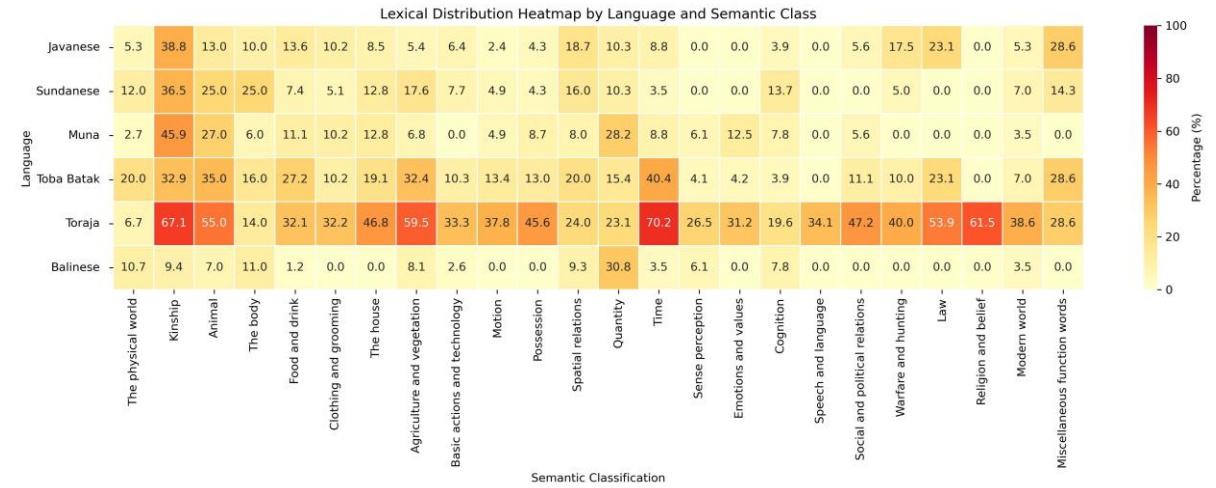
melahirkan bahasa nasional, sehingga jumlah kata yang dimiliki sangat terbatas.

Berangkat dari jumlah persebaran kata pada klasifikasi tertentu tersebut, adanya fenomena kata pinjaman, duplikasi, dan replikasi dapat menunjukkan bagaimana keberadaan kata-kata dalam ruang lingkup arti tersebut (menurut bahasanya); seperti contoh, apabila dalam sebuah bahasa ditemukan persentase kata pinjaman yang sangat kecil pada ruang lingkup semantik kognisi, dapat diasumsikan bahwa masyarakat penuturnya memiliki keterkaitan yang erat dengan kata-kata yang digunakan untuk menjelaskan kognisi, sehingga upaya yang lebih (dari segi frekwensi penggunaan dan pelestarian bahasa) telah diterapkan untuk menjaga keutuhan artinya.

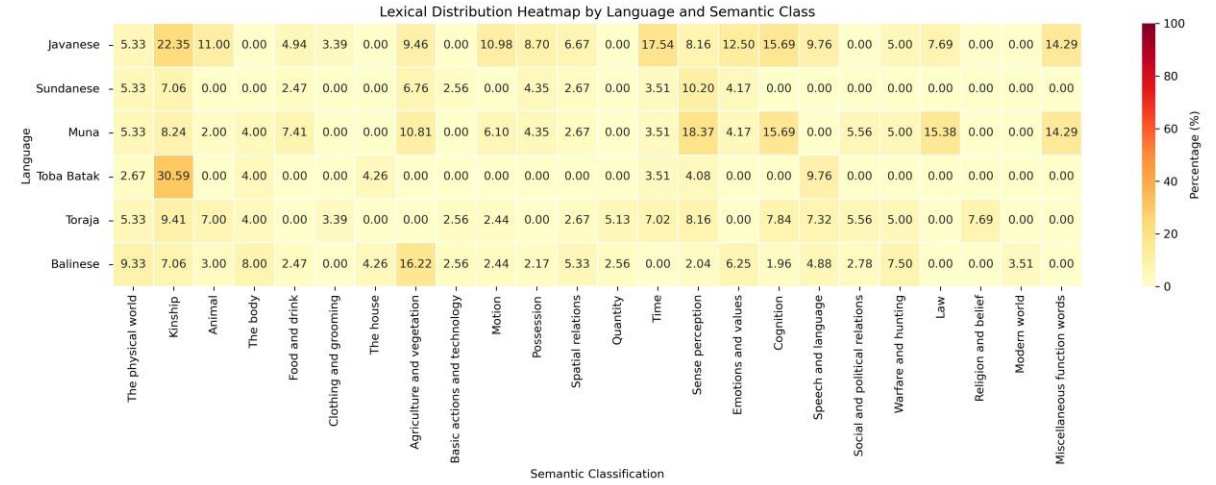
Hal serupa juga dapat didefinisikan dalam penggunaan duplikasi atau replikasi. Adanya pengulangan sebuah kata secara menyeluruh pada arti yang berbeda, dapat menunjukkan terbatasnya kedalaman arti bagi representasi makna tersebut. Sedangkan duplikasi yang terjadi pada beberapa kata dengan perluasan bentuk (contohnya dari kata menjadi frasa, namun memakai kata inti yang sama, atau perubahan imbuhan) juga dapat menunjukkan bagaimana keterkaitan beberapa kata secara semantik, sehingga tidak membutuhkan representasi tunggal dalam kosa kata penutur bahasa. Setelah penggabungan seluruh data, secara garis besar hasil perbandingannya dapat dilihat pada 3 gambar berikut



Gambar 2. Jumlah kata pinjaman (*borrowing*) dalam persentase (%)
Sumber: Data diolah, 2025



Gambar 3. Jumlah kata duplikasi (*duplicate*) dalam persentase (%)
Sumber: Data diolah, 2025



Gambar 4. Jumlah kata replikasi (*replicate*) dalam persentase (%)
Sumber: Data diolah, 2025

4. Pembahasan

Berdasarkan penelitian sebelumnya di bahasa Bali (Darmasetiyawan, 2025), tingginya kata pinjaman, duplikasi, dan replikasi pada ruang lingkup semantik angka dan waktu. Melalui perbandingan bahasa Bali dengan data secara umum, tingginya jumlah ketiga fenomena ini ditemukan pada ruang lingkup agama dan kepercayaan, pakaian dan busana, serta hukum. Namun perbedaan mencolok ditemukan pada ruang lingkup dunia fisik, hubungan sosial budaya, serta benda milik. Dengan hasil perbandingan antar bahasa di Indonesia seperti gambar 2 diatas, dapat dilihat bahwa ruang lingkup angka dan waktu tidak lagi memiliki hasil tertinggi. Dengan adanya perbandingan 6 bahasa ini, terlihat bahwa bahasa tertentu (contohnya bahasa Batak Toba) memang cenderung memiliki tingkat kata pinjaman yang lebih tinggi. Secara umum, dari ke 6 bahasa ini, ruang lingkup semantik yang memiliki tingkatan kata

pinjaman tinggi secara berurutan adalah dunia modern, agama dan kepercayaan, pakaian dan busana, perang dan perburuan, dunia fisik, makanan dan minuman, serta hukum. Sementara ruang lingkup dunia fisik cenderung kekal/tidak berubah secara makna pada setiap bahasa (kecuali memiliki korelasi erat dengan budaya), pemaknaan dunia modern dan hukum lebih berupa keberterimaan bahasa terhadap kata-kata dari bahasa Indonesia yang sebelumnya secara konseptual tidak bahkan ada pada bahasa daerah tersebut. Di lain sisi, tingginya kata pinjaman dari ruang lingkup agama dan kepercayaan, pakaian dan busana, serta makanan dan minuman, merupakan hasil dari persebaran yang cukup masif di Indonesia, sehingga banyak kata-kata yang langsung berterima oleh penutur bahasa daerah. Berangkat dari asumsi tersebut, salah satu ruang lingkup yang cukup tinggi pada perang dan perburuan hanya dapat dipahami sebagai pemaknaan pada budaya penuturnya yang cenderung tidak ada (aktivitas tersebut sangatlah jarang) atau sangat spesifik (secara prosedur dan peralatan sangat jauh berbeda).

Lebih jauhnya, dari angka-angka hasil tersebut, apabila kita lihat pemetaan lokasi geografis bahasanya pada gambar 1, terdapat kecenderungan bahwa semakin ke barat lokasi bahasa tersebut di Indonesia, semakin tinggi tingkat kata pinjamannya, dengan urutan paling tinggi ke rendah; bahasa Batak Toba, Sunda, Jawa, Bali, Toraja, Muna. Hal ini juga sejalan dengan studi linguistik historis yang menyatakan bahwa pergerakan bahasa Indonesia (atau bahasa Melayu pada waktu tersebut) memang dari barat (Sumatera) dan menyebar ke timur melalui jalur maritim (Collins, 2005). Berdasarkan perkiraan dari berbagai sumber, jumlah penutur bahasa masing-masing adalah lebih dari 1 juta untuk bahasa Batak Toba, lebih dari 30 juta untuk bahasa Sunda, berkisar 60 sampai 80 juta untuk bahasa Jawa, diatas 3 juta untuk bahasa Bali, lebih dari 500 ribu untuk bahasa Toraja, dan sekitar 300 ribu untuk bahasa Muna. Apabila dilihat dari rata-rata jumlah penutur tersebut, asumsi perkembangan penutur dwibahasa (karena seluruhnya adalah penduduk Indonesia) akan merujuk pada semakin tinggi jumlah penuturnya, maka akan semakin tinggi pula tingkat kata pinjamannya. Namun berangkat dari perbedaan yang jelas pada ruang lingkup kata-katanya (Nababan, 1993), maka dapat diasumsikan bahwa bahasa Jawa yang paling tinggi jumlah penuturnya, memiliki ketahanan bahasa (dan budaya) yang tinggi dengan berhasilnya upaya pemisahan ranah penggunaan bahasa daerah dengan bahasa Indonesia.

Apabila dibandingkan dengan hasil duplikasi pada gambar 3, terlihat korelasi yang berbanding terbalik dengan tingkat persentase kata pinjaman di gambar 2, dimana tingkat duplikasi yang tinggi, muncul pada bahasa yang cenderung memiliki tingkat kata pinjaman yang rendah (contohnya bahasa Toraja). Setidaknya secara teoritis, dari pengertian kontak

bahasa, bahasa daerah mengalami kontak dengan bahasa Indonesia dimana bahasa Indonesia sebagai bahasa nasional memiliki peran yang dominan, sementara bahasa daerah mengalami tekanan secara kontak bahasa, sehingga duplikasi pada bahasa daerah mulai digunakan untuk fungsi jamak, intensitas, atau generalisir pada perubahan morfologisnya (Thomason dan Kaufman, 1988). Dari sudut pandang ruang lingkup semantik, hal ini juga dapat terlihat pada kekeluargaan, dimana tingginya kata pinjaman pada gambar 2 berbanding terbalik dengan rendahnya duplikasi pada gambar 3. Sebaliknya, rendahnya kata pinjaman pada ruang lingkup kata fungsi tertentu memberikan hasil duplikasi yang tinggi. Rendahnya kata pinjaman pada ruang lingkup kekeluargaan dapat menunjukkan adanya pemertahanan bahasa daerah (serta pergeseran ke bahasa Indonesia) yang ditunjukkan oleh bagaimana kosa kata pada ruang lingkup kekeluargaan sangat dianggap penting (contohnya Idaryani, 2023). Maka dari itu, tingkat duplikasi yang cenderung berupa keseragaman bentuk morfologis ini dapat diartikan sebagai perluasan makna yang mengarah pada tingkat pemertahanan bahasa daerah tersebut.

Hal serupa juga terjadi dengan hasil replikasi yang ada pada gambar 4, dimana bahasa yang menunjukkan tingkat duplikasi tinggi (contohnya bahasa Toraja pula), cenderung memiliki tingkat replikasi rendah. Korelasi yang jelas terlihat antara fenomena duplikasi dan replikasi disini adalah bagaimana proses kebertahanan bahasa terjadi secara organik – secara perlahan, kata-kata yang mengalami duplikasi bisa saja menjadi replikasi saat perbedaan secara semantik yang dibawa oleh perbedaan leksikal pada kata-kata dalam duplikasi tersebut perlahan menghilang. Berangkat dari korelasi ini, tingkat replikasi dapat diartikan sebagai fenomena serupa dengan kata pinjaman (dari bahasa Indonesia), dimana keduanya merefleksikan rendahnya tingkat pemertahanan bahasa-bahasa daerah tersebut – semakin tinggi tingkat kata pinjaman dan replikasi yang terjadi, diasumsikan semakin rendah pula pemertahanan bahasanya. Beberapa penelitian sebelumnya (de Cillia dkk, 1998; Tsunoda, 2006) mulai melihat indikasi adanya penurunan (atau tidak stabilnya) bentuk morfologi kompleks dan hilangnya fitur khas bahasa asal sebagai argumen untuk menurunnya pemertahanan bahasa. Hal ini sejalan dengan hasil penelitian ini, dimana bentuk kata pinjaman dan replikasi yang secara akumulatif merupakan salah satu tanda dari perubahan bentuk morfologis dapat diartikan sebagai tingkatan menurunnya pemertahanan bahasa daerah.

5. Simpulan

Secara keseluruhan, tingkatan kata pinjaman, duplikasi, dan replikasi (berurutan dari tinggi ke rendah) pada bahasa Batak Toba, Sunda, Jawa, Bali, Toraja, dan Muna memberikan

hasil yang beragam. Berdasarkan pemetaan geografis dan hasil persentase tersebut, terdapat kecenderungan semakin di barat lokasi geografis bahasa tersebut, semakin tinggi kata pinjamannya. Tingginya tingkat pemertahanan bahasa dapat dilihat melalui penggabungan jumlah persentase dengan perkiraan jumlah penutur bahasa daerah tersebut. Persentase kata pinjaman dan replikasi dapat menunjukkan rendahnya pemertahanan bahasa, sedangkan persentase duplikasi menunjukkan tingginya pemertahanan bahasa.

6. Daftar Pustaka

de Cillia, R., Menz, F., Dressler, W. U., & Cech, P. (1998). Linguistic minorities in Austria. In C. B. Paulston & D. Peckham (Eds.), *Linguistic minorities in Central and Eastern Europe* (pp. 18–36). Multilingual Matters [27, 34].

Collins, J. T. (2005). Bahasa Melayu, bahasa daerah, dan bahasa nasional: Kasus Ambon. *Linguistik Indonesia, 23*(2), 141–156.

Darmasetiyawan, I. M. S., Aryawibawa, I. N., and Mulyawan, I. W. (2024). *Psycholinguistics*. Jawa Barat: Mega Press Nusantara.

Darmasetiyawan, I. M. S. (2025). Balinese cultural representation from the perspective of the Balinese wordlist. *LingTera, 12*(2). https://doi.org/10.21831/lt.v12i2.86302

Haspelmath, M. & Tadmor, U. (2009). *Loanwords in the World's Languages: A Comparative Handbook.* Berlin, New York: De Gruyter Mouton. https://doi.org/10.1515/9783110218442

Heggarty, P. (2010). Beyond lexicostatistics: How to get more out of 'word list' comparisons. *Diachronica, 27*(2), 301–324. https://doi.org/10.1075/dia.27.2.07heg

Idaryani. (2023). The usage of kinship terms across three generations in Indonesia. *Journal of English Language and Education, 8*(1), 396. https://doi.org/10.31004/jele.v8i1.396

Kanojia, D., Dabre, R., Dewangan, S., Bhattacharyya, P., Haffari, G., & Kulkarni, M. (2020). *Harnessing cross-lingual features to improve cognate detection for low-resource languages*. In D. Scott, N. Bel, & C. Zong (Eds.), Proceedings of the 28th International Conference on Computational Linguistics (COLING 2020) (pp. 1384–1395). International Committee on Computational Linguistics. https://doi.org/10.18653/v1/2020.coling-main.119

Kaplan, J. (2017). From lexicostatistics to lexomics: Basic vocabulary and the study of language prehistory. *Osiris, 32*(1), 202–223. https://doi.org/10.1086/694093

Klamer, M., Edwards, O., Fricke, H., Gialitaki, Z., Moro, F., Palmér, A., Saad, G., Sulistyono, Y., Visser, E., and Wu, J. (2021). Practicalities of language data collection and management in and around Indonesia. Wacana: *Journal of the Humanities of Indonesia, 22*(2), 467-521.

Klamer, M., & Moro, F. (Eds.). (2023). *Traces of Contact in the Lexicon*. Leiden, The Netherlands: Brill. https://doi.org/10.1163/9789004529458

Lailiyah, N., & Wijayanti, F. I. (2022). Kekerabatan bahasa Jawa, Bali dan Bima: Perspektif linguistik historis komparatif. *Linguistik Indonesia, 40*(2), 327–345. https://doi.org/10.26499/li.v40i2.308

Meelen, M., Hill, N. W., & Fellner, H. (2022). What are cognates? Papers in *Historical Phonology, 7*, 44–80. https://doi.org/10.2218/pihph.7.2022.7405

Nababan, P. W. J. (1993). *Sosiolinguistik: Suatu pengantar*. Jakarta: Gramedia Pustaka Utama.

Ntelu, A., & Djou, D. N. (2017). The language family relation of local languages in Gorontalo Province: A lexicostatistic study. *Journal of Arts & Humanities, 6*(11), 48–59. https://doi.org/10.18533/journal.v6i11.1285

Pye, C. (2017). *The Comparative Method of Language Acquisition Research*. Chicago and London: The University of Chicago Press.

Rahman, Z. A., & Agus, M. (2022). Kekerabatan bahasa Makassar dan bahasa Selayar: Analisis leksikostatistik dan glotokronologi. *Gema Wiralodra, 13*(1), 215–232.

Sulistyono, Y., & Fernandez, I. Y. (2015). Penerapan teknik leksikostatistik dalam studi komparatif bahasa Baranusa, Kedang, dan Lamaholot di Nusa Tenggara Timur. *Jurnal Penelitian Humaniora, 16*(1), 1–13. https://doi.org/10.23917/humaniora.v16i1.1506

Tadmor, U., Haspelmath, M., & Taylor, B. (2012). *Borrowability and the notion of basic vocabulary*. Diachronica, 27, 226-246.

Thomason, S. G., & Kaufman, T. (1988). *Language contact, creolization, and genetic linguistics*. Berkeley, CA: University of California Press.

Tsunoda, T. (2006). *Language Endangerment and Language Revitalization*. OAPEN Library.

Uban, A.-S., Ciobanu, A.-M., & Dinu, L. P. (2020). A computational approach to measuring the semantic divergence of cognates. In *Proceedings of the 20th International Conference on Computational Linguistics and Intelligent Text Processing (CICLing 2019)* (pp. 96–108). Springer.

Wiemer, B., Wälchli, B., & Hansen, B. (Eds.). (2012). Grammatical replication and borrowability in language contact (*Trends in Linguistics, Studies and Monographs, Vol. 242*). De Gruyter Mouton

Wu, M.-S., & List, J.-M. (2023). Annotating cognates in phylogenetic studies of Southeast Asian languages. *Language Dynamics and Change, 13*(2), 161–197. https://doi.org/10.1163/22105832-bja10023