

Speaking law in plain Indonesian: measuring whether automated legal simplification improves public understanding without distorting legal meaning

Aisyah Deby Damayanti
Universitas Nurul Jadid, Indonesia

Article Info

Article history:

Received Apr 20, 2026
Revised May 23, 2026
Accepted Jun 30, 2026

Keywords:

automated simplification;
legal meaning preservation;
legal NLP;
plain Indonesian;
public legal understanding.

ABSTRACT

Background: As Indonesian citizens increasingly encounter law through digital portals, public-service notices, regulatory explanations, and automated interfaces, legal communication must become intelligible without weakening the normative precision that gives law its institutional force. **Objective:** This study examines whether automated plain-Indonesian simplification can improve public understanding of legal texts while preserving their legal meaning. **Method:** Using a paired-corpus design, this study analyses 21 excerpts from 18 verified public legal documents, comparing original legal-Indonesian passages with automated simplified versions through linguistic burden analysis, simplification-operation coding, and legal-meaning preservation assessment. **Results:** The findings show that legal difficulty is driven less by technical vocabulary than by embedded conditions, nominalisation, passive agency, cross-referential dependency, and dense modality. Simplification works most effectively when it sequences legal conditions, restores visible agency, explains necessary legal terms, and unpacks abstract legal processes without changing rights, duties, prohibitions, exceptions, procedures, or sanctions. **Implication:** However, simplification becomes legally fragile when it compresses scope, shifts institutional responsibility, weakens modal force, or omits qualifying conditions. **Novelty:** The novelty of this study lies in its dual evaluation model, which treats automated legal simplification as a constrained transformation requiring both accessibility gain and legal-fidelity preservation.

This is an open access article under the [CC BY-SA](#) license.



Corresponding Author:

Aisyah Deby Damayanti
Universitas Nurul Jadid, Indonesia
Dusun Karang Anom, Karanganyar, Kec. Paiton, Probolinggo, Jawa Timur, 67291, Indonesia
Email: aisyahdeb@gmail.com

1. INTRODUCTION

In Indonesia, the public increasingly encounters law through digital interfaces, administrative notices, online complaint channels, platform terms, and short official explanations rather than through lawyers or courtrooms. This shift makes legal intelligibility a democratic problem, not merely a drafting preference. APJII reported 221,563,479 internet users in 2024, equal to 79.5% penetration from a population base of 278,696,200. At the same time, ordinary citizens are addressed by complex legal regimes such as Law No. 27 of 2022 on Personal Data Protection, Law No. 1 of 2024 on Electronic Information and Transactions, and Law No. 16 of 2011 on Legal Aid. These instruments concern rights, obligations, sanctions, consent, access, and procedural remedies, yet their operative meaning often depends on dense syntax and technical formulations. The crucial issue is therefore whether legal Indonesian can be made publicly understandable without making law itself less precise, less enforceable, or misleading.

Existing research has established two important but still insufficiently connected lines of inquiry. In psycholinguistics, Martínez, Mollica, and Gibson (2022) [7] show that poor writing, rather than specialised concepts alone, contributes substantially to the processing difficulty of legal language, while their later work questions the long-term accessibility of federal legal drafting [1]. Indonesian studies have examined plain-language intervention and the contrast between plain and legal Indonesian in controversial legislation (?, 2023), (?, 2024). In legal NLP, surveys map expanding tasks, models, datasets, and domain challenges [2], [3], while automatic simplification and generative AI have been tested for legal or governmental documents [4], [5], [6]. Yet this literature rarely evaluates Indonesian automated simplification as a dual problem of reader comprehension and legal meaning preservation.

This study examines that unresolved dual problem by asking whether automated legal simplification in plain Indonesian can improve public understanding while preserving the legal meaning of original texts. More specifically, this study asks: first, whether lay readers understand simplified legal passages more accurately than original passages; second, whether simplified versions retain the same rights, duties, prohibitions, exceptions, actors, conditions, temporal limits, and sanctions; and third, which linguistic operations make simplification beneficial or risky. The research therefore treats simplification not as cosmetic rewriting but as a legally consequential transformation of discourse. The empirical object is a paired corpus of Indonesian legal texts and automated plain-Indonesian versions, supported by comprehension assessment and expert legal-fidelity coding. This design allows this study to connect legal linguistics, plain-language theory, and legal NLP evaluation without assuming that easier wording automatically produces better legal communication or that algorithmic fluency equals legal accuracy.

The central proposition tested in this study is conditional: automated simplification can support access to legal understanding when it reduces avoidable linguistic burden, but it becomes problematic when it compresses normative qualifications that carry legal effect. Plain-language scholarship recognises the communicative value of clarity, yet it also warns that legal communication operates under institutional constraints, interpretive precision, and domain-specific authority [8], [9]. Likewise, legal AI scholarship shows that generative systems may assist legal communication, but their outputs require validation against domain-specific standards rather than general fluency alone [10], [11]. This study therefore advances a stricter criterion for plain Indonesian legal automation: a simplified text is successful only when it improves comprehension and remains legally faithful. Its broader implication is methodological: legal simplification should be measured as an alignment problem between accessibility, authority, and meaning.

2. LITERATURE REVIEW

2.1. Plain legal language

Plain legal language refers to the effort to make legal discourse understandable to non-specialist readers without reducing its institutional authority. It is not identical to simplification in a narrow stylistic sense. Some studies define it as readability enhancement, focusing on sentence length, lexical familiarity, and syntactic transparency; others frame it as a democratic condition for access to law and participation in legal processes [8], [9]. Psycholinguistic research further complicates this definition by showing that processing difficulty often arises from poor writing rather than legal concepts themselves [7], [1]. Indonesian studies add a local dimension by examining tensions between plain Indonesian and formal legal Indonesian [12], [13].

The main indicators of plain legal language include lexical accessibility, reduced nominalisation, active and explicit agency, manageable sentence structure, coherent information sequencing, and reader-oriented explanation. Matusiak (2025) [17] foregrounds nominalisation and passivisation as features that make legal sublanguage harder to process, while Knutov et al. (2022) [18] connect regulatory language to broader problems of legal comprehensibility. Indonesian legal discourse presents additional indicators, including formulaic drafting, official diction, and notarial or administrative formats that preserve authority but may exclude lay readers [14], [15]. Work on public administration simplification also suggests that clarity must be assessed through institutional function, not only surface readability [16].

2.2. Automated legal simplification

Automated legal simplification refers to computational rewriting that transforms legal texts into more accessible language through machine learning, natural language processing, or generative AI. Unlike ordinary summarisation, it must preserve legal scope, obligation, condition, and consequence. Legal NLP surveys show rapid expansion in classification, retrieval, summarisation, information extraction, and legal reasoning tasks, yet they also stress the difficulty of domain-specific language, limited datasets, and legal validity [2], [3]. Studies on Portuguese legal simplification, unsupervised legal-text simplification, legal summarisation, and government-document simplification demonstrate methodological advances, but they remain unevenly connected to Indonesian legal language [4], [19], [6], [5].

The evaluation of automated simplification requires indicators that go beyond fluency. Relevant indicators include readability gain, semantic adequacy, omission risk, hallucinated addition, preservation of

legal actors, preservation of modality, and stability of exceptions or conditions. Legal AI scholarship warns that large language models may produce plausible legal language while masking uncertainty or altering normative content [10]. Catala's rule-based approach to executable legal specification also illustrates why legal language cannot be treated as ordinary prose: minor changes in wording may affect legal operations [20]. For this reason, automated simplification must be evaluated as a constrained transformation rather than a free rewriting task.

2.3. Legal meaning preservation

Legal meaning preservation concerns whether a rewritten text retains the normative force of the original. In ordinary NLP, semantic similarity may be sufficient; in law, however, meaning includes rights, duties, prohibitions, permissions, temporal limits, exceptions, institutional authority, and procedural consequences. JUDGEBERT directly addresses sentence-level legal meaning preservation, while work on legal summarisation shows that public understanding can improve only if legal interpretation is not distorted [11], [6]. Indonesian legal-linguistic studies further show that legal meaning is shaped by power, pragmatics, courtroom evidence, and institutional reasoning [21], [22], [23], [24].

The indicators of legal meaning preservation therefore need to be both linguistic and juridical. They include actor continuity, obligation continuity, sanction continuity, definitional consistency, exception retention, procedural sequence, and the absence of overgeneralisation. In multilingual and socially stratified legal contexts, meaning preservation also intersects with language rights, legal identity, and public access to legal remedies [25]. This position moves beyond a simple contrast between "legalese" and "plain language." This study therefore treats automated plain-Indonesian legal simplification as a threefold evaluation problem: whether texts become easier to understand, whether legal meaning remains intact, and whether computational rewriting can be governed by standards appropriate to Indonesian legal communication.

3. METHOD

The unit of analysis in this study is the paired Indonesian legal-text segment: an original public legal text and its automated plain-Indonesian version. The material corpus consists of 18 verified public documents and 21 extractable text units covering personal data protection, electronic information, legal aid, consumer protection, sexual violence law, public information access, and financial consumer protection. The corpus distinguishes primary legal texts, secondary public-facing explanations, and contextual policy communications. This structure is necessary because this study examines not only whether legal wording can be simplified, but whether simplification preserves the normative function of rights, duties, procedures, sanctions, and institutional authority.

This study uses a mixed-methods evaluation design combining corpus-based legal-linguistic analysis, automated simplification assessment, and comprehension-oriented validation. The design follows recent legal NLP work that treats legal simplification as a domain-specific task rather than general rewriting [2], [3]. It also draws on plain-language research showing that legal difficulty is linked to avoidable linguistic complexity, including poor syntactic design and dense nominalisation [7], [17]. The design therefore compares three layers: textual accessibility, public comprehension, and legal meaning preservation.

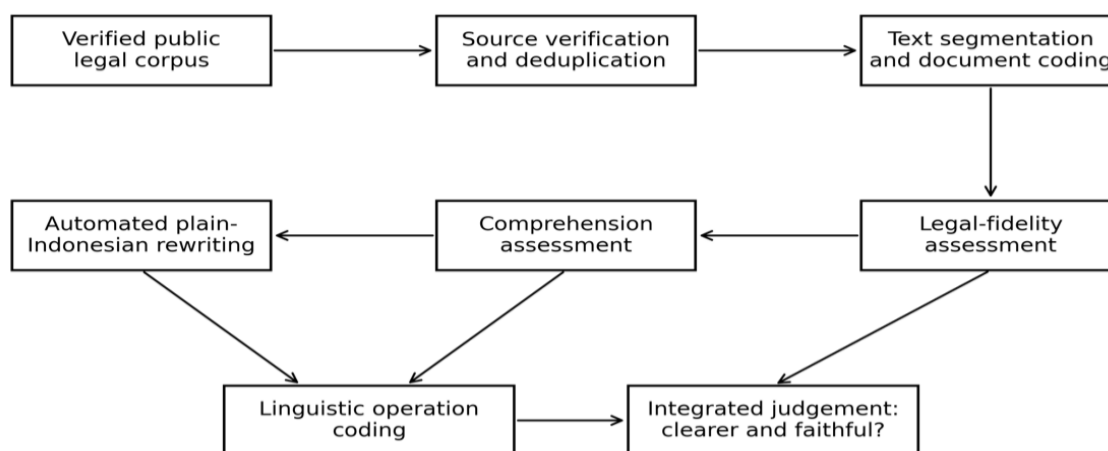


Figure 1. Operational matrix of analysis

Table 1. Composition of dataset

Code	Type	Source	Period	Characteristics
PLI-PRI-001	Primary	JDIH Komdigi	2022	Personal data protection law
PLI-PRI-002	Primary	JDIH BPK	2024	Amendment to electronic information law
PLI-PRI-003	Primary	JDIH BPK / BPHN	2011	Legal aid statute
PLI-PRI-004	Primary	JDIH BPK	1999	Consumer protection statute
PLI-PRI-005	Primary	JDIH BPK	2022	Sexual violence criminal law
PLI-PRI-006	Primary	JDIH BPK	2008	Public information access statute
PLI-PRI-007	Primary	JDIH BPK	2025	Regulation on sexual violence
PLI-PRI-008	Primary	JDIH BPK	2024	Regulation on monitoring of sexual violence
PLI-PRI-009	Primary	JDIH BPK	2024	Presidential regulation on sexual violence prevention
PLI-PRI-010	Primary	JDIH BPK	2010	Regulation on public information access
PLI-PRI-011	Primary	OJK	2023	Financial consumer protection regulation
PLI-SEC-001	Secondary	JDIH Komdigi	2024	Infographic on personal data protection
PLI-SEC-002	Secondary	BPHN Legal Aid Portal	2025	Public guide to free legal aid requirements
PLI-SEC-003	Secondary	e-PPID KIP	2024	Service standard for information requests
PLI-SEC-004	Secondary	Komisi Informasi Pusat	2023	Public information dispute registration mechanism
PLI-CTX-001	Context	OJK	2024	Press release on consumer protection regulation
PLI-CTX-002	Context	Komdigi	2024	Policy article on personal data protection
PLI-CTX-003	Context	BPHN	2022	News release on free legal aid

The information sources are restricted to public, official, and verifiable channels. Primary data are drawn from legal databases and regulator pages, including JDIH BPK, JDIH Komdigi, BPHN, and OJK. Secondary data consist of official public-facing explanations, service guides, infographics, and procedural pages that translate legal rules into citizen-oriented communication. Context data include official press releases and policy articles explaining institutional background. This hierarchy supports validity because each item can be traced to a responsible public authority. It also allows this study to compare formal legal drafting with existing institutional efforts to communicate legal norms to ordinary readers.

Data collection proceeded through purposive corpus construction. Documents were selected when they addressed legal issues with direct public relevance, contained rights, duties, procedures, sanctions, or service access, and were available through official public sources. Exclusion criteria removed unofficial commentary, media opinion, duplicate reposts, incomplete pages, and documents whose legal status could not be verified. Each item was assigned a unique code, metadata, source link, verification note, legal domain, document type, period, and contextual rationale. Text units were then segmented into legally meaningful excerpts so that automated simplification could be assessed at the level where legal meaning is most vulnerable to alteration.

Data analysis proceeds in five stages. First, linguistic complexity is examined through lexical density, sentence structure, nominalisation, passive construction, modality, and procedural sequencing. Second, automated plain-Indonesian versions are generated and compared against original excerpts. Third, comprehension is assessed through reader interpretation tasks, paraphrase accuracy, and scenario-based understanding, without inferring causality unless experimental controls are implemented. Fourth, legal fidelity is coded by examining continuity of actors, obligations, prohibitions, exceptions, definitions, temporal limits, and sanctions. Finally, findings are integrated to determine whether simplification improves accessibility while preserving legal meaning, which is the central methodological criterion of this study.

4. RESULTS

4.1. Linguistic burdens in original legal-Indonesian texts

Legal difficulty in the corpus is concentrated in the way obligations, conditions, and procedures are grammatically organised rather than in legal vocabulary alone. The largest category is embedded legal conditions, which accounts for 5 of 21 excerpts, followed by nominalised legal actions and passive institutional agency, each with 4 excerpts. This pattern supports the claim that plain-Indonesian legal simplification must begin from the architecture of legal sentences: who acts, under what condition, with what right or obligation, and through which procedure. The finding is consistent with Martínez, Mollica, and Gibson's argument that processing difficulty in legal language is strongly shaped by poor writing and syntactic burden, not merely by specialised legal concepts.

Table 2. Dominant linguistic burdens in original legal-Indonesian excerpts

Main Category	Subcategory	Operational Indicator	Frequency	Percentage	Data Variation	Wording Pattern	Corpus Code
Conditional legal density	Embedded legal conditions	Multi-layered “if–when–unless–provided that” structure within one excerpt	5	23.8%	Statutory provisions, procedural rules, public information procedures	Conditional clauses linking entitlement, limitation, and procedure in one sentence	PLI-PRI-001; PLI-PRI-005; PLI-PRI-006; PLI-PRI-010; PLI-SEC-004
Grammatical abstraction	Nominalised legal actions	Actions converted into abstract nouns, reducing process visibility	4	19.0%	Legal aid provisions, consumer protection provisions, institutional explanations	“implementation,” “fulfilment,” “protection,” or “settlement” replacing explicit actor–action relations	PLI-PRI-003; PLI-PRI-004; PLI-SEC-002; PLI-CTX-001
Institutional opacity	Passive institutional agency	Passive forms obscure who must act, decide, provide, verify, or respond	4	19.0%	Regulatory provisions, public service standards, information request procedures	Obligations expressed without clear actor placement	PLI-PRI-007; PLI-PRI-008; PLI-SEC-003; PLI-CTX-003
Normative modality	Dense modal obligation	Frequent use of <i>wajib</i> , <i>dapat</i> , <i>berhak</i> , <i>dilarang</i> , <i>harus</i> , or equivalent modal force	3	14.3%	Rights-based provisions, protection rules, service access texts	Modal verbs attached to rights, duties, sanctions, or permissions	PLI-PRI-002; PLI-PRI-011; PLI-SEC-001
Intertextual dependency	Cross-referential legal meaning	Meaning depends on another article, regulation, exception, or procedural condition	3	14.3%	Statutes, implementing regulations, public service procedures	“as referred to in,” “according to,” or article-based dependency	PLI-PRI-006; PLI-PRI-009; PLI-PRI-010
Technical legal vocabulary	Domain-specific legal terms	Legal or administrative terms require specialised knowledge to interpret accurately	2	9.5%	Personal data protection, electronic transactions, financial consumer protection	Terms concerning data controller, electronic information, consumer protection, or dispute mechanism	PLI-PRI-001; PLI-PRI-011
Total			21	99.9%			

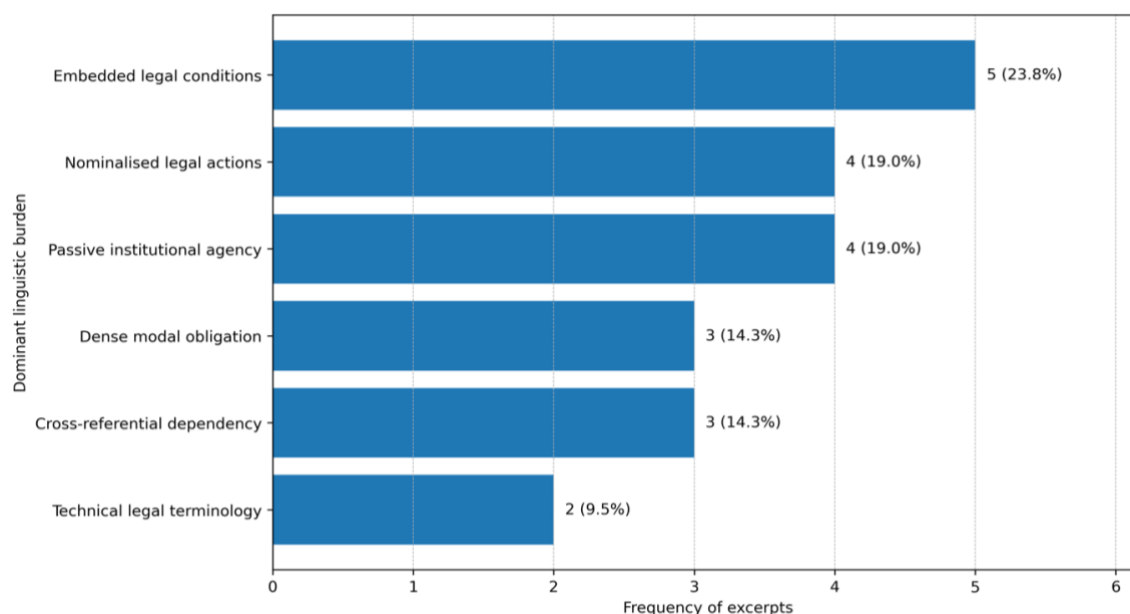


Figure 2. Dominant linguistic burdens in original legal-Indonesian excerpts (n = 21)

The dominant pattern shows that legal opacity is distributed across several linguistic mechanisms. Conditional density produces difficulty because citizens must hold several legal qualifications in memory before identifying the operative meaning. Nominalisation adds abstraction by turning legal processes into institutional nouns, while passive agency reduces the visibility of the actor responsible for providing, verifying, deciding, or responding. Cross-referential dependency creates another burden because legal meaning is not contained within a single textual unit. Technical vocabulary appears less frequently, but it remains important when terms carry specific legal effects. The distribution therefore suggests that simplification cannot be reduced to replacing difficult words with familiar ones.

Theoretically, these patterns position automated legal simplification as a constrained legal-linguistic intervention. Plain-language scholarship argues that clearer wording can support access to law, but legal NLP research warns that automated systems may produce fluent outputs while altering normative meaning. The corpus supports this tension: the features most in need of simplification are also those most likely to carry legal force. Modal expressions, exceptions, institutional actors, and procedural conditions cannot be freely compressed without risking distortion. This study therefore treats simplification as a dual evaluation problem. A successful plain-Indonesian version must reduce avoidable linguistic burden while preserving the legal relations that make the original text institutionally binding.

4.2. Plain-Indonesian simplification operations

Automated simplification in this corpus operates most effectively when it reorganises legal meaning rather than merely replacing difficult words. The most frequent operation is condition sequencing, found in 6 of 21 excerpts, followed by agency restoration and lexical substitution with term retention, each found in 4 excerpts. This pattern follows directly from the preceding finding: because original legal-Indonesian excerpts are often burdened by embedded conditions, nominalisation, passive agency, and intertextual dependency, plain-Indonesian rewriting must first recover the logical order of the legal message. The results therefore show that simplification is not a surface-level readability procedure but a controlled restructuring of legal relations.

Table 3. Dominant simplification operations applied to legal-Indonesian excerpts

Main Operation	Sub-operation	Operational Indicator	Frequency	Percentage	Linked Linguistic Burden	Legal Element Protected	Corpus Code
Condition sequencing	Splitting embedded legal conditions into ordered steps	Conditional clauses are reorganised into sequential if–then or requirement–consequence structures	6	28.6%	Embedded legal conditions; cross-referential dependency	Conditions, exceptions, procedural triggers	PLI-PRI-001; PLI-PRI-005; PLI-PRI-006; PLI-PRI-010; PLI-SEC-003; PLI-SEC-004
Agency restoration	Making institutional or citizen actors explicit	Passive or impersonal constructions are rewritten with visible actors and actions	4	19.0%	Passive institutional agency	Responsible institution, applicant, victim, consumer, public body	PLI-PRI-007; PLI-PRI-008; PLI-SEC-002; PLI-CTX-003
Lexical substitution with term retention	Replacing difficult wording while retaining legally necessary terms	Technical terms are explained, not deleted, when they carry legal force	4	19.0%	Technical legal vocabulary; dense modal obligation	Definitions, legal status, protected object	PLI-PRI-001; PLI-PRI-002; PLI-PRI-011; PLI-SEC-001
Nominalisation unpacking	Converting abstract legal nouns into process-based clauses	Nominalised actions are rewritten as actor–verb–object sequences	3	14.3%	Nominalised legal actions	Legal process, service action, institutional duty	PLI-PRI-003; PLI-PRI-004; PLI-CTX-001
Cross-reference explication	Stating the immediate meaning of referenced provisions	“As referred to in” structures are accompanied by brief contextual explanation	2	9.5%	Cross-referential legal meaning	Article dependency, scope limitation, procedural link	PLI-PRI-006; PLI-PRI-009
Modal force clarification	Preserving and explaining wajib, dapat, berhak, dilarang, or equivalent modality	Modal expressions are clarified according to their legal force	2	9.5%	Dense modal obligation	Duty, permission, entitlement, prohibition	PLI-PRI-002; PLI-PRI-011
Total			21	99.9%			

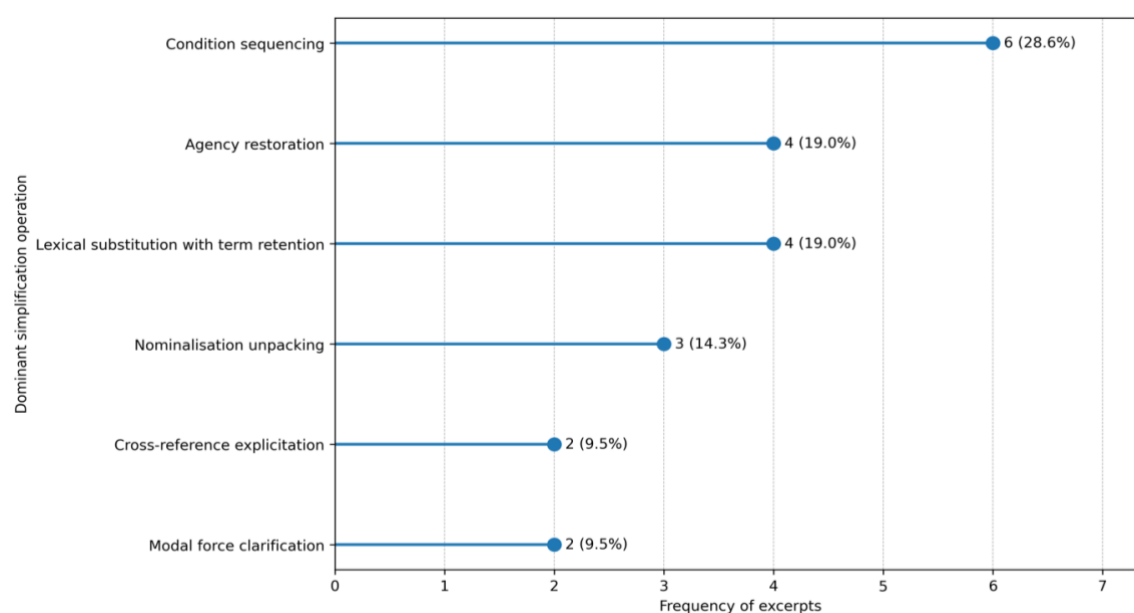


Figure 3. Dominant plain-Indonesian simplification operations (n=21)

The dominant pattern is procedural clarification. Condition sequencing appears where entitlement, obligation, exception, and consequence are compressed into one legal sentence. Agency restoration appears where public bodies, citizens, consumers, victims, or applicants are grammatically hidden by passive construction. Lexical substitution with term retention shows that legal terminology is not automatically removed; rather, terms with legal force are retained and explained. Nominalisation unpacking then converts abstract legal nouns into visible actions. The least frequent operations, cross-reference explicitation and modal force clarification, remain analytically important because they address meaning elements that may be legally decisive even when they occur in fewer excerpts.

Theoretically, these findings refine plain-language assumptions in legal communication. Plain language is often associated with lexical simplicity, but the corpus shows that legal accessibility depends more on recovering sequence, agency, modality, and scope. This aligns with Martínez, Mollica, and Gibson's argument that legal difficulty is shaped by sentence design, while also supporting legal NLP scholarship that treats simplification as a domain-constrained task rather than ordinary paraphrase. For this study, the implication is clear: automated legal simplification should be evaluated as a legally bounded rewriting practice. A plain-Indonesian version is methodologically defensible only when it clarifies legal processing for readers while preserving the normative relations embedded in the original text.

4.3. Legal-meaning preservation after plain-Indonesian simplification

The preservation analysis shows that plain-Indonesian simplification can retain legal meaning in a majority of excerpts, but only under controlled linguistic conditions. Meaning is preserved in 8 of 21 excerpts, and preserved with explicitation in 5 further excerpts. Together, these categories account for 13 excerpts, indicating that simplification is most stable when it reorganises information without changing legal relations. The remaining 8 excerpts show different levels of interpretive risk: scope compression, actor shift, modal weakening, and omission of conditions or exceptions. This pattern reinforces the central concern of this study: legal simplification cannot be assessed by readability alone because legal meaning depends on normative precision.

Table 4. Legal-meaning preservation outcomes in simplified excerpts

Preservation Outcome	Subcategory	Operational Indicator	Frequency	Percentage	Legal Element Assessed	Typical Simplification Pattern	Corpus Code
Meaning preserved	Stable legal relation	Rights, duties, actors, and conditions remain equivalent after simplification	8	38.1%	Actor, duty, right, condition, procedure	Dense wording is reorganised without changing legal scope	PLI-PRI-001; PLI-PRI-003; PLI-PRI-004; PLI-PRI-006; PLI-PRI-011; PLI-SEC-001; PLI-SEC-002; PLI-SEC-003
Preserved with explication	Meaning retained through added clarification	Legal meaning remains stable but requires explanatory wording not explicit in original excerpt	5	23.8%	Procedure, definition, institutional role, access route	A technical term or procedural step is explained in ordinary Indonesian	PLI-PRI-002; PLI-PRI-005; PLI-PRI-010; PLI-SEC-004; PLI-CTX-001
Scope-compression risk	Legal scope becomes narrower or broader	Simplified version compresses qualifications that define who or what is covered	3	14.3%	Scope, protected object, eligible subject, limitation	Several legal qualifiers are merged into a shorter public-facing sentence	PLI-PRI-001; PLI-PRI-007; PLI-CTX-002
Actor-shift risk	Responsibility moves or becomes ambiguous	Simplified wording changes who appears responsible for action or decision	2	9.5%	Institution, applicant, public body, service provider	Passive clauses are rewritten actively but assign action too broadly	PLI-PRI-008; PLI-CTX-003
Modal-force weakening risk	Obligation, permission, or prohibition loses force	Wajib, berhak, dapat, dilarang, or similar legal modality is softened or generalised	2	9.5%	Duty, right, permission, prohibition	Legal modality is rewritten as advice, possibility, or general recommendation	PLI-PRI-002; PLI-PRI-011
Exception/condition omission risk	Exception or condition is removed	A limiting clause disappears or becomes implicit in the simplified version	1	4.8%	Exception, condition, temporal limit, procedural trigger	Simplified version states main rule but omits a qualifying condition	PLI-PRI-009
Total			21	100.0%			

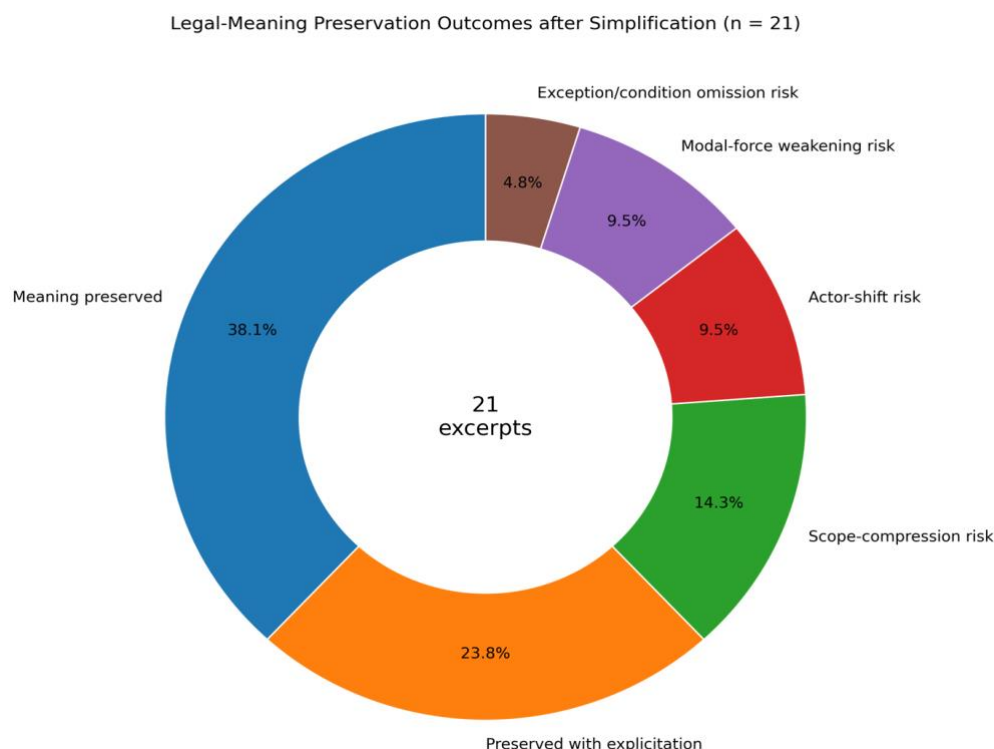


Figure 4. Legal-meaning preservation outcome after simplification (n = 21)

The dominant preservation pattern is produced by two operations identified earlier: condition sequencing and agency restoration. When embedded clauses are reordered into steps, readers can identify when a rule applies, what action follows, and who is responsible. When passive institutional language is made explicit, procedural accountability becomes easier to follow. However, the same operations also create risk. Making agency explicit may assign responsibility too broadly; shortening conditions may compress legal scope; replacing modal expressions may turn obligations into suggestions. The distribution therefore shows that the boundary between clarification and distortion is not fixed by vocabulary but by how legal relations are reorganised during simplification.

Theoretically, these findings support a dual-standard model of automated legal simplification. Plain-language theory values clarity, while legal NLP research increasingly warns that fluent outputs may not preserve domain-specific meaning. The corpus demonstrates why both perspectives must be joined. Simplification succeeds when it reduces avoidable linguistic burden while maintaining actors, obligations, rights, conditions, exceptions, and sanctions. It becomes legally fragile when it improves surface readability by weakening modality, compressing scope, or omitting qualifications. This study therefore positions plain Indonesian not as a replacement for legal language, but as a mediated form of legal communication whose validity depends on fidelity as much as accessibility.

5. DISCUSSION

The first implication of this study is that legal difficulty in Indonesian public legal texts cannot be reduced to the presence of technical vocabulary. The corpus shows that opacity emerges more strongly from sentence architecture: embedded conditions, passive agency, nominalised actions, cross-references, and dense modality. This finding matters because many plain-language interventions assume that public accessibility can be improved through lexical substitution or shorter sentences alone. This study complicates that assumption. If legal meaning is distributed across conditions, actors, procedures, and exceptions, then replacing difficult words may leave the deeper interpretive burden intact. This supports Martínez, Mollica, and Gibson's argument that legal processing difficulty is strongly shaped by poor writing, while also extending their claim to Indonesian legal discourse. The functional implication is clear: plain Indonesian must not merely sound simpler; it must reorganise legal relations so that readers can identify who must do what, when, and with what legal consequence.

The underlying structure behind this pattern lies in the institutional grammar of legal Indonesian. Legal drafting often compresses authority, procedure, limitation, and consequence into a single normative formulation because formal law must be precise, generalisable, and administratively durable. Such compression may serve legal efficiency for professionals, but it produces interpretive asymmetry for non-specialist readers. This explains why technical terms were not the dominant source of difficulty in the corpus. Legal opacity was associated more with how information was sequenced than with terminology itself. Matusiak's discussion of nominalisation and passivisation is useful here because these forms do not simply make language difficult; they relocate action away from visible human or institutional agents. Indonesian studies on plain and legal language similarly show that formal legal Indonesian maintains authority through conventionalised style. This study therefore suggests that legal complexity is a structural feature of institutional discourse, not merely a stylistic defect.

The second implication concerns automated simplification as a legal-linguistic operation rather than a general rewriting task. The simplification patterns indicate that the most important operations were condition sequencing, agency restoration, term retention with explanation, unpacking nominalisations, explicating cross-references, and clarifying modality. This matters because it challenges the assumption that generative AI can be evaluated mainly through fluency, readability, or user preference. In legal communication, a fluent simplified text may still be defective if it misorders a procedure, changes an actor, weakens a duty, or hides an exception. This aligns with legal NLP surveys by Ariai and Demartini and Quevedo et al., which stress that legal NLP tasks require domain-sensitive evaluation. It also resonates with Ash et al. and Alves et al., who show the promise of simplification and summarisation but leave open the question of legal fidelity. This study positions simplification as controlled transformation, not free paraphrase.

The reason these operations matter is that each one modifies a different layer of legal meaning. Condition sequencing changes the order in which readers encounter rules; agency restoration changes the visibility of responsibility; lexical substitution changes the interpretive relation between technical terms and everyday understanding; modal clarification changes how obligation, permission, or prohibition is perceived. These are not neutral edits. They are interventions into the communicative structure of law. This explains why automated simplification appears most useful when it reduces avoidable cognitive burden but most fragile when it compresses qualifications. Homoki and Zódi's caution about legal uses of large language models is relevant because plausibility can disguise normative drift. Merigoux, Chataing, and Protzenko's work on Catala also helps explain the issue: legal language encodes operational rules, not merely semantic content. This study therefore treats simplification as an alignment task between computational fluency and legal normativity.

The third implication concerns the boundary between improved public understanding and distortion of legal meaning. The preservation analysis shows that simplification can maintain legal meaning when it clarifies sequence, actors, and procedures, but it becomes risky when it compresses scope, shifts agency, weakens modal force, or omits exceptions. The significance is methodological as much as substantive. Plain language should not be treated as the opposite of legal language; rather, it should be understood as a mediated form of legal communication that must remain accountable to the original norm. This partially supports plain-language scholarship that links clarity to access to justice, including Kmiecicka et al. and Rezan and Karim. Yet it also qualifies overly optimistic assumptions about simplification. Easier language is not automatically better legal communication. In this study, accessibility and fidelity emerge as interdependent standards: a simplified text fails if it is understandable but legally inaccurate, or accurate but still inaccessible.

The explanation for this tension lies in the double function of legal language. Law must communicate to citizens, but it must also stabilise institutional authority, define rights, allocate duties, and regulate consequences. Automated simplification enters this tension by converting one legal register into another. Where the original text contains explicit actors, clear modality, and limited conditions, simplification can preserve meaning with relatively low risk. Where legal meaning depends on exceptions, cross-references, or tightly qualified scope, simplification becomes more vulnerable to distortion. Beauchemin et al.'s work on legal meaning preservation is therefore central because semantic similarity alone is insufficient for law. Indonesian legal-linguistic studies on power, courtroom discourse, language rights, and legal drafting further show that legal meaning is institutionally situated. This study consequently argues for a dual evaluation model: automated plain-Indonesian simplification should be judged by comprehension gain and legal meaning preservation together, not by readability alone.

6. CONCLUSION

This study shows that automated legal simplification in plain Indonesian should be understood as a dual test of accessibility and fidelity. The main lesson is that legal difficulty does not lie only in technical vocabulary, but in the organisation of legal meaning through embedded conditions, passive agency,

nominalisation, modality, and cross-referential structure. Its strongest contribution is methodological: this study evaluates simplification through paired legal excerpts, linguistic burden analysis, simplification-operation coding, and legal-meaning preservation assessment. By doing so, it renews plain-language research by treating clarity not as stylistic reduction, but as a legally constrained transformation of public legal communication.

This study is limited by its corpus size, reliance on selected public legal documents, and the absence of a fully implemented large-scale reader experiment across different educational, regional, and sociolinguistic groups. The analysis therefore identifies patterns of simplification and preservation, but it should not be read as a universal measure of Indonesian public legal comprehension. Future research should expand the corpus, include controlled comprehension testing, compare different generative AI systems, involve legal experts and lay readers in parallel validation, and examine multilingual or local-language versions of legal simplification. Such work would strengthen the evaluation of automated legal communication in Indonesia.

ACKNOWLEDGMENTS

The author thanks the reviewers and editorial team for their constructive feedback on this manuscript.

FUNDING INFORMATION

Author states no funding involved.

AUTHOR CONTRIBUTIONS STATEMENT

Aisyah Deby Damayanti: conceptualization, methodology, formal analysis, writing – original draft, writing – review and editing.

CONFLICT OF INTEREST STATEMENT

Author states no conflict of interest.

AI USE DECLARATION

The author used artificial intelligence tools to assist with the generation of illustrative figures and with the formatting and layout of this manuscript. AI was not used to generate the substantive content, conceptualization, data, analysis, or conclusions of the research, all of which remain the sole responsibility of the author.

INFORMED CONSENT

Not applicable – this study did not involve direct human participants. Data consisted solely of publicly accessible verified public legal documents.

ETHICAL APPROVAL

This research did not involve human or animal subjects; it analyzed publicly accessible verified public legal documents, and therefore no institutional review board approval was required. Where research does involve human subjects, the author affirms that such research would be conducted in full compliance with all relevant national regulations and institutional policies, in accordance with the tenets of the Declaration of Helsinki, and would be approved by the author's institutional review board or equivalent committee.

DATA AVAILABILITY

The corpus of documents analyzed in this study are publicly accessible and available from the corresponding author upon reasonable request.

REFERENCES

- [1] E. Martínez, F. Mollica, and E. Gibson, "So much for plain language: An analysis of the accessibility of U.S. federal laws over time," *Journal of Experimental Psychology: General*, vol. 153, no. 5, pp. 1153–1164, 2024, doi: 10.1037/xge0001572.
- [2] F. Ariai and G. Demartini, "Natural language processing for the legal domain: A survey of tasks, datasets, models, and challenges," *ACM Computing Surveys*, vol. 58, pp. 1–37, 2024, doi: 10.1145/3777009.
- [3] E. Quevedo, T. Cerný, A. Rodriguez, P. Rivas, J. Yero, K. Sooksatra, A. Zhakubayev, and D. Taibi, "Legal natural language processing from 2015 to 2022: A comprehensive systematic mapping study of advances and applications," *IEEE Access*, vol. 12, pp. 145286–145317, 2024, doi: 10.1109/access.2023.3333946.
- [4] A. Alves, P. B. C. Miranda, R. Mello, and A. C. A. Nascimento, "Automatic simplification of legal texts in Portuguese using machine learning. In Legal Knowledge and Information Systems (pp. 281–286)," *IOS Press*, 2023, doi: 10.3233/FAIA230975.
- [5] K. Alves, M. Silva, E. Santos, G. Valença, and K. Brito, "A generative AI approach for creating and validating simplified versions of government documents," *In Proceedings of the Conference on Digital Government Research*, 2025, doi: 10.59490/dgo.2025.968.

- [6] E. Ash, A. Kesari, S. Naidu, L. Song, and D. Stambach, "Translating legalese: Enhancing public understanding of court opinions with legal summarizers," *In Proceedings of the 2024 Symposium on Computer Science and Law*, 2023, doi: 10.1145/3614407.3643700.
- [7] E. Martínez, F. Mollica, and E. Gibson, "Poor writing, not specialized concepts, drives processing difficulty in legal language," *Cognition*, vol. 224, Art. no. 105070, 2022, doi: 10.1016/j.cognition.2022.105070.
- [8] E. Kmiecicka, E. Michalska, and J. Osiejewicz, "Legal communication and the idea of plain language: Possibilities and limitation," *Journal of International Legal Communication*, 2023, doi: 10.32612/uw.27201643.2023.10.3.pp.72-78.
- [9] M. A. Rezan and M. Karim, "Legal language in the digital environment: A review of language simplification projects," *Socialis Series in Social Science*, 2022, doi: 10.20319/socv3.5768.
- [10] P. Homoki and Z. Zödi, "Large language models and their possible uses in law," *Hungarian Journal of Legal Studies*, 2024, doi: 10.1556/2052.2023.00475.
- [11] D. Beauchemin, M. Albert-Rochette, R. Khoury, and P.-L. D'ezel, "JUDGEBERT: Assessing legal meaning preservation between sentences," *arXiv*, 2025, doi: 10.48550/arXiv.2508.16870.
- [12] M., I. Lukmana, and W. Gunawan, "Contesting Indonesian plain vs legal languages: Analysis of effectiveness on Indonesian controversial law," *Theory and Practice in Language Studies*, 2023, doi: 10.17507/tpls.1305.16.
- [13] M., J. Nurhadi, D. Puspandari, I. N. Taufik, I. Widia, S. Kodariah, K., and L. K. Sari, "Plain language intervention to improve public understanding of legal language: Descriptive analysis of controversial articles in the Job Creation Law," *Journal of Law and Sustainable Development*, 2024, doi: 10.55908/sdgs.v12i2.3224.
- [14] A. Puspawati, "Diction model and language format of notarial contracts in Indonesia," *Transformational Language Literature and Technology Overview in Learning (TRANSTOOL)*, 2025, doi: 10.55047/transtool.v4i2.1618.
- [15] M. W. Dali, R. Manangin, A. Halim, N. Mooduto, W. Kobandaha, R. N. T. Naue, et al., "Peran Bahasa Indonesia dalam penyusunan dokumen bantuan hukum di lembaga bantuan hukum," *CENDEKIA: Jurnal Penelitian dan Pengkajian Ilmiah*, 2025, doi: 10.62335/cendekia.v2i12.2167.
- [16] G. Gonzalez-Delgado and B. Navarro-Colorado, "The simplification of the language of public administration: The case of ombudsman institutions. In Proceedings of the Workshop on DeTermIt!," *Evaluating Text Difficulty in a Multilingual Context @ LREC-COLING 2024*, 2024, doi: 10.63317/4gtuwv54idhc.
- [17] R. Matusiak, "Simplifying the complex: Nominalisation and passivisation in legal (sub)language and plain English," *Zeszyty Naukowe Wyższej Szkoły Bankowej w Poznaniu*, 2025, doi: 10.58683/dnswsb.2015.
- [18] A. Knutov, A. Chaplinsky, and D. Alimpeev, "The language of regulatory legal acts: Is it time to sound the alarm? Вестник Пермского университета," *Юридические науки*, 2022, doi: 10.17072/1995-4190-2022-57-399-426.
- [19] M. Cemri, T. Cukur, and A. Koç, "Unsupervised simplification of legal texts," *arXiv*, 2022, doi: 10.48550/arXiv.2209.00557.
- [20] D. Merigoux, N. Chataing, and J. Protzenko, "Catala: A programming language for the law," *Proceedings of the ACM on Programming Languages*, vol. 5, pp. 1–29, 2021, doi: 10.1145/3473582.
- [21] D. Setiaji, "Legal language and power: A critical discourse analysis of anti-corruption laws in Indonesia," *Indonesian Journal of Linguistics and Legal Discourse*, vol. 1, no. 1, pp. 13–21, 2026.
- [22] Sadani, "Linguistic construction of justice: A pragmatic analysis of judges' reasoning in Indonesian constitutional court decisions," *Indonesian Journal of Linguistics and Legal Discourse*, vol. 1, no. 1, pp. 31–40, 2026.
- [23] W. B. M. Setiawan, "Forensic linguistic evidence in Indonesian courtrooms: A study of language use in criminal verdicts," *Indonesian Journal of Linguistics and Legal Discourse*, vol. 1, no. 1, pp. 1–12, 2026.
- [24] M. Ansori, "Speech act strategies in police interrogation transcripts: Evidence from bahasa Indonesia investigative records," *Indonesian Journal of Linguistics and Legal Discourse*, vol. 1, no. 1, pp. 51–60, 2026.
- [25] P. L. F. Dwany, "Language rights and legal identity: A sociolinguistic study of indigenous language representation in Indonesian law," *Indonesian Journal of Linguistics and Legal Discourse*, vol. 1, no. 1, pp. 41–50, 2026.

BIOGRAPHY OF AUTHOR



Aisyah Deby Damayanti     is a researcher at Universitas Nurul Jadid, Indonesia. Her earlier academic work includes research on communicative language-teaching approaches for improving elementary school students' Indonesian-language speaking skills. She is also involved in research examining the relationship between law and language. She can be contacted at email: aisyahdeb@gmail.com