



Linguistic Complexity and Readability in Corporate Privacy Policies: A Corpus-Based Comparison of Apple and Samsung

Zainab Shakoor^{*a}, Fariha Yasmeen^a, Syed Shafqat Hussain Zaidi^b

a. Department of English, University of Education, Lahore, Punjab, Pakistan

b. Department of English, Riphah International University, Lahore, Punjab, Pakistan

ARTICLE INFO

Received:

June 14, 2026

Revision Received:

July 27, 2026

Accepted:

July 31, 2026

Available Online:

July 31, 2026

Keywords:

privacy policy, corpus linguistics, readability analysis, lexical density, digital discourse, Apple, Samsung.

ABSTRACT

Corporate privacy policies are legally binding documents that govern how technology companies collect, process, and share user data. Despite their legal significance, these documents are frequently characterised by dense, inaccessible language that hinders consumer comprehension. This study employs a corpus-based approach to conduct a comparative readability and lexical analysis of the official privacy policies of two leading global technology corporations, Apple Inc. and Samsung Electronics Co., Ltd. The corpus consists of two authentic policy texts extracted from their respective official websites. Thirteen readability indices were computed using automated computational tools, including the Flesch Reading Ease score, the Flesch-Kincaid Grade Level, the Gunning Fog Index, the SMOG Index, the Coleman-Liau Index, the Automated Readability Index (ARI), the Dale-Chall Score, and the Linsear Write Formula, alongside lexical measures including lexical density, type-token ratio, and long-word frequency. Findings reveal that both policies operate at a college-graduate reading level, substantially exceeding the eighth-grade plain-language benchmark recommended for consumer documents. Samsung exhibits consistently greater linguistic complexity across all indices, with a higher proportion of polysyllabic words compared to Apple. The study concludes that both corporations employ language of considerable complexity that effectively limits lay-reader comprehension, with implications for digital privacy literacy, informed consent, and regulatory policy in the context of global data governance.

1. Introduction

The rapid expansion of digital services has made privacy policies one of the most widely encountered yet least understood forms of legal discourse in contemporary society. In the digital landscape, privacy policies constitute the primary contractual interface between technology corporations and their users. These documents serve to inform users about how their personal data is collected, processed, stored, and shared. In legal terms, user engagement with a product or service typically expressed through a checkbox or a button labelled "I Agree", implies informed consent premised on the assumption that users have read and understood the policy in question. However, a substantial and growing body of scholarly research has called this assumption into question, demonstrating that the linguistic complexity of such documents renders them effectively incomprehensible to the average consumer (Ermakova et al., 2015; McDonald & Cranor, 2008; Reidenberg et al., 2015).

The challenge of privacy policy comprehension is not merely a matter of document length, though length is certainly a contributing factor. Rather, it is the density of legal jargon, the syntactic complexity of sentence structures, and the heavy

* Correspondence to: Department of English, University of Education, Lahore, Punjab, Pakistan

E-mail address: zainabshakoor014@gmail.com (Z. Shakoor).



reliance on polysyllabic nominalisations that together create what Solove (2013) has termed the "privacy paradox" users nominally consent to conditions they are structurally incapable of comprehending. This paradox is particularly acute in the context of technology companies, whose products and services are used by consumers across a wide spectrum of educational backgrounds and reading abilities. In this context, readability analysis offers a rigorous and reproducible method for quantifying the degree of linguistic difficulty embedded in corporate privacy discourse.

Apple Inc. and Samsung Electronics Co., Ltd. represent two of the most influential technology companies in the world, together accounting for the majority of global smartphone market share. Both companies maintain extensive privacy policies that comply with international data protection legislation, including the European Union's General Data Protection Regulation (GDPR) and the California Consumer Privacy Act (CCPA). A comparative analysis of the linguistic features of these two policies is therefore not only academically justified. Still, it has direct implications for consumer rights, regulatory compliance, and the broader discourse of digital governance.

The present study employs a corpus-based methodology to address this gap in the literature. By extracting authentic policy texts directly from the companies' official websites and submitting them to automated readability analysis, the study generates a comprehensive, multi-index quantitative profile of each document's linguistic accessibility. This approach is consistent with the broader tradition of corpus linguistics, which holds that language patterns are most reliably identified and described through systematic analysis of naturally occurring texts within a defined corpus (Biber et al., 1998; McEnery & Wilson, 2001). The following sections outline the study's objectives and research questions before proceeding to a review of the relevant scholarly literature and a description of the methodology employed.

1.1. Statement of the Problem

In contemporary digital society, personal data has become one of the most valuable resources collected, processed, and exchanged by technology corporations. Privacy policies are intended to function as the primary mechanism through which users are informed about how their personal information is gathered, stored, shared, and utilised. However, growing evidence suggests that these documents are frequently written in highly technical and legally complex language that exceeds the reading ability of ordinary consumers. Consequently, many individuals consent to data-processing practices without fully understanding the implications of their decisions.

This issue represents more than a linguistic concern; it is a significant digital governance and consumer-rights challenge. When users cannot comprehend privacy disclosures, meaningful informed consent becomes difficult to achieve. Such limitations may expose individuals to privacy risks, extensive data collection practices, behavioural profiling, targeted advertising, and potential misuse of personal information. In an era increasingly shaped by artificial intelligence, big data analytics, and global digital platforms, inaccessible privacy communication may contribute to widening information asymmetries between technology corporations and the users who depend upon their services.

Although previous studies have highlighted the complexity of privacy policies in general, limited research has provided a detailed corpus-based comparison of the privacy policies of Apple and Samsung, two of the world's most influential technology companies. Given their substantial global user base and their central role in the digital ecosystem, examining the readability and lexical complexity of their privacy disclosures is both academically relevant and socially important. Therefore, the present study investigates the extent to which the privacy policies of Apple and Samsung are accessible to ordinary users and explores the implications of their linguistic complexity for transparency, informed consent, and digital privacy governance.

1.2. Research Objectives

The present study is guided by two principal objectives:

- To quantify and compare the readability of the Apple and Samsung privacy policies using thirteen automated indices, thereby establishing the grade level and difficulty band of each document.
- To analyse the lexical characteristics of each policy, encompassing lexical density, type-token ratio, and polysyllabic word frequency—and to evaluate their implications for consumer comprehension, informed consent, and regulatory plain-language compliance.

1.3. Research Questions

To fulfil these objectives, the study addresses the following research questions:

1. To what degree do the Apple and Samsung privacy policies differ in terms of quantitative readability and lexical complexity as measured by multiple automated indices?
2. What implications do the observed readability and lexical patterns have for consumer comprehension, informed consent, and regulatory compliance?

2. Literature Review

The following review situates the present study within three interconnected bodies of scholarship: readability research applied to digital legal texts, corpus-based analysis of corporate privacy policies as a distinct discourse genre, and the theoretical and methodological foundations of the readability formulae employed in this study.

2.1. Readability Research and Digital Legal Texts

Readability, defined broadly as the ease with which written text may be read and understood, has been a central concern in applied linguistics and educational psychology for over a century (Klare, 1963). Traditional readability formulae, developed originally for educational materials, have increasingly been applied to legal and contractual documents as researchers have sought to quantify the accessibility of texts that citizens are routinely required to engage with (Flesch, 1948; Gunning, 1952). In the domain of digital legal language, readability research has gained significant momentum in the post-GDPR era, as regulators and consumer rights advocates have begun to scrutinise the linguistic design of privacy-related documents (Degeling et al., 2019).

McDonald and Cranor (2008) conducted a landmark study estimating that if American internet users were to read all the privacy policies they encounter in a year, the cumulative time required would amount to approximately 76 work days per person—a figure that underscores the fundamental impracticality of the notice-and-consent model. More recently, Ermakova et al. (2015) conducted a systematic review of privacy policy research, finding consistent evidence that these documents are written at a reading level far exceeding the average educational level of their intended audiences. Milne and Culnan (2004) similarly found that consumers tend to engage with privacy policies selectively and superficially, relying on heuristic cues rather than careful reading.

The application of corpus linguistics to legal discourse has a well-established precedent. Bhatia (1993) pioneered the genre analysis of legal texts, demonstrating that legal language is characterised by distinctive rhetorical conventions including the heavy use of nominal groups, passive constructions, and qualification strategies that contribute to its opacity. Tiersma (1999) provided a comprehensive account of the linguistic features that distinguish legal English from ordinary English, many of which are directly observable in contemporary privacy policies. More recently, scholars such as Lippi et al. (2019) have applied computational linguistic methods to privacy policy corpora, identifying linguistic markers associated with potentially unfair contractual terms.

2.2. Corporate Privacy Policies as a Discourse Genre

Privacy policies occupy a distinctive position within the broader genre of corporate legal discourse. Unlike other legal documents such as contracts or statutes, privacy policies are addressed directly to lay consumers and are typically presented as accessible, user-friendly documents. However, research consistently demonstrates a significant gap between this presentational stance and the actual linguistic complexity of the texts (Reidenberg et al., 2015; Waldman, 2018). The policy documents of technology companies, in particular, tend to exhibit a hybrid register that combines legal formality with marketing-inflected language, creating what Solove (2013) describes as a form of "notice theatre"—the performance of transparency without its substance.

Corpus-based approaches to privacy policy analysis have grown considerably in recent years. Raschke et al. (2014) developed one of the first large-scale annotated corpora of privacy policies, enabling systematic comparison of linguistic features across multiple corporate actors. Liu et al. (2016) examined the evolution of privacy policies over time, finding evidence of increasing length and complexity corresponding to expanding data collection practices. The present study contributes to this emerging literature by providing a fine-grained, multi-index readability comparison of two of the world's most widely used technology brands.

2.3. Readability Formulae: A Critical Overview

The present study employs multiple readability formulae in recognition of the fact that no single index provides a complete or uncontested measure of textual complexity. The Flesch Reading Ease formula (Flesch, 1948), which produces a score between

0 and 100 (with higher scores indicating greater ease), is the most widely cited readability measure in the empirical literature. The Flesch–Kincaid Grade Level formula (Kincaid et al., 1975) translates Flesch scores into an estimated US grade level, providing an intuitively accessible metric for comparison. The Gunning Fog Index (Gunning, 1952) and the SMOG Index (McLaughlin, 1969) place particular emphasis on polysyllabic word frequency, which is especially relevant in the context of legal discourse.

The Coleman–Liau Index (Coleman & Liau, 1975) and the Automated Readability Index (Senter & Smith, 1967) base their calculations on character counts rather than syllable counts, thereby providing complementary data points that triangulate the findings of syllable-based formulae. The Dale–Chall formula (Dale & Chall, 1948) draws on a validated list of familiar words, making it particularly sensitive to the presence of technical vocabulary. Together, these eight primary indices provide a robust multi-dimensional account of textual difficulty that is well suited to the demands of the present comparative analysis.

Lexical measures, including lexical density and type–token ratio, supplement these formulae by providing information about the richness and concentration of the vocabulary employed. Lexical density the proportion of content words relative to total words—is associated with information load and register formality, with higher density indices characteristic of written academic and legal registers (Halliday, 1985). The type–token ratio (TTR) measures lexical diversity by calculating the proportion of unique words relative to the total word count; a lower TTR may indicate repetitive or formulaic language, while a higher TTR suggests lexical variety.

Although previous studies have established that privacy policies are generally difficult for consumers to understand (McDonald & Cranor, 2008; Reidenberg et al., 2015), comparatively little research has conducted detailed corpus-based readability analyses of the privacy policies of leading smartphone manufacturers using multiple readability indices simultaneously. Furthermore, few studies have compared Apple and Samsung, two corporations that collectively dominate the global smartphone market. This study addresses this gap by providing a systematic comparison of their privacy policies using both readability and lexical measures.

3. Research Methodology

This section describes the research design, corpus construction procedures, and analytical tools employed in the study. Together, these decisions reflect a commitment to empirical rigour, replicability, and methodological transparency.

3.1. Research Design

This study adopts a quantitative corpus linguistics design, consistent with the established tradition of computational stylistics and automated text analysis (Biber et al., 1998). The research design is grounded in the theoretical framework of corpus linguistics, which prioritises empirical description of naturally occurring language use over intuitive or prescriptive accounts of language (McEnery & Wilson, 2001). The comparative design enables systematic identification of similarities and differences between the two texts across multiple linguistic dimensions, directly addressing the research questions stated above.

3.2. Corpus Construction

The corpus was constructed by extracting the full text of two corporate privacy policies directly from their official online sources. Apple's privacy policy was retrieved from apple.com/legal/privacy/en-ww/ and Samsung's from samsung.com/uk/info/privacy/. Both texts were retrieved in their current, publicly available versions (Apple: updated 30 July 2025; Samsung: effective 1 July 2024). Text extraction was performed programmatically using the Python requests library in Google Colaboratory, with explicit encoding correction applied to the Samsung text to resolve Unicode character mapping artefacts arising from the server's encoding header. Navigation elements, headers, footers, and cookie consent banners were excluded from the analysis through targeted text-cleaning procedures. Although the corpus comprises only two documents, each text represents a complete and authoritative privacy policy issued by a major multinational technology corporation. The study therefore employs a small specialised corpus design, which is common in corpus-assisted discourse studies focusing on specific institutional texts. The objective is not statistical generalisation to all privacy policies but a detailed comparative examination of two influential corporate documents.

The final corpus comprised 4,412 words for Apple and 3,232 words for Samsung, yielding a total corpus size of 7,644 words. This corpus size, while relatively modest, is appropriate for the focused comparative purpose of the study, as both texts constitute complete, discrete textual artefacts rather than samples drawn from a larger population.

3.3. Analytical Tools and Procedures

Readability analysis was conducted using the textstat Python library (version 0.7.0), which implements thirteen validated readability formulae. The following metrics were computed for each text: Flesch Reading Ease, Flesch–Kincaid Grade Level, Gunning Fog Index, SMOG Index, Coleman–Liau Index, Automated Readability Index (ARI), Dale–Chall Score, Linsear Write Formula, average sentence length in words, average word length in characters, percentage of polysyllabic words more than 7 characters, lexical density, and type–token ratio. Lexical density was calculated as the ratio of unique content words (nouns, verbs, adjectives, adverbs) to total words, expressed as a percentage. The type–token ratio was calculated as the ratio of unique word forms (types) to total word tokens, expressed as a percentage. All computations were performed in a reproducible Python environment and are fully replicable using the publicly accessible policy URLs and the analytical scripts documented in this paper.

3.4. Limitations

Several limitations of the present study warrant acknowledgement. First, the corpus is limited to two texts, and the findings cannot be generalised to the privacy policies of other technology companies without further investigation. Second, automated readability formulae, while well-validated and widely used, are necessarily reductive: they capture surface-level textual features (sentence length, word length, syllable count) but do not directly measure cognitive processing difficulty or comprehension. Third, the study does not examine the semantic or pragmatic dimensions of the texts, including the use of hedging, vague language, or nominalisation as discursive strategies; these dimensions would repay systematic investigation in future research. Fourth, the policy texts were retrieved at a specific point in time and may have been updated subsequently, which limits the generalisability of the specific numerical findings to the current versions of the policies. An additional limitation concerns the use of conventional type-token ratio, which is known to be influenced by corpus size. Although the difference in corpus length between the two policies was moderate, future research should employ length-independent measures of lexical diversity.

4. Results and Discussion

The results are presented in three sub-sections: an overview of corpus characteristics, a discussion of the readability indices, and an analysis of lexical features. Table 1 below presents the complete quantitative profile of both policy texts.

4.1. Overview of Corpus Characteristics

Table 1

Readability and Lexical Metrics for Apple and Samsung Privacy Policies

Metric	Apple Privacy Policy	Samsung Privacy Policy
Total Words	4,412	3,232
Total Sentences	195	137
Unique Words (Types)	902	706
Flesch Reading Ease	30.70	29.40
Flesch–Kincaid Grade Level	14.60	15.02
Gunning Fog Index	17.47	18.42
SMOG Index	15.84	16.60
Coleman–Liau Index	13.01	13.83
ARI (Automated Readability Index)	14.73	16.22
Dale–Chall Score	11.02	11.44
Linsear Write Formula	14.80	15.60
Avg. Sentence Length (words)	22.63	23.59
Avg. Word Length (characters)	5.06	5.24

Long Words ≥7 Characters (%)	30.53%	34.71%
Lexical Density (%)	59.72%	58.94%
Type-Token Ratio (%)	19.94%	21.55%

Note. All readability scores computed using the textstat Python library (vo.7.0). Corpus extracted from official corporate websites in 2025.

The Apple privacy policy contains 4,412 words distributed across 195 sentences, while Samsung's policy consists of 3,232 words across 137 sentences. Apple's text is therefore approximately 36% longer than Samsung's, though both documents fall within the range typically observed for corporate privacy policies of multinational technology companies (Lippi et al., 2019). The greater total word count in Apple's policy does not, however, translate into proportionally greater lexical variety: Apple's type-token ratio (19.94%) is slightly lower than Samsung's (21.55%), suggesting that the additional length is achieved in part through greater lexical repetition rather than increased vocabulary breadth.

4.2. Readability Indices

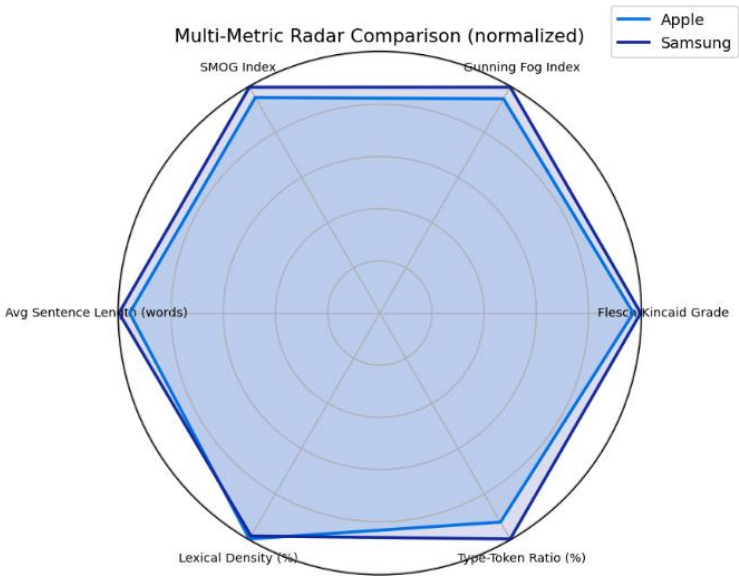
Both privacy policies score within the "Difficult" to "Very Difficult" range across all readability indices. The Flesch Reading Ease scores—Apple (30.70) and Samsung (29.40)—place both documents in the "College" to "College Graduate" difficulty band, meaning the texts are accessible only to readers who have completed or are currently engaged in university-level education. These scores are markedly lower than the score of 60–70 typically associated with standard written English and substantially below the score of 70–80 commonly recommended for consumer-facing documents (Flesch, 1948).

The Flesch–Kincaid Grade Level scores of 14.60 (Apple) and 15.02 (Samsung) correspond to approximately the second year of a four-year university programme, confirming that both texts are targeted—whether deliberately or incidentally—at a highly educated readership. The Gunning Fog Index values of 17.47 for Apple and 18.42 for Samsung are particularly instructive, as this formula is specifically sensitive to the density of polysyllabic words. A score above 17 on the Gunning Fog Index is generally considered characteristic of highly specialised professional discourse, equivalent in difficulty to academic or technical writing intended for expert audiences (Gunning, 1952). The SMOG Index reinforces this interpretation: Apple scores 15.84 and Samsung 16.60, both corresponding to a post-graduate reading level.

The Automated Readability Index presents the largest absolute difference between the two texts, with Apple scoring 14.73 and Samsung 16.22—a gap of nearly 1.5 grade levels. This discrepancy is consistent with Samsung's higher average sentence length (23.59 words) relative to Apple's (22.63 words) and its greater proportion of long words (34.71% vs. 30.53%). Across all eight primary readability indices, Samsung's text registers as the more linguistically complex of the two, though the differences are generally of modest magnitude and both texts remain firmly within the same broad difficulty band. Figure 1 presents the Flesch Reading Ease scores for both texts as donut charts, while Figure 2 provides a grade-level comparison across six primary readability indices.

Figure 1

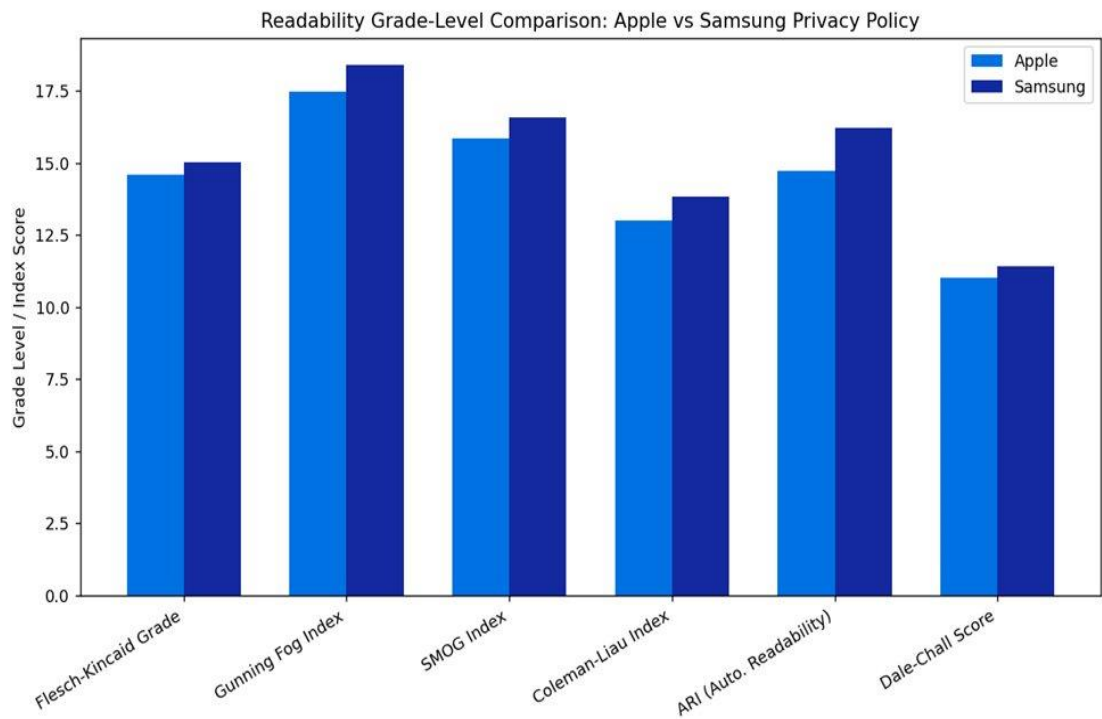
Flesch Reading Ease Scores: Apple and Samsung



Note. The shaded segment represents the Flesch Reading Ease score (out of 100) for each policy. Higher scores indicate greater readability.

Figure 2

Readability Grade-Level Comparison: Apple vs Samsung Privacy Policy



Note. Grade-level scores across six primary readability indices. Higher scores indicate greater reading difficulty.

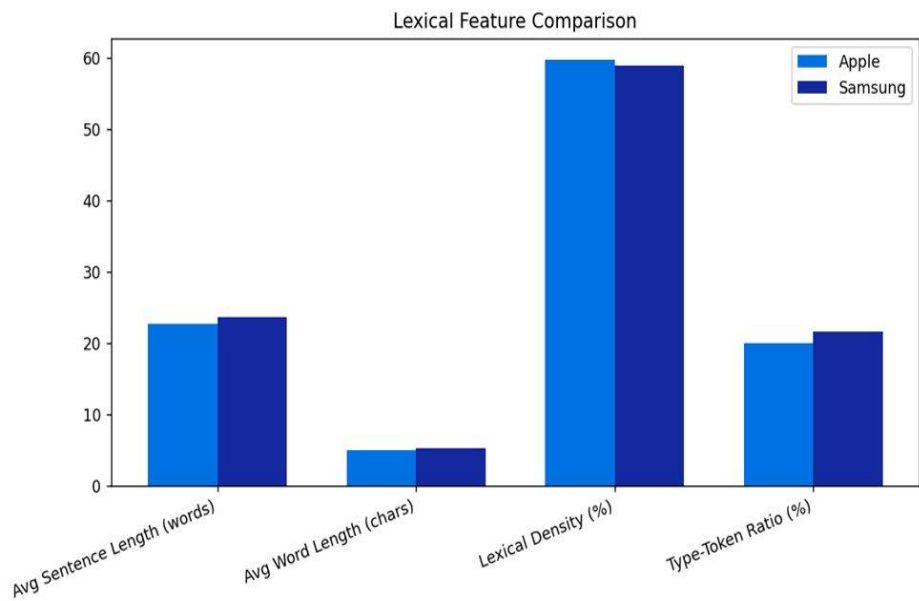
4.3. Lexical Characteristics

The lexical profiles of the two policies reveal comparable but distinct patterns of vocabulary use. Apple's text contains 902 unique word types against a total of 4,412 tokens, while Samsung's comprises 706 types from 3,232 tokens. As noted above, Samsung's slightly higher type-token ratio suggests marginally greater lexical diversity relative to text length. Both texts exhibit high lexical density: Apple at 59.72% and Samsung at 58.94%. These values fall at the upper end of the range typical for written expository prose (Halliday, 1985), confirming that both policies are information-dense texts that make heavy demands on the reader's processing capacity. Lexical density was calculated using part-of-speech tagging implemented through the Natural Language Toolkit (NLTK). Content words were operationalised as nouns, lexical verbs, adjectives, and adverbs, while function words were excluded from the calculation.

The long-word frequency measure—the percentage of words containing seven or more characters—provides a further index of lexical complexity that is directly relevant to the comprehensibility of legal language. Samsung's policy contains a substantially higher proportion of long words (34.71%) compared to Apple's (30.53%), a difference of approximately four percentage points. Given that polysyllabic words are disproportionately represented among technical legal terms, this finding suggests that Samsung's policy makes more extensive use of complex nomenclature. Illustrative examples from the texts include terms such as "personalisation," "authentication," "pseudonymisation," "cross-contextual," and "de-identified," all of which are polysyllabic and unlikely to be familiar to the average consumer. Figure 3 illustrates the key lexical metrics for both texts side by side.

Figure 3

Lexical Feature Comparison: Apple and Samsung Privacy Policies



Note. Comparison of average sentence length, average word length, lexical density, and type-token ratio for each policy.

4.4. Discussion

The results reported in the preceding section converge to present a clear and consistent picture: both Apple and Samsung privacy policies are written at a level of linguistic complexity that substantially exceeds the reading ability of the majority of their intended users. The following discussion interprets these findings in relation to the study's two research questions, considering first the implications for consumer comprehension and informed consent, and then the broader theoretical significance of the linguistic patterns identified.

4.4.1. Implications for Consumer Comprehension and Informed Consent

The findings of this study converge with a substantial body of prior research in demonstrating that corporate privacy policies are systematically written at a level of complexity that exceeds the reading ability of the majority of their intended users. Estimates suggest that the average adult reading ability in most developed countries corresponds to a reading level equivalent to approximately the eighth to tenth grade of the US educational system (Kutner et al., 2006). The Flesch–Kincaid Grade Level scores identified in the present study—14.60 for Apple and 15.02 for Samsung—are therefore between four and seven grade levels above what the average user can be expected to comfortably process. This gap is not trivial. It has direct implications for the legal and ethical validity of consent obtained under current notice-and-consent frameworks, which presuppose that the user has had a meaningful opportunity to read and understand the terms to which they are agreeing (Solove, 2013).

These implications become particularly salient when considered in the context of global regulatory frameworks. The GDPR, for example, requires that privacy information be provided "in a concise, transparent, intelligible and easily accessible form, using clear and plain language, in particular for any information addressed specifically to a child" (Regulation (EU) 2016/679, Article 12). The present analysis suggests that the policies of both Apple and Samsung fall significantly short of this standard with respect to plain language, notwithstanding their general compliance with other formal requirements such as disclosure of data categories and legal bases for processing.

The finding that Samsung's policy is marginally but consistently more complex than Apple's across virtually all indices is noteworthy, particularly given that Samsung's policy is shorter. This indicates that the additional complexity in Samsung's text is achieved through higher concentrations of polysyllabic vocabulary and longer sentence structures rather than through greater overall volume of text. Apple's policy, while more accessible in relative terms, nonetheless remains firmly within the difficult-to-very-difficult range and cannot be considered plain-language compliant by any standard regulatory benchmark.

4.4.2. Theoretical Implications: Obfuscation and Power in Corporate Privacy Discourse

The linguistic patterns identified in this study can be productively interpreted through the lens of critical discourse analysis (Fairclough, 1989). From this perspective, the complexity of privacy policy language is not simply a reflection of unavoidable legal technicality but may be understood as a discursive strategy that serves the interests of corporations over those of consumers. Highly complex language creates informational asymmetries that work to the advantage of the drafter—in this case, well-resourced legal teams operating on behalf of powerful technology companies—at the expense of the ordinary user who is structurally disadvantaged in their capacity to understand, negotiate, or contest the terms presented.

This reading is consistent with Waldman's (2018) concept of "privacy law's false promise," according to which the performative transparency of privacy policies, their provision of detailed information—is paradoxically a mechanism for obscuring rather than revealing the nature of data practices. The complexity of the language serves to deter engagement, ensuring that consent is formal rather than substantive. The high lexical density values observed in both texts (approximately 59–60%) are particularly consistent with this interpretation, as information-dense written language is characterised in the applied linguistics literature as deliberately foregrounding propositional content at the expense of interactional accessibility (Halliday, 1985).

5. Conclusion

This corpus-based study has provided a systematic, empirically grounded account of the readability and lexical characteristics of the privacy policies of two major global technology corporations, Apple Inc. and Samsung Electronics Co., Ltd. Drawing on thirteen quantitative indices computed from authentic policy texts extracted from official corporate websites, the study has demonstrated that both documents are written at a college-graduate level of difficulty—substantially exceeding plain-language benchmarks and the average reading ability of their intended users. Samsung's policy is marginally but consistently more complex than Apple's across all measures, characterised by a higher proportion of polysyllabic words, longer average sentences, and a higher grade-level score on all primary readability formulae.

In addressing the study's two research questions, the findings show that, while the quantitative differences between the two policies are modest, they are consistent across all thirteen indices. More significantly, both policies operate at a level of linguistic complexity that is incompatible with the plain-language standards mandated by major data protection regulations, including the GDPR. These findings carry significant implications for regulators, consumer rights advocates, and scholars of digital privacy, suggesting that the current notice-and-consent model—as implemented by leading technology companies—fails to provide the meaningful transparency that it ostensibly promises.

Future research should extend this analysis to larger corpora of privacy policies across a wider range of industries and jurisdictions, and should complement quantitative readability analysis with qualitative discourse analysis of specific rhetorical and pragmatic features. There is also a clear need for intervention research examining whether simplified or restructured privacy disclosures can meaningfully improve consumer comprehension and informed consent in digital contexts. Based on these findings, technology companies should consider adopting plain-language drafting practices, shorter sentence structures, and simplified terminology when preparing privacy disclosures. Such measures may enhance transparency, strengthen informed consent, and better align corporate privacy policies with contemporary regulatory expectations regarding accessible communication.

References

- Bhatia, V. K. (1993). *Analysing genre: Language use in professional settings*. Longman.
- Biber, D., Conrad, S., & Reppen, R. (1998). *Corpus linguistics: Investigating language structure and use*. Cambridge University Press.
- Coleman, M., & Liao, T. L. (1975). A computer readability formula designed for machine scoring. *Journal of Applied Psychology*, 60(2), 283–284. <https://doi.org/10.1037/h0076540>
- Dale, E., & Chall, J. S. (1948). A formula for predicting readability. *Educational Research Bulletin*, 27(1), 11–20.
- Degeling, M., Utz, C., Lentzsch, C., Hosseini, H., Schaub, F., & Holz, T. (2019). We value your privacy... now take some cookies: Measuring the GDPR's impact on web privacy. *Proceedings of the Network and Distributed System Security Symposium*. <https://doi.org/10.14722/ndss.2019.23378>
- Ermakova, T., Fabian, B., & Babina, E. (2015). Readability of privacy policies of healthcare websites. *Proceedings of the 13th International Conference on Wirtschaftsinformatik*, 1085–1099.
- Fairclough, N. (1989). *Language and power*. Longman.

- Flesch, R. (1948). A new readability yardstick. *Journal of Applied Psychology*, 32(3), 221–233. <https://doi.org/10.1037/h0057532>
- Gunning, R. (1952). *The technique of clear writing*. McGraw-Hill.
- Halliday, M. A. K. (1985). *Spoken and written language*. Deakin University Press.
- Kincaid, J. P., Fishburne, R. P., Rogers, R. L., & Chissom, B. S. (1975). Derivation of new readability formulas (Automated Readability Index, Fog Count and Flesch Reading Ease formula) for Navy enlisted personnel (Research Branch Report 8-75). Naval Technical Training Command.
- Klare, G. R. (1963). *The measurement of readability*. Iowa State University Press.
- Kutner, M., Greenberg, E., Jin, Y., & Paulsen, C. (2006). The health literacy of America's adults: Results from the 2003 National Assessment of Adult Literacy (NCES 2006-483). National Center for Education Statistics.
- Lippi, M., Palka, P., Contissa, G., Lagioia, F., Micklitz, H. W., Sartor, G., & Torroni, P. (2019). CLAUDETTE: An automated detector of potentially unfair clauses in online terms of service. *Artificial Intelligence and Law*, 27(2), 117–139. <https://doi.org/10.1007/s10506-019-09243-2>
- Liu, F., Eugenio, B., Jin, H., & Strater, L. (2016). Automatic identification of quasi-identifiers in unstructured and structured data for de-identification. *Proceedings of the IEEE CBMS*, 163–168.
- McDonald, A. M., & Cranor, L. F. (2008). The cost of reading privacy policies. *I/S: A Journal of Law and Policy for the Information Society*, 4(3), 543–568.
- McEnery, T., & Wilson, A. (2001). *Corpus linguistics: An introduction* (2nd ed.). Edinburgh University Press.
- McLaughlin, G. H. (1969). SMOG grading: A new readability formula. *Journal of Reading*, 12(8), 639–646.
- Milne, G. R., & Culnan, M. J. (2004). Strategies for reducing online privacy risks: Why consumers read (or don't read) online privacy notices. *Journal of Interactive Marketing*, 18(3), 15–29. <https://doi.org/10.1002/dir.20009>
- Raschke, P., Kübler, A., & Bosch, F. (2014). Towards transparency and accountability in automated decision-making: Developing privacy policy analysis tools. *Proceedings of the Transparency, Accountability, and Trust Workshop at ECAI*.
- Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data (General Data Protection Regulation). (2016). *Official Journal of the European Union*, L 119, 1–88.
- Reidenberg, J. R., Breaux, T., Cranor, L. F., French, B., Grannis, A., Graves, J. T., Liu, F., McDonald, A., Norton, T. B., Ramanath, R., Russell, N. C., Sadeh, N., & Schaub, F. (2015). Disagreeable privacy policies: Mismatches between meaning and users' understanding. *Berkeley Technology Law Journal*, 30(1), 39–68.
- Samsung Electronics Co., Ltd. (2024, July 1). Samsung privacy policy. <https://www.samsung.com/uk/info/privacy/>
- Senter, R. J., & Smith, E. A. (1967). Automated readability index (AMRL-TR-66-22). Wright-Patterson Air Force Base.
- Solove, D. J. (2013). Introduction: Privacy self-management and the consent dilemma. *Harvard Law Review*, 126(7), 1880–1903.
- Tiersma, P. M. (1999). *Legal language*. University of Chicago Press.
- Waldman, A. E. (2018). Privacy, notice, and design. *Stanford Technology Law Review*, 21(1), 129–197.