Plain Language Assessment of Statutes

Wolfgang ALSCHNER,[a,1] Daniel D'ALIMONTE,[a] Giovanni C. GIUGA[a] and Sophie GADBOIS[a]

[a] *University of Ottawa*

Abstract. Legislative drafters use plain language drafting techniques to increase the readability of statutes in several Anglo-American jurisdictions. Existing readability metrics, such as Flesch-Kincaid, however, are a poor proxy for how effectively drafters incorporate these guidelines. This paper proposes a rules-based operationalization of the literature's readability measures and tests them on legislation that underwent plain language rewriting. The results suggest that our readability metrics provide a more holistic representation of a statute's readability compared to traditional techniques. Future machine-learning classifications promise to further improve the detection of complex features, such as nominalizations.

Keywords. Plain Language, Readability, Rules-based approach, Flesch-Kincaid.

1. Introduction

Statutes often use Latin "legalese" or complex sentences that make text difficult to understand for non-experts [1]. In response, several Anglo-American jurisdictions have recently passed guidelines and laws that require statutes to be written in "plain language".[2] These reforms built on decades of scholarly work that emphasizes the need for legal drafters to employ shorter, simpler sentences, use ordinary words in their normal sense and write in the active voice [2][3]. However, there is a disconnect between the principles developed in plain legal language laws, guidelines, or scholarship and the operationalization of legal readability checks in practice. While the former creates rules *specific* to the legal domain, the latter employs *generic* metrics developed outside of the legal context to assess readability, such as Flesch-Kincaid (FK) scores. North Carolina, Florida, and Oregon, for example, have enacted legislation that requires government documents to meet a minimum FK score.[3]

Using simplistic general-purpose metrics on statutes, such as FK scores that assess readability by counting syllables per words and words per sentences, is problematic for several reasons. First, peculiarities of legal texts, such frequent cross-references, lists, or unusual punctuation conventions, skew measures like FK, which require the clear sentence boundaries found in prose. Second, general-purpose metrics, at best, only indirectly capture plain legal language recommendations and, at worst, may be negatively correlated with them, *e.g.* when replacements of legalese with wordier ordinary terms reduce FK scores. Third, beyond using shorter words and sentences,

---

[2] For example, the Government of Canada's Department of Justice requires that new legislative texts follow its *Legistics* guidelines, the United States' *Plain Writing Act of 2010*, the United Kingdom's *The Good Law Initiative*, and the European Union's *Better Regulation Agenda*.

[3] NC Gen Stat § 58-66-30(b) (2013); FL Stat § 627.4145 (2019); OR Rev Stat § 316.364 (2019).

generic metrics, like FK, fail to provide specific guidance to drafters on how to write more readable texts [4]. This contribution seeks to remedy this disconnect by operationalizing legal plain language guidelines for statutory readability. Our metrics describe different lexical, grammatical, stylistic, and structural properties of statutory texts that, according to the plain language literature, make legal texts more readable. To validate our measures, we test them against an original dataset of before-and-after plain language rewrites from five Anglo-American jurisdictions. The results illustrate that our metrics offer a more holistic understanding of readability compared to traditional measures, but also point to the need for future research to combine rules-based and machine-learning approaches to devise readability measures specific to the legal domain.

2. Methodology

We systematically reviewed the recommendations developed in English-speaking plain language scholarship and drafting guidelines.[4] We then ranked recommendations by their frequency to identify drafting principles that enjoy widespread support across Anglo-American jurisdictions. Finally, among the top-ranking principles, we focused on those that are difficult to evaluate manually.[5] Based on these considerations we operationalized guidelines through a rules-based approach that detects a set of lexical, grammatical, stylistic and structural properties of statutory texts summarized in Table 1.

Table 1. Readability Metrics Developed from Plain Language Guidelines and Scholarship

Metric Type	Explanation
Lexical	**(1) Shall/must:** searches for instances of "shall", "shan't", "must", and "mustn't". **(2) Legalese:** calculated based on the occurrence of a subset of terms from the Black legal dictionary that are classified as non-English by the hunspell R-package dictionary spell check function and thus likely either Latin or legalese, *e.g.* "offeror".
Grammar	**(1) Compound phrases:** counted when coordinating conjunctions identified by a POS tagger and common independent marker words are recognized. These terms typically denote when clauses have been combined, which can create complex and wordy sentences. **(2) Conditional phrase:** counts sufficient and standalone core conditional indicators. Dual use conditionals, such as "any", are not considered to limit overcounting. Sufficient and necessary conditionals are not separately counted to avoid double counting paired conditionals like "if-then". **(3) Nominalizations**: counted by identifying words with common nominalization endings that are not proper nouns, legalese or statute-specific words (*e.g.* "provision"). To limit overcounting, only problematic nominalizations that are paired with the passive voice or that are preceded by a proposition and thus contribute to wordiness are counted.
Style	**(1) Passive voice:** looks for conjugations of "to be" and verbs with the past participle tense POS. Once all indices are located, a custom made matching function is employed to determine the closest past participle match to a "to be" conjugation given the "to be" conjugation must occur first, the matches have to be within a user specified proximity of one another, and the matches have to be unique with preference given to the closer matching pair. Matching pairs are counted as instances of the passive voice. **(2) All-caps:** checks whether words are written in all caps using regular expressions.
Structure	**Counts words, sentences, and syllables**. Texts are preprocessed to improve sentence boundary detection by eliminating external references with problematic punctuation, list elements and numerical characters.

---

[4] To this end, we reviewed 34 plain legal language textbooks, journal articles and official drafting guidelines from several Anglo-American jurisdictions.

[5] For example, plain language scholars recommend that statutory texts contain a table of content and are written in active voice. While the former can be easily evaluated by a human reviewer, instances of passive voice are more difficult to detect efficiently.

3. Validation

To validate how well our metrics identify plain language text characteristics, we created an original dataset of statutes from five Anglo-American jurisdictions in two versions: the originally enacted legislation ("before") and a plain language rewrite ("after"). The dataset's "after" legislation includes texts (1) written by academics but not enacted (Equality Act, Takeover Codes, Timeshares Act) and (2) enacted by a government to replace legislation (Minneapolis City Charter and Contract & Commercial Law Bill).

Table 2. Before-after Legislation Used in Dataset

Piece of Legislation	Jurisdiction
Promotion of Equality and Prevention of Unfair Discrimination Act ("Equality Act") Section 12	South Africa
Timeshares Act	United Kingdom
Contract and Commercial Law Bill	New Zealand
Minneapolis City Charter	United States
Takeover Codes	Australia

Validation proceeded in two stages. The first stage of validation consisted of iteratively comparing human and automated feature identification results for a sample of 10% of every legislation to identify errors and to refine our identification rules. As expected, the simple shall/must, total number of words, and all caps functions performed well during this initial validation. Sentence counts were initially problematic due to incorrect sentence boundary detection, but these errors could be addressed by eliminating confounding punctuations through text pre-processing. The identification of legalese was improved by stemming words in the text and the dictionary before checking for matches, which helped detect small variations of the same term. Limitations of a rules-based identification emerged most clearly with the more complex nominalization, compound phrase, and conditional phrase metrics, where our metrics, even after refinements, approximated but did not perfectly match manual feature detection (see Section 4).

Once we were confident that our metrics captured the most common categories of compound phrases, conditionals, and nominalizations we encountered during our sampling, we proceeded to the second stage of the validation to compare the before-after texts across our metrics. The results reproduced in Table 3 validate that our metrics succeed in tracking changes between the versions. The shall/must measure show a significant decrease of shall and concomitant increase of must in the plain language rewrite. In addition, the legislations' plain language versions use fewer compound phrases, nominalizations, less passive voice and fewer total words and legalese compared to the original versions. Our metrics thus capture plain language text modifications.

4. Discussion and Limitations

Our metrics offer a more nuanced representation of a statute's readability compared to FK scores and help drafters to review or rewrite statutes based on plain language criteria.

Table 3. Original and Plain Language Legislation's Readability Measures Scores

	Minneapolis City Charter		NZ Commercial Bill		SA Equality Act s.12		AU Take-over Codes		UK's Time-shares Act	
Metric	Orig.	PL	Orig.	PL	Orig.	PL	Orig.	PL	Orig.	PL
Shall	1763	0	283	0	0	0	155	0	32	0
Must	8	121	25	124	0	0	0	106	7	14
Compound Phrases	4976	837	2095	1883	8	4	1668	792	168	135
Conditional Phrases	445	122	384	509	0	0	249	198	28	38
Nominalization	490	83	412	354	1	0	491	116	82	62
Passive Voice	1138	74	652	609	3	0	716	237	41	29
Total Words	65554	12865	35066	33523	73	40	31635	13764	3531	2600
Sentence Number	1188	676	520	780	1	3	282	267	70	86
Syllables per 100 words	161	176	157	158	211	188	162	161	163	165
Legalese	186	46	159	125	1	0	317	216	50	3
All Caps	6	0	0	0	0	0	3	2	0	0

At the same time, our approach comes with limitations. First, since plain language guidelines and the formatting of statutory texts vary across jurisdictions, our rules likely require adaptation for different jurisdictions. Second, our rules-based approach could be complemented with machine learning to correctly identify more complex text features. While some of the most prominent plain language guidelines lend themselves to a simple rules-based implementation (*e.g.* shall/must), more complex features require a more nuanced approach. Rules-based detection of nominalizations based on typical word endings, for example, leads to overcounting words that have nominalization endings but do not have verbs as root words ("business") and valid nominalizations that are not used in problematic ways ("information"). Devising rules that comprehensively capture this distinction is challenging. Human expert labelling of problematic nominalizations scaled through machine learning provides an alternative avenue for identifying relevant text features. For example, Sugisaki has shown how machine learning classifiers can identify text parameters including complex noun phrases in legal texts [5]. Future research could thus combine rules-based approaches that tackle low-hanging fruit (*e.g.* shall/must) with machine learning for more challenging feature identification tasks. In combination, they provide a scalable means to operationalize plain language assessments of statutes.

References

[1]   Gail S Dykstra. Plain Language, Legal Documents and Forms: Background Information. *Can L Information Council*. 1987.

[2]   Michael EJ Masson, Mary Anne Waldron. Comprehension of Legal Contracts by Non-Experts: Effectiveness of Plain Language Redrafting. 8 (1994) *Applied Cognitive Psychology* 67.

[3]   Peter Butt, Richard Castle. *Modern Legal Drafting: A Guide to Using Clearer Language*, 2nd ed. Cambridge University Press. 2006.

[4]   Crossley, Scott A., Stephen Skalicky, and Mihai Dascalu. Moving beyond Classic Readability Formulas: New Methods and New Models. 42 (2019) *Journal of Research in Reading* 541.

[5]   Kyoko Sugisaki. Towards Data-Driven Style Checking: An Example for Law Texts. In: Florian Bex and Serena Villata (eds.). *Frontiers in Artificial Intelligence and Applications. 29th International Conference on Legal Knowledge and Information Systems (JURIX)*. 2016. p. 93-100.