Legal Knowledge and Information Systems
E. Francesconi et al. (Eds.)

doi:10.3233/FAIA220475

Measuring the Complexity of Dutch Legislation

Tim VAN DEN BELT [a] and Henry PRAKKEN [b]

[a] *Faculty of Law, University of Groningen, The Netherlands*
[b] *Department of Information and Computing Sciences, Utrecht University & Faculty of Law, University of Groningen, The Netherlands & European University Institute, Florence, Italy*

Abstract. For legislation to be effective, it should not be too complex; otherwise, it cannot be sufficiently understood by those who have to apply the law or comply with it. This paper adds to the research in AI & law on developing precise mathematical complexity measures for legislation and applying these measures by computational means. The framework of Katz & Bommarito (2014) is applied to measure the complexity of Dutch legislation. The aim is twofold: first, to investigate whether this framework is meaningfully more widely applicable by applying it to a different jurisdiction and a corpus of larger size; and second, to identify possible improvements to the framework.

Keywords. legislation, complexity analysis, Dutch law

1. Introduction

For legislation to be effective, it should not be too complex; otherwise, it cannot be sufficiently understood by those who have to apply the law or comply with it. In law and politics, the desire to constrain the complexity of legislation is often discussed but these discussions could benefit from precise measures of the complexity of the legislation. Accordingly, in AI & law research exists on developing precise mathematical complexity measures for legislation and applying these measures by computational means [1,3,2,5]. The hope underlying this research is that it will aid academic and policy discussions about the complexity of the law, resulting in more accessible and understandable legislation.

Bourcier and Mazzega [1] made a distinction between *structure-based* and *content-based* measures of complexity and discussed some possible measures of these kinds. Waltl and Matthes [5] applied several quantitative metrics of these kinds for analysing the complexity of German law. Katz & Bommarito [2] refined Bourcier and Mazzega's classification into *structure-*, *language-* and *interdependence-based* complexity measures. They then proposed a comprehensive computational framework in which several such measures are combined into an overall measure of the complexity of the legislation. They then applied the framework to measure the complexity of the *United States Code*. The framework was “motivated by the specific contours of the United States Code”, but the authors hypothesised that it is more widely applicable.

Accordingly, this paper presents an application of the framework of [2] to measure the complexity of Dutch legislation. The aim of this is twofold. First, we want to investigate Katz & Bommarito’s [2] hypothesis that their framework is meaningfully more widely applicable by applying it to a different jurisdiction, a different language and a corpus of larger size. A second aim is to identify possible additions to or improvements of their framework as used in [2].

To summarise our findings, we found that the framework of Katz & Bommarito can be applied both mathematically and computationally in our corpus of Dutch legislation. However, we found reasons to recommend that complexity measures that strongly correlate with the structural size of legislation are less useful since they may be beyond the legislator’s control.

The rest of this paper is organised as follows. In Section 2 we describe the corpus of Dutch legislation that was our study’s object and summarise the way we applied Katz & Bommarito’s framework to measure its complexity. In Section 3 analyse our complexity results and compare them with the results of Katz & Bommarito. We conclude in Section 4. The full extent of our analysis goes beyond a conference paper. Therefore, we can in this paper only present a summary of the data, method and results; the full details are available on Github.[1]

2. Corpus and Method

In this section we describe the corpus of Dutch legislation that was our study’s object and summarise Katz & Bommarito’s method of measuring complexity and our additions to and modifications of their method.

According to Katz & Bommarito, the United States Code is only a small portion of existing US law. By contrast, our data set consists of essentially the entire corpus of Dutch legislation limited to acts. To analyse its complexity, structured data is required. Our dataset consisted of a structured XML version of the corpus made available by KOOP, the knowledge and exploitation center for official Dutch government publications

The underlying idea of Katz & Bommarito’s approach is that complexity can be measured using a knowledge acquisition process where someone wants to decide whether to comply with the law. This idea is operationalised into three features: *structure*, *interdependence* and *language* of legislation.

Structure The structure of a piece of legislation is represented as a tree, where the nodes represent the elements of the act and the links capture their hierarchical relations. For the Dutch legislation we distinguished the elements ’book’, ’department’, ’title’, ’chapter’, ’paragraph’, ’subparagraph’, ’section’, ’subsection’, ’sub’. This tree is then used to define two structure-based measures. *Structural size* is the number of nodes in the tree, while *Graph depth* (by Katz & Bommarito called *Element depth distribution*) is the mean distance of all nodes to the root of the tree. In addition to Katz & Bomarrito’s, we also measured the element depth dis-

[1] github.com/TimvandenBelt/Complexity-Dutch-Legislation.

tribution of only the leaf nodes. We observed little difference in this measure by comparing the correlation and results, which differ by just 0.018. When ordering the results from highest to lowest, the ranking differs minimally.

Language Katz & Bommarito define the following measures in terms of the language of legislation. *Size* is the number of tokens within the text of an element. *Average word length* is the average number of characters of words in the text of an element (disregarding 'stop' words of several kinds). It should be noted here that, all other things being equal, average word length will be lower for English than for languages like Dutch and German, which combine words into single longer words. For instance, 'word length' translates to 'woordlengte' in Dutch. Finally, Katz & Bommarito use *Word entropy*, which informally measures the amount of textual variance of an element: does it use many different words and concepts, or is it homogeneous in these respects? They measure this in terms of the information-theoretic concept of Shannon entropy [4]. All other things being equal, the higher the word entropy of an element, the more complex it is. We also applied lemmatisation through the use of natural language processing.[2] The idea was that identical verbs or nouns might be used but in different forms and thus increasing the entropy. With lemmatisation, we morphed all words to their base form, providing, in our eyes, a better representation of the homogeneity of the text. However, we observed minimal differences with regular word entropy. In addition to the framework of Katz & Bommarito, we also use a measure of *Readability* of an element. For this, we use the so-called *Flesch reading ease* measure. It rates the readability of a text on a scale from 0 to 100, based on the average sentence length and the average number of syllables per word.[3] We use this measure as we believe it provides a more accurate representation of language complexity as it considers both word complexity and sentence complexity.

Interdependence Katz & Bommarito measure the interdependence within legislation in terms of the number of citations from one element to another. The higher the number of citations, the higher the complexity. Interdependence can be both internal (within an act) and external (between acts). Citations are represented in a directed citation graph, where the nodes are in [2] sections, and citations below-section nodes are attributed to section level from all 'titles' in the corpus, while in our case, they are at section level and below. The reason for this difference is that we believe that some below-section nodes may be of similar size or larger than some section nodes. We also believe it provides a more factual representation and may yield a more accurate network analysis. The links in the citation graph are citations from one element to another. Within-element citations in a title (in [2]) or act (in our analysis) are represented by subgraphs where all nodes are from the same title, respectively, act. Katz & Bommarito distinguish between explicit citations and the use of definitions from one element by another element. Due to time constraints and limits in the data, we have only considered explicit citations, excluding definitions. We measure *internal interdependence* of an act by counting the number of citations that cite another element in the same act.

[2]For which we used Spacy: https://spacy.io/.

[3]For detecting syllables, we used Spacy along with a community package: https://spacy.io/universe/project/spacy_syllables

We then normalise this against the structural size of the act by dividing the number of citations by the number of nodes in the hierarchical graph of that act. For measuring *external interdependence* between titles, Katz & Bommarito distinguish between titles exporting information (by being cited by another title) and titles importing information (by citing another title). They then measure the numerical difference ("net flow") between the number of imports and exports of a title. They also consider a normalised version "net flow per section" relative to title size. We apply the same methods to acts and their sections.

Waltl and Matthes [5] used several of the above-discussed measures, namely, section-nodes, number of words, element depth, internal interdependence, and a variation of external interdependence. In addition, they measured language complexity in terms of indeterminacy and vocabulary variety. Vocabulary variety can be compared to word entropy. Indeterminacy was outside our scope due to time constraints. Unlike [2] and us, [5] did not use a composite complexity measure.

Composite measures Katz & Bommarito then use these measures to define two composite measures. Both choose one measure from each of the three categories structure, language and interdependence. For their *unnormalised composite measure* they choose structural size, word entropy and net flow while for their *normalised composite measure* they choose mean element depth, word entropy and net flow per section. For both composite measures they then rank each title with each of these individual measures. Finally, they combine the three rankings thus obtained by computing the average rank of each title, acknowledging that other methods might be more suitable.

We used the same unnormalised composite measure, but we replaced word entropy with Flesch readability in their normalised composite measure. The reason for this is that, in our opinion, word entropy is not suitable for a normalised composite since it correlates too strongly with the size of the legislation.

3. Results & Analysis

We gauged each measure and calculated the correlation of most in relation to the structural size of legislation. Thereafter, just as [2], we used two composites to rank the legislation, with some minor adjustments. These results can be found on Github.[4] In total, 1120 acts were analysed.

In this section we analyse our results and compare them to those of Katz & Bommarito. As regards the normalised and unnormalised rankings, it is interesting to observe that as in [2], some acts rank similarly in these two rankings while for other acts there are considerable differences in rank (although still within the same region). Apart from this, an absolute comparison between [2] and our analysis on the various criteria is not very informative, because of the differences between the Dutch and English languages and the differences in legislation style between the Dutch and US jurisdictions. We, therefore, focus on correlation analysis. While Katz & Bommarito performed two correlation analyses, we did several more. Table 1 summarises our correlation results. We in particular investigated

[4] github.com/TimvandenBelt/Complexity-Dutch-Legislation.

Table 1. Correlation results ordered by highest to lowest R value.

Correlation	P value	R value	R squared %
Size & text nodes	0.000	0.998	99.68
Size & number of words	0.000	0.978	95.59
Size & number of tokens	0.000	0.977	95.48
Size & non-text nodes	0.000	0.973	94.65
Size & below-section nodes	0.000	0.964	92.95
Size & section nodes	0.000	0.961	92.37
Size & mean depth	0.000	0.926	85.82
Size & word entropy	0.000	0.925	85.58
Size & lemmatised word entropy	0.000	0.921	84.77
Size & citations total	0.000	0.919	84.42
Size & mean leaf depth	0.000	0.908	82.39
Size & internal citations	0.000	0.894	80.00
Sections & below-section nodes	0.000	0.870	75.64
Sections & above-section nodes	0.000	0.868	75.41

Correlation	P value	R value	R squared %
Size & above-section nodes	0.000	0.867	75.24
Size & external citations	0.000	0.843	71.13
Above-section & below-section nodes	0.000	0.795	63.31
Size & tokens per section	0.000	0.580	33.66
Size & net flow	0.000	0.408	16.66
Size & word length	0.000	0.111	1.24
Size & net flow per section	0.281	-0.032	0.10
Flesch & tokens per section	0.046	-0.060	0.36
Size & Flesch	0.000	-0.105	1.11
Flesch & number of words	0.000	-0.120	1.44
Flesch & word length	0.000	-0.604	36.53

the correlation of the various other measures with the structural size of the legislation. The motivation for this is that if a measure strongly correlates with the size of legislation, the measure may be beyond the legislator's control. A legislator can, of course, attempt to lessen the size of the legislation, but this might render the legislation less effective in practice, which harms instead of improves the quality of legislation. It may therefore be argued that measures that strongly correlate with the size of legislation are less useful as measures of the complexity of legislation. After all, a practical motivation for developing complexity measures is to support legislators in making legislation more accessible and understandable.

Katz & Bommarito found that size was at best weakly correlated with mean element depth. Our results show a stronger correlation with more statistical significance. Katz & Bommarito found that size strongly correlates with the number of sections. Our results are nearly identical with more statistical significance. Additionally, we observed that the measures text nodes, number of words, number of tokens, non-text nodes, below-section nodes, section nodes, mean depth, word entropy, lemmatised word entropy, citations total, mean leaf depth, above-section nodes and external citations either strongly or decently correlate with the size of legislation. Size and tokens per section very weakly correlate with the structural size of legislation. Net flow, word length, net flow per section and Flesch do not seem to correlate with the (structural) size of legislation. [5] also found that Flesch does not correlate with the number of words.

4. Conclusion

In this paper we have reported on an experiment to investigate whether the complexity framework of Katz & Bommarito [2] can be meaningfully used to analyse the complexity of Dutch legislation. We found that this is possible both mathematically and computationally. We also compared our results to those of Katz & Bommarito. Since an absolute comparison in terms of the complexity numbers is not very informative because of differences between the Dutch and English language and legislation style, we mainly focused on correlation analysis. By and large, our correlation results were similar to the results in [2] but with higher statistical significance because of a higher number of legislative documents.

We also did several correlation analyses not done by Katz & Bommarito, particularly to see which complexity measures correlate with the size of legislation. This was motivated by the idea that complexity measures that strongly correlate with the structural size of legislation are less useful as measures of complexity since they are largely beyond the legislator's control. This means that the underlying idea of Katz & Bommarito to measure complexity in knowledge acquisition costs requires refinement, especially since the underlying aim of their work is to support legislators in making legislation less complex. In fact, our recommendation can be generalised to any measure that strongly correlates with features of legislation that are beyond the legislator's control. In light of this, we believe that our additional correlation analyses are a vital addition to the analysis of Katz & Bommarito, who did a correlation analysis for just two of their measures, namely, size versus sections & mean element depth. The results of our correlation analysis motivated us to recommend the replacement of the word entropy measure with the Flesch readability score in the normalised ranking composite since, unlike word entropy, the Flesch readability score only negligibly correlated with legislation size.

We end by mentioning two limitations that our approach shares with that of Katz & Bommarito. First, the framework gives only relative measures of complexity and no measures of when legislation is too complex. Second, the choice of composite measures is not yet governed by clear and convincing criteria. These issues should be addressed in future research.

References

[1] D. Bourcier and P. Mazzega. Toward measures of complexity in legal systems. In *Proceedings of the Eleventh International Conference on Artificial Intelligence and Law*, pages 211–215, New York, 2007. ACM Press.

[2] D.M. Katz and M.J. Bommarito. Measuring the complexity of the law: The United States Code. *Artificial Intelligence and Law*, 22:337–374, 2014.

[3] M. Palmirani and L. Cervone. Measuring the complexity of law over time. In P. Casanovas, U. Pagallo, M. Palmirani, and G. Sartor, editors, *AI Approaches to the Complexity of Legal Systems: AICOL 2013 International Workshops, Revised Selected Papers*, number 8929 in Springer Lecture Notes in AI, pages 82–99, Berlin, 2013. Springer Verlag.

[4] C. E. Shannon. Prediction and entropy of printed English. *The Bell System Technical Journal*, 30(1):50–64, 1951.

[5] B. Waltl and F. Matthes. Towards measures of complexity: Applying structural and linguistic metrics to German laws. In R. Hoekstra, editor, *Legal Knowledge and Information Systems. JURIX 2014: The Twenty-seventh Annual Conference*, pages 153 – 162. IOS Press, Amsterdam etc., 2014.