Measuring the Complexity of Dutch Legislation (Extended Version)

Tim van den Belt[1] and Henry Prakken[2]

[1] Dutch Authority for Digital Infrastructure, The Netherlands
[2] Department of Information and Computing Sciences, Utrecht University, The Netherlands

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.
It was found that Katz & Bommarito's framework can be applied to Dutch legislation. However, it is argued that complexity measures that strongly correlate with the structural size of legislation are less useful since they may be beyond the legislator's control. To this end, additional correlation results to those of Katz & Bommarito are reported and it is recommended that their knowledge-acquisition-cost approach to measuring the complexity of legislation is refined by taking the possibility of the legislator's control into account.

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. For example, in a study done by the Dutch Ombudsperson, 38% of the respondees pointed towards the legislation as a cause of the incomprehensibility of the government, of which 31.5% found it to be incomprehensible [10]. Accordingly, in AI & law research exists on developing precise mathematical complexity measures for legislation and applying these measures by computational means [4,11,7,15]. 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 [4] made a distinction between *structure-based* and *content-based* measures of complexity and discussed some possible measures for both kinds. Waltl and Matthes [15] applied several quantitative metrics for these kinds of complexity of German law. Katz & Bommarito [7] refined Bourcier and Mazzega's classification into *structure-*, *language-* and *interdependence-based* measures. They then proposed a comprehensive computational framework in which several structural, interdependence-based and language-based 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 [7] to measure the complexity of Dutch legislation.[3] The aim of this is twofold. First, we want to investigate Katz & Bommarito's [7] hypothesis that their framework is 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 [7].

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 we present our results and in Section 4 we analyse these results and compare them with the results of Katz & Bommarito. We conclude in Section 6. The full extent of our analysis goes beyond a workshop paper. Therefore, we can in this paper only present a summary of the data, method and results; the full details are available on Github.[4]

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.

2.1 The corpus of Dutch legislation

According to Katz & Bommarito, the United States Code is only a small portion of existing US law. By contrast, our data set consists of the entire corpus of Dutch legislation limited to acts. To analyse, structured data is required. Our

[3] This paper is an extended version of [1].
[4] `https://github.com/TimvandenBelt/Complexity-Dutch-Legislation`.

dataset consisted of a structured XML version of the corpus made available by KOOP, the knowledge and exploitation center for official Dutch government publications. For an analysis of the complexity of law over time as induced by explicit modification provisions in legislation see [11].

2.2 The method

We next summarise Katz & Bommarito's framework and how we have used it in our analysis. The underlying idea of the framework 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 research, we also measured the element depth distribution of only the leaf nodes. We observed little difference in this measure by comparing the correlation result, which differs by just 0.018. When ordering the results from highest to lowest, the ranking differs minimally. Katz & Bommarito theorise that a larger corpus or one with more depth requires more effort to assimilate the necessary knowledge. A larger corpus means more text to process. A deeper corpus could indicate a higher legal specificity, meaning more technical legal concepts, and therefore a higher cost. However, this approach does not take into account that a hierarchical structure could also lessen knowledge acquisition costs as it provides oversight and the possibility of an index. Additional investigation is required to address this paradox.

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[5] *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. Additionally, this and the other language metrics are aimed at western

[5] Katz & Bommarito define tokens as following: "*contiguous string of text*". For clarity, in addition to that definition, we explicitly exclude punctuation, spaces, brackets or other symbols. We recognise their importance in language but they provide no substance on their own and are therefore excluded to count towards the structural size of an act.

and Latin-based languages and may not function accurately for other linguistic systems. More investigation is needed in this field. Finally, Katz & Bommarito use *Word entropy*, which informally measures the amount of textual variance of an element: does it use many different words and therefore more different legal concepts, or is it homogeneous in these respects? They measure this in terms of the information-theoretic concept of Shannon entropy [13]. All other things being equal, the higher the word entropy of an element, the more complex it is as it might indicate more diverse legal concepts. We also applied lemmatisation through the use of natural language processing.[6] 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 in ranking and correlation 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, earlier used by [9,15]. 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.[7] We use this measure as we believe it provides a more accurate representation of language complexity as it considers both word 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. For each citation, one needs to exert more effort to acquire the necessary knowledge by traversing to the directed element and processing said element. However, Katz & Bommarito did not take into account that using citations can simplify legislation, because using citations can prevent recursive texts and therefore lower the structural (text) size of a corpus. Additional investigation is required to address this paradox. 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 [7] 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 [7]) 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. For example, an explicit citation is section 12 of title 51 (patents), which states

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

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

> *... at the rate for each year's issue established for this purpose in **section 41(d)**.*

An example of the use of a definition from another element is the use of section 351, paragraph 1 of title 35 (patents), which states

> *The term "**treaty**" means the Patent Cooperation Treaty done at Washington, on June 19, 1970.*

which is used in paragraph b:

> *The term "Regulations", when capitalised, means the Regulations under the treaty, done at Washington on the same date as the* **treaty**. *The term "regulations", when not capitalised, means the regulations established by the Director under this title.*

Due to time constraints and limits in the research data, we have only considered explicit citations, excluding definitions. We measure *internal interdependence* within an act by counting the number of citations that cite another element in the same act. Additionally, we describe a third form of extracting citations—an implicit form of citation by utilising intermediaries. For example, the Dutch Penal Code consists of three books. The second book governs crimes, while the third book governs misdemeanours. The Dutch Criminal Code of Procedures sometimes provides different regimes for crimes and misdemeanours, which means that many crimes and misdemeanours have an intermediary interdependence with citations through a non-explicit relation with in this case an above section element, which is required to read for a thorough knowledge acquisition process. 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 [15] 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 [7] and us, [15] 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

Below a summarised overview of the top results can be found. We noted the top 5 highest and lowest scoring acts for some measures as presented in Katz & Bommarito's research and some of our additional measures. Here, V is as in [7] the set of all nodes of the act's structural hierarchy, while V^* is the set of above-section nodes, V_S is the set of section-level nodes, and V_* is the set of below-section nodes. Every act consists of a root node that is the tree hierarchy's top. It is not included in the number of above-section, below-section and section-level nodes but is included in the total number of nodes of legislation. All English act names are informal translations by us. A complete list of results can be found on Github.[8] In total, 1120 acts were analysed.

Table 1. Totals and averages of the Dutch legislation.

Variable	**Totals**	**Average**
Acts	1,120	n.a.
V	254,973	227
V^*	12,098	10
V_S	182,005	53
V_*	59,750	162
Words	6,726,185	6,006
Tokens	6,840,400	6,108
Sentences	294,015	263
Citations	128,954	115

Table 2. Top 5 highest and lowest scoring acts based on structural size. We omitted part of the lowest as the next 18 acts consisted of 2 nodes, of which 1 section node.

Act	V	V^*	V_S	V_*
Financial Supervision Act	7,597	407	1,257	5,932
Code of Criminal Procedure	5,573	224	1,112	4,236
Code of Civil Procedure (applies to non-digital litigation)	4,111	165	1,488	2,457
Civil Code Book 7	4,054	132	838	3,083
Code of Civil Procedure (applies in the case of digital litigation)	3,631	163	998	2,469
Act on the acceptance of the Statute for the Kingdom of the Netherlands	2	0	1	0
Foundation LOTT Act	2	0	1	0

[8] `https://github.com/TimvandenBelt/Complexity-Dutch-Legislation`.

Table 3. Top 5 highest scoring acts based on mean element depth. Similar to structural size, we omitted the lowest scoring as they all scored 0.5 or 0.67.

Act	Mean element depth
BES Civil Code Book 8	5.671
Financial Supervision Act	5.617
BES Primary Education Act	5.515
Primary Education Act	5.494
Secondary Education Act	5.474

Table 4. Top 5 highest and lowest scoring acts based on tokens.

Act	Tokens	Tokens per section
Financial Supervision Act	227,762	181.19
Code of Criminal Procedure	156,720	140.94
Civil Code Book 7	122,716	146.44
Civil Code Book 2	110,797	188.75
Code of Civil Procedure (applies to non-digital litigation)	106,712	71.72
Act on the acceptance of the Statute for the Kingdom of the Netherlands	16	16
Implementation law on the registration of beneficial owners of trusts and similar legal arrangements	20	6.67
Act on Electronic communications with the administrative court	35	8.75
Act on Extension of Total Duration of Conditional Cessation of State-Owned Nursing	44	14.67
Land-Based Growth Milk Cattle Farming Act	48	16

Table 5. Top 5 highest and lowest scoring acts based on average word size.

Act	Avg. word size
Judicial Classification Act	8.22
Act on Adjustment of benefits for government employees	7.92
Workforce Participation Incentive Act	7.67
LNV Legislation (Law framework on Independent Administrative Bodies)	7.57
BES Ministerial Regulations Approval Act	7.47
Act on derogation from adaptation mechanism of the Individual Rent Subsidy Act subsidy period 1-7-1995 until 30-6-1996	4.71
Act on continuation of temporary increase in contributions granted under the Individual Rent Allowance Act	4.73
Act on temporary increase of contributions Law on individual rent subsidy	4.73
Repeal Act on the Regulation of State Education in the Visual Arts	4.77
Land-Based Growth Milk Cattle Farming Act	4.77

Table 6. Top 5 highest and lowest acts based on Flesch score.

Act	**Flesch score**
Act on the Public Administration Council	54.56
Act on the Weather corps	53.69
Act on the Acceptance of the Statute for the Kingdom of the Netherlands	53.12
Swearing-in of Ministers and Members of the States General Act	52.31
Act on Extension of Total Duration of Conditional Cessation of State-Owned Nursing	49.72
Act on Provincial division of the Wadden Sea	-58.31
Kingdom Act containing rules concerning the legal status of some military judicial officers	-35.82
Act on Adjustment of Benefits for Government Employees	-34.95
Dutch Code of Criminal Procedure Implementation Act	-29.29
Foundation LOTT Act	-24.77

Some of the Flesch results are outside the designed range of $0-100$. The formula allows scoring below zero. Results below 0 are dreadful to read. For example, the 'Act on the provincial division of the Wadden Sea' scored -58.3.[9] Sections 2, 3 and 4 consists of a single sentence averaging 356 words per sentence.

Table 7. Top 5 highest and lowest scoring acts based on entropy.

Act	**Entropy**
Civil Code Book 7	7.21
Civil Code Book 6	7.02
Code of Civil Procedure (applies in the case of digital litigation)	6.98
Code of Civil Procedure (applies to non-digital litigation)	6.97
Civil Code Book 3	6.96
Act on the acceptance of the Statute for the Kingdom of the Netherlands	1.79
Implementation Act on registration of beneficial owners of trusts and similar legal arrangements	2.21
Code of Criminal Procedure Implementation Act	2.43
Act on electronic communications with the administrative court	2.65
Hague Convention Implementation Act	2.67

Table 8. Top 5 highest and lowest scoring acts based on net flow.

Act	Net flow	Net flow per section
Economic Offences Act	1,482	18.76
Extradition Act	355	4.93
Government Employees Benefits Act	260	2.83
Income Tax Implementation Act 2001	236	7.15
Act on Transitional arrangements for incapacity insurance	207	2.96
General Administrative Law Act	-3,111	-5.88
Penal Code	-1,569	-2.32
Code of Criminal Procedure	-1,099	-0.99
Code of Civil Procedure (applies in the case of digital litigation)	-1,050	-1.05
Civil Code Book 2	-606	-1.03

Table 9. Top 5 highest and lowest scoring acts based on unnormalised complexity composite score.

Act	Rank
Code of Civil Procedure (applies in the case of non-digital litigation)	1
Environmental Management Act	2
Telecommunications Act	3
Bankruptcy Act	4
Electricity Act 1998	5
Act to lower the WNT remuneration ceiling	1,120
Act on the acceptance of the Statute for the Kingdom of the Netherlands	1,119
Approval Act regulation on the disapplication of energy investment deduction and environmental investment deduction	1,118
Act on territorial jurisdiction of some notaries	1,117
Act authorising participation by the Netherlands in the sixth resource addition to the African Development Fund	1116

Table 10. Top 5 highest and lowest scoring acts based on normalised complexity composite score.

Act	Rank
Sustainable Energy and Climate Transition Storage Act	1
Act on the municipal division of the Wadden Sea	2
Empty Premises Act	3
Act on temporary measures for deduction of costs of major maintenance and painting	4
Curaçao and Sint Maarten Financial Supervision Kingdom Act	5
Act to lower the WNT remuneration ceiling	1,120
BES Hazardgames Act I	1,119
BES Foreign Companies Recognition of Legal Personality Act	1,118
Electronic Signatures Act	1,117
Services Directive Adaptation Act	1116

4 Analysis

Table 11. Correlation results ordered by highest to lowest.

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
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

In this section we analyse our results and compare them to those of Katz & Bommarito [7]. We gauged each measure and calculated the correlation of most in relation to the structural size of legislation. Thereafter, just as [7], we used two composites to rank the legislation, with some minor adjustments. As regards the normalised and unnormalised rankings, it is interesting to observe that as in [7], 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 [7] and our analysis of 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 11 summarises our correlation results. We in particular investigated 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. Having said so, this observation does not imply that structural size has no value as a criterion. Omitting structural size in attempts to lessen complexity may result in an unwanted increase in legislation's size.[10]

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. [15] also found that Flesch does not correlate with the number of words.

5 Related & future research

In this section, we discuss other related work and provide some suggestions for future research.

Several studies exist on applying network theory to law [2,3,6,8]. These theories encompass betweenness centrality, eigenvector and other community metrics. These measures can be used to denote the role of specific acts and therefore their possible importance and complexity. This knowledge can possibly further enhance both the definition of complexity as well as assist the legislator in setting priorities in policy making.

Waltl and Matthes [15] used the indeterminacy of words to indicate possible complex legal concepts. Indeterminacy describes words or concepts which are multi-interpretable or, sometimes as intended, vague. This enables the term or concept to fit an interpretation suitable for a given situation. According to Waltl and Matthes it increases the difficulty of interpretation. "*Consequently, those words indicate the complexity of legal texts*" [15]. For example, section 6:11 of the 'General Administrative Law Act' dictates that:

> *With regard to a notice of objection or appeal submitted after the expiry of the time limit, it shall not be declared inadmissible on that ground if it cannot* ***reasonably*** *be considered that the petitioner was in default.*

[10] In 1991, the Australian Social Security Act was revised to simplify the legislation for individuals. However, as [14] and [5] described, due to among other things, its sheer volume, it was not very successful.

What 'reasonably' encompasses can differ widely on a case-to-case basis. Their study was limited to 62 identified legal terms. This measure would expand the scope of Katz & Bommarito's model with normative complexity. We suggest investigating whether NLP is sufficiently advanced to identify such indeterminacy, possibly assisted by research conducted on the normative costs of legal words or concepts.

Furthermore, we suggest expanding the refinement of language complexity in the legal context. Language is very tangible for a legislator to adjust. Some patterns may be characteristic of the legal language. With the improvements in NLP, perhaps a legal model can be developed to identify these characteristics. Some characteristics we believe contribute to complexity are two or more (collated) negations and the ambiguity of long segmented sentences with unclarity on the intended application target of a segment of the sentence.

Finally, it would be interesting to perform user studies to see if their performance agrees with our complexity measures, or to consult with drafters and lawyers to investigate to which extent the various criteria are meaningful to them.

6 Conclusion

In this paper we have reported on an experiment to investigate whether the complexity framework of Katz & Bommarito [7] 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 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 [7] but with higher statistical significance because of a higher number of legislative documents.

A valid question is what additional insights are gained by a rather direct application of the method of Katz & Bommarito to another corpus of legislation. As we mentioned in the introduction, their framework was "motivated by the specific contours of the United States Code", but the authors hypothesised that it is more widely applicable. One of our aims was to test this hypothesis by applying their approach to a different jurisdiction, a different language and a corpus of larger size. This is not exactly the same as empirical replication studies in the social sciences. There the aim is to test whether significant results found with one group of test subjects can be replicated for another group to increase confidence in the results [12]. The experiment of Katz & Bommarito did not yield experimental results of that kind so a 'standard' replication study was impossible. Nevertheless, we believe that our repeated application of their method on another corpus has yielded additional insights. First, the fact that in our experiment we did not run into unforeseen problems is some evidence that Kats & Bommarito's method is more widely applicable. In addition, we performed several correlation analyses not done by Katz & Bommarito.

We also did several correlation analyses not done by Katz & Bommarito, particularly to see which complexity measures correlate with the size of the legislation. This was motivated by the idea that 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. For example, simply averaging scores ignores the possibility that some criteria should receive more weight than other criteria. These issues should be addressed in future research.

References

1. Van den Belt, T., Prakken, H.: Measuring the complexity of Dutch legislation. In: Francesconi, E., Borges, G., Sorge, C. (eds.) Legal Knowledge and Information Systems. JURIX 2022: The Thirty-fifth Annual Conference. p. 243–248. Frontiers in Artificial Intelligence and Applications, IOS Press, Amsterdam, The Netherlands (2022)
2. Boulet, R., Mazzega, P., Bourcier, D.: A network approach to the French system of legal codes—part i: analysis of a dense network. Artificial Intelligence and Law **19**(4), 333–355 (11 2011)
3. Boulet, R., Mazzega, P., Bourcier, D.: Network approach to the French system of legal codes part II: The role of the weights in a network. Artificial Intelligence and Law **26**(1), 23–47 (3 2018)
4. Bourcier, D., Mazzega, P.: Toward measures of complexity in legal systems. In: Proceedings of the Eleventh International Conference on Artificial Intelligence and Law. pp. 211–215. ACM Press, New York (2007)
5. Federal Court of Australia: Secretary, Department of Family and Community Services v Geeves (Jun 2004), `https://www8.austlii.edu.au/cgi-bin/viewdoc/au/cases/cth/FCAFC/2004/166.html`, jurisdiction: Commonwealth
6. Fowler, J.H., Johnson, T.R., Spriggs, J.F., Jeon, S., Wahlbeck, P.J.: Network analysis and the law: Measuring the legal importance of precedents at the U.S. Supreme Court. Political Analysis **15**(3), 324–346 (2007)

7. Katz, D., Bommarito, M.: Measuring the complexity of the law: The United States Code. Artificial Intelligence and Law **22**, 337–374 (2014)
8. Lyte, A., Slater, D., Michel, S.: Network measures of the United States Code. Tech. rep., The MITRE Corporation (2015)
9. Morgenstern, L.: Toward automated international law compliance monitoring (TAILCM). Tech. rep., Defense Technical Information Center (2014). https://doi.org/10.21236/ADA608547
10. Nationale Ombudsman: Onze onbegrijpelijke overheid. Bureau Nationale Ombudsman (2013), ombudslezing
11. Palmirani, M., Cervone, L.: Measuring the complexity of law over time. In: Casanovas, P., Pagallo, U., Palmirani, M., Sartor, G. (eds.) AI Approaches to the Complexity of Legal Systems: AICOL 2013 International Workshops, Revised Selected Papers. pp. 82–99. No. 8929 in Springer Lecture Notes in AI, Springer Verlag, Berlin (2013)
12. Pashler, H., Wagenmakers, E.: Editors' introduction to the special section on replicability in psychological science: A crisis of confidence? Perspectives on Psychological Science **7**(6), 528–530 (2012), pMID: 26168108
13. Shannon, C.E.: Prediction and entropy of printed English. The Bell System Technical Journal **30**(1), 50–64 (1951)
14. Sutherland, P.: Social Security Overpayments and Debt Recovery: Key Developments (Jul 2017). https://doi.org/10.2139/ssrn.3289416, https://papers.ssrn.com/abstract=3289416
15. Waltl, B., Matthes, F.: Towards measures of complexity: Applying structural and linguistic metrics to German laws. In: Hoekstra, R. (ed.) Legal Knowledge and Information Systems. JURIX 2014: The Twenty-seventh Annual Conference, pp. 153 – 162. IOS Press, Amsterdam, The Netherlands (2014)