On the Reform of the Italian Constitution: an Interdisciplinary Text Readability Analysis

Calogero Jerik Scozzaro[1,2,*], Matteo Delsanto[1], Antonio Mastropaolo[2], Enrico Mensa[1], Luisa Revelli[2] and Daniele Paolo Radicioni[1,*]

[1]*Università degli Studi di Torino, Italy*
[2]*Università della Valle d'Aosta, Italy*

Abstract
This work can be considered as an instant paper: on June 18, 2024 the Constitutional Reform Bill presented by the Italian Government last November and reviewed by the Senate's Constitutional Affairs Committee on the "premiership" received its first approval by the Italian Senate. We present an analysis aimed at linguistically and computationally characterizing the readability of the text amendments now being discussed. It puts together evidences from different perspectives: legal and linguistic analysis, traditional readability indices, a novel attempt to define readability through the prediction of reading times; all such perspectives are compared with the output obtained by prompting GPT to take into consideration also the output of that language model. The proposed analyses can be intended as a technical contribution to the reflection on issues fundamental to democracy in Italy and beyond, concerning the need to analyze the quality of the writing of such fundamental documents for the democratic life of states.

Keywords
Text Readability, Text Simplification, Reading Times, Constitutional Reform Analysis.

1. Introduction

The Italian Constitution is the cornerstone of the country's legal and political system: it establishes the framework for government, delineates the separation of powers, and guarantees the fundamental rights and freedoms of individuals. Its role in the Italian legislative system is multifaceted, serving as the supreme law and acting as a source of legitimacy for all laws and regulations: any law or regulation that contradicts the Constitution can be declared unconstitutional and void by the Constitutional Court (Corte Costituzionale). Such supremacy ensures that all legislative and executive actions conform to constitutional principles. The Constitution also embeds the principles of democracy, ensuring that the government is elected by the people and that sovereignty is exerced by citizens through the Italian Parliament. Additionally, it includes provisions to protect political pluralism and to prevent the concentration of power.

This study proposes an analysis of the Constitutional Reform Bill, under deliberation by the Italian Parliament since mid-November 2023: the proposed reform impacts on Articles 59, 88, 92, and 94 of the Constitution affecting relevant topics, such as the direct election of the President of the Council of Ministers, and deeply modifies the role of the President of the Republic.[1] Ensuring the readability of this text would be of the utmost importance, since the comprehensibility of the basic democratic elements being modified is a basic pillar of the democratic system. This study thus provides an analysis of how clear and readable is the set of articles as reformulated in the present Constitutional Reform Bill. In general, legislative and regulatory text documents contain complex, highly specialized language, lengthy sentences that are typically considered as difficult to understand. It is featured by specific semiotic and linguistic conventions, vocabulary, semantics, syntax and morphology that may result as difficult to understand by laypeople with no domain expertise. The drafting process of the Italian

---

[1]https://www.senato.it/leg/19/BGT/Schede/Ddliter/testi/57694_testi.htm

CEUR-WS.org/Vol-3877/paper14.pdf

Constitution, however, also reflects a linguistic refinement process, targeted at simplifying the text as much as possible: normative clarity was reputed a necessary device for making the constitutional text understandable and knowable according to democratic ideals [2].

In the following we present a pool of analyses and measures to assess the readability of the Constitutional Reform Bill. We start by briefly recalling a bit of history of the Italian Constitution (Section 2); in Section 3 our analyses are presented: we first provide a linguistic and legal analysis of the meaning of the proposed amendments to lay the foundations for comparisons with automatic approaches (Section 3.1). In Section 3.2 we illustrate a first assessment of the readability of the proposed amendments based on well-known readability indices, and introduce a novel measure of text readability based on the prediction of reading times, and report and comment the output of a prompting session involving Chat-GPT4o [3, 4].

These are the main contributions of the work: unfortunately enough, it attests converging evidence that the text being proposed is poor, regards as the quality of the writing; secondly, it provides a working example of how different disciplines may join efforts providing as many (possibly complimentary) analyses. Finally, we introduce a novel readability metric, whereby the readability of a text is estimated based on the prediction of reading times. This metric is fully different from existing approaches, as it relies on training a system on eye-tracking (low-level behavioral) data, also refined through the adoption of a language model fine-tuned on a corpus of Regional regulations, under the assumption that reading times are a direct function of the cognitive load implied in text understanding, and can thus be employed as a proxy for text readability.

2. Background: History and Main Linguistic Traits of the Italian Constitution

After the election of the Constituent Assembly in June 1946, the process of creating the Italian Constitution began with the establishment of a Committee tasked with drafting and proposing the initial Constitution project. This Committee was supported by a drafting committee responsible for refining the text, addressing both legal and linguistic issues, and ensuring sufficient normative clarity necessary to make the constitutional text understandable and knowable according to democratic ideals [2, 136]. The Italian Constitution is the result of this legal and linguistic clarification effort, aimed at reaching a community of speakers that was still largely dialect-speaking and poorly literate [5].

From a syntactic perspective, the brevity of sentences in the Italian Constitution, averaging 19 words: structures are deliberately kept as simple as possible. Sentences composed of a single clause are predominant, and exhibit a regular arrangement of components [6]. In cases of complex sentences, coordination is preferred over subordination [7]. The commitment of the drafters to create a linear and accessible text is also evidenced by the sparing use of the subjunctive mood, employed only 26 times, and primarily in contexts where it is mandatory [8]. Regarding the lexicon, $1,002$ of the $1,357$ lemmas used in the Italian Constitution belong to the basic vocabulary, which comprises the most frequent and familiar words for Italian speakers. Their overall recurrence, accounting for 92.13% of the text ($9,369$ tokens in the original text; $10,717$ in the currently effective version) to a good extent ensures lexical readability [9]. The overall semantic accessibility also depends on collocations and contextual meanings. The presence of legal phrases with technical-scientific connotations and a significant number of redefinitions –common words used with specific sectoral meanings– may create comprehension difficulties or interpretive misunderstandings, particularly for younger or less competent speakers [10, 11, 12].

The following analysis focuses on the set of articles with the proposed amendments, as presently (as of June, 2024) approved by the Italian Senate.[2]

---

[2]These are available at the URL https://www.senato.it/service/PDF/PDFServer/BGT/01414450.pdf.

3. Analysis of the Constitutional Reform Bill

The proposed amendments to the five articles of the Italian Constitution result in an overall textual and structural increase. The total number of paragraphs increases from 14 to 20, the number of words from 279 to 581, and the number of clauses from 16 to 26 (excluding transitional provisions). The average sentence length of 22.3 words is higher than for the corresponding portion of the original text (18.4 words per sentence), though it still adheres to the controlled writing guidelines, which recommend not exceeding 20-25 words per sentence [13, 14]. The deviation from the brevity principles that guided the original Constitution drafters is more evident when considering that 4 of the newly proposed clauses are composed of more than 40 words, and 2 of these exceed 50 words.

3.1. Analysis of the Amendments to Articles 92 and 94

The first clause of the second paragraph of Article 92 consists of 41 words. It requires the reader to understand the legal phrase "universal and direct suffrage."[3] Within the same Article 92, the third paragraph (53 words) provides details on the procedures for the election of the Chambers and the President of the Council of Ministers through a long clause (26 words) introduced by a gerund and marked by a closing comma, which disrupts the syntactic-semantic linearity of the sentence.

The twenty paragraphs included in the amendment proposals vary from of a minimum of 13 words (Art. 88, paragraph 1) to a maximum of 69 words (Art. 94, paragraph 3). The paragraphs containing the main innovations compared to the original constitutional text are designed to introduce procedural methods intended to be legally unambiguous. Such an objective is pursued in some cases through the use of redundancy mechanisms: role designations –in their complete formulation, considering that the title of 'president' is attributed to multiple offices– are therefore repeated. The resulting cumbersome effect of this solution is notable in paragraph 3 of Art. 94, presented below along with a possibly simplified rewrite.

> [Art. 94, paragraph 3: 3 propositions; 68 words; 381 chars][4] Within ten days since its formation, the Government presents itself to the Chambers to obtain their trust vote. If the trust in the Government led by the elected President is not voted, the President of the Republic renews the mandate to the elected President to form the Government. If, even in this case, the Government does not obtain the trust vote of the Chambers, the President of the Republic proceeds to dissolve the Chambers.

Simplified rewrite (3 propositions; 59 words; 301 characters):[5]

> After being formed, the Government has ten days to present itself to the Chambers and seek their trust vote. If trust is not granted, the President of the Republic once again tasks the elected President with forming the Government. If, even in this second case, the Government does not obtain favourable trust vote, the President of the Republic dissolves the Chambers.

From a methodological point of view, one of the authors (expert in linguistic sciences) produced the simplified version of the proposed amendment, while the remaining authors approved the simplified rewrite as semantically equivalent to the original.[6]

---

[3]This is an unavoidable technical expression, already present in four other articles of the original text (Article 56, Article 58, Article 122, and Article 126).

[4]The original Italian formulation is: "Entro dieci giorni dalla sua formazione il Governo si presenta alle Camere per ottenerne la fiducia. Nel caso in cui non sia approvata la mozione di fiducia al Governo presieduto dal Presidente eletto, il Presidente della Repubblica rinnova l'incarico al Presidente eletto di formare il Governo. Qualora anche in quest'ultimo caso il Governo non ottenga la fiducia delle Camere, il Presidente della Repubblica procede allo scioglimento delle Camere".

[5]"Dopo essere stato formato il Governo ha dieci giorni di tempo per presentarsi alle Camere e chiedere la loro fiducia. Se la fiducia non viene concessa il Presidente della Repubblica incarica di nuovo il Presidente eletto di formare il Governo. Se anche in questo secondo caso il Governo non ottiene la fiducia il Presidente della Repubblica scioglie le Camere".

[6]More specifically, the simplified texts were drafted by trying to keep them as structurally faithful as possible to the original. At the same time the simplified version was modified at different levels: from a *lexical* point of view, technical or potentially ambiguous terminology was replaced by lexemes belonging to the basic vocabulary of standard Italian; from a *morpho-syntactic* point of view, complex formulations were reworked through structurally more linear solutions and dislocation of components.

A different example is provided in the fifth paragraph of Art. 92, which states: "The President of the Republic entrusts the elected President of the Council of Ministers with the task of forming the Government; appoints and dismisses ministers upon *this* proposal." In this case, the implicitness resulting from the anaphoric use of "this" makes it grammatically acceptable for "this" to refer to the antecedent "Government" rather than to the intended referent, the "elected President of the Council of Ministers," which is presumably the substitution intended by the Legislator.[7]

The following paragraph consists of a single proposition (30 words) that includes two parenthetical statements, a relative clause, and two anaphoric references ('who', [orders] 'it'):

> In the event of the resignation of the elected President of the Council of Ministers, following parliamentary notification, he may request the dissolution of the Chambers to the President of the Republic, who orders it.

The content of the paragraph could be presented through a simpler utterance (24 words):

> In the event of resignation, the elected President of the Council of Ministers has seven days to ask the President of the Republic to dissolve the Chambers.

The failure to exercise the power to which the Legislator presumably refers is the proposal to dissolve the Chambers by the elected President of the Council of Ministers, and not the dissolution of the Chambers by the President of the Republic, as a literal interpretation of the text might suggest. This reading would even leave open the possibility that the President of the Republic might not accept the dissolution request made by the President of the Council of Ministers. The first sentence could be integrated and reformulated as follows, with the parenthetical statement moved forward to preserve the unity of the phrase "conferire l'incarico" (to assign the task):

> If the resigning President of the Council of Ministers does not exercise this power, the President of the Republic may, once during the legislative term, *assign* to him or another parliamentarian elected in connection with him *the task* of forming the Government.

The solutions to be provided in cases of "death, permanent disability, or removal from office," which present completely different scenarios and are incompatible with the prospect of a second mandate for the elected President of the Council of Ministers, should be separated and included in a separate paragraph. The phrase illustrating a permanent disability can be considered an example of under-determination [15], as it is subject to various plausible interpretations and therefore a potential source of disputes during the application phase.

From a lexical standpoint, the overall examination of the amendment texts –leaving apart semantic redefinitions and specialized locutions ('seduta comune', 'revoca della fiducia', 'informativa parlamentare': respectively "joint session", "revocation of trust", "information to lawmakers") that cannot be replaced in the legal field– reveals defects that could be avoided, such as the use of bureaucratic terms ('avere luogo' instead of 'svolgersi', both meaning 'to take place'; 'conferire' instead of 'assegnare' or 'attribuire', both meaning 'to give') and collocations that are not part of the standard Italian language, such as the transitive use of the verb 'importare' in the phrase 'importare obbligo' ('to involve an obligation') (Art. 94, paragraph 4: "Il voto contrario [...] non importa obbligo di dimissioni").

3.2. Readability Analysis based on the Prediction of Reading Times

3.2.1. Readability Indices

Readability indices are used to estimate the difficulty of reading a text [16]. These indices are calculated based on various linguistic elements, including the number of syllables, words, and sentences. The most popular readability index is the Flesch Reading Ease Score [17], which assigns a score between 0 (hardest) and 100 (easiest) based on the number of syllables per 100 words and the number of words per sentence. This reading ease score can be converted into a grade level, resulting in the Flesch-Kincaid

---

[7]We tried to preserve in the translation the ambiguity stemming from the original sentence "Il Presidente della Repubblica conferisce al Presidente del Consiglio eletto l'incarico di formare il Governo; nomina e revoca, su proposta *di questo*, i ministri", where the demonstrative pronoun "di questo" has been translated with "this".

Table 1

Indices assessing the readability of the considered articles of the Italian Constitution along with their proposed amendments. While both Flesch-Vacca and GulpEase are *readability* scores (such that a higher score is preferable, which is shown by the symbol ↑), the READ-IT score grasps the difficulty, so a lower score is preferable in this case (↓). By 'in force' we indicate an article currently in force, 'proposed reform' refers to an amended article according to the governative proposal; while 'simplified prop.ref.' indicates our simplification of the proposed reform.

Articles	Flesch-Vacca ↑	GulpEase ↑	READ-IT ↓
Art. 92 in force	43.32	49.26	0.10%
Art. 92 proposed reform	32.81	45.34	100.00%
Art. 94 in force	49.44	55.97	48.00%
Art. 94 proposed reform	35.62	47.36	95.10%
Art. 94 simplified prop.ref.	45.03	52.26	52.40%

Grade Level [18]. In 1972, Franchina and Vacca created an Italian adaptation of the Flesch Reading Ease Score [19]. Additionally, in 1986, another readability index was developed for the Italian language, the GulpEase Index [20]. Both indices follow a scoring scale similar to the Flesch Reading Ease Score, where higher scores indicate greater readability. More recently, a readability index specifically designed for text simplification has been devised: READ–IT [21]. This index combines traditional raw text features with lexical, morpho-syntactic and syntactic information, and allows computing the readability (namely, a difficulty) score for entire texts, and sentences therein.

The Vacca Index scores computed for the texts at hand show a decrease in readability for the proposed versions of the Constitution articles with respect to the in force versions. Specifically, the readability score for Article 92 drops from 43.32 (version currently in force) to 32.81 for the newly proposed version. Likewise, the readability score for the Article 94 decreased from 49.44 in the former version to 35.62, characterizing the amended version of the article.[8] The Vacca index scores computed for articles 92 and 94 are reported in Table 1. The GulpEase Index scores are consistent with the Vacca Index, showing a significant reduction in readability, for both Article 92 and 94: the scores of the former drop from 49.26 in the version in force to 45.34 in the amended version, and from 55.97 to 47.36 in the amended version for the latter article.[9] Moreover, the simplified version of Article 94 introduced in Section 3.1 shows an increase in readability compared to the new version on both the Vacca Index and GulpEase, with values of 45.03 and 52.26, respectively. In Table 1 we also provide the READ-IT difficulty scores that, similar to previous indices, show that the proposed amended texts contain elements worsening the overall readability. Finally, we note that the proposed simplified text consistently received more favorable scores, showing that the readability associated to the proposed amendments can be substantially improved.

Figure 1 illustrates the Flesch-Vacca and GulpEase scores for all articles in the Constitution, along with the average score for each index. The plotted points collectively describe the readability of the whole Constitution, and provide a context to the scores computed for Article 92 –for which we propose a comparison between the in-force and proposed version– and Article 94, for which we additionally report the readability score of our simplified text. For the Vacca Index, both the in force and proposed versions of articles 92 and 94 are above the average value. Conversely, for the GulpEase Index, the in force versions are above the average, while the proposed versions fall below it, thereby resulting in a reduced readability.

3.2.2. Reading Times as a Proxy for Readability

To date, eye tracking systems allow to collect precise data in form of timestamped fixations that describe and to a good extent allow to reconstruct readers' behavior and difficulty throughout the reading task.

---

[8]According to the Vacca Index, text associated with scores between 30-50 is understandable for university students, while scores between 50-60 characterize text suited for high school students.

[9]According to the GulpEase Index, texts with scores below 40 are difficult to understand for high school graduates, while those with scores below 60 are difficult for people with a middle school diploma.
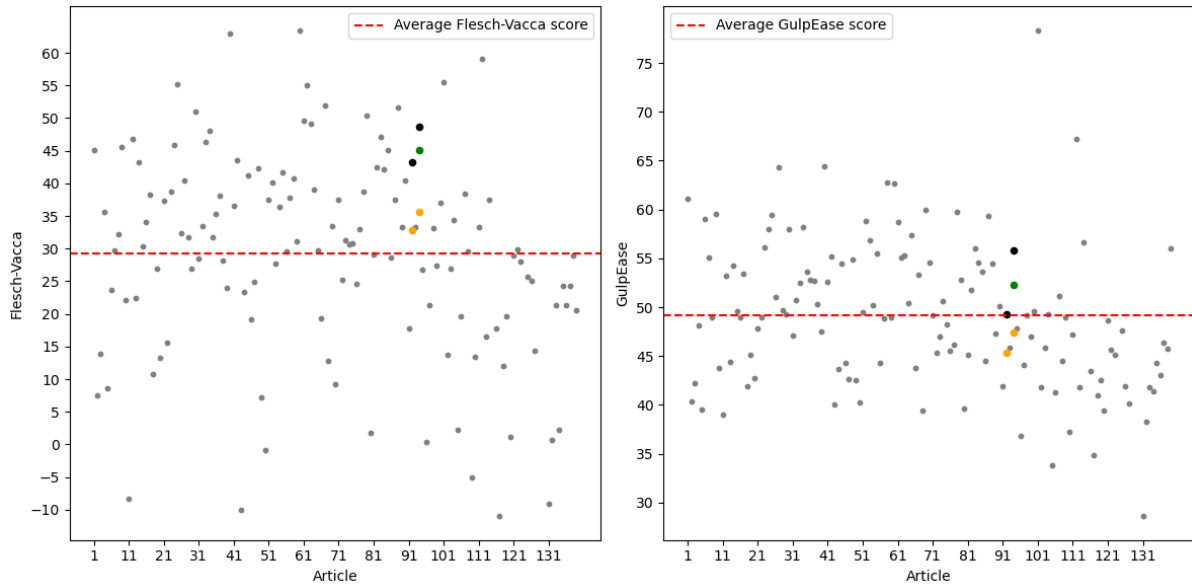
Figure 1: Flesch-Vacca Index scores (on the left) and GulpEase Index scores (on the right) for all the 139 articles in the Constitution. Regards as the Articles 92 and 94, points marked in orange report the values obtained for the proposed modifications, while scores for the in-force versions are marked in black. Green points report the values associated to the simplified version (only available for the Article 94).

On the other side, the refinement and spread of language models allows to automatically perform subtle forms of linguistic analysis, such as determining the semantic coherence between a term and its surrounding context, thereby determining the predictability of words given their preceding context.

To give some background on how eye-tracking works, two main eye movements are commonly individuated throughout the reading task, *fixations* and *saccades*. Fixations are brief stops (with duration ranging from 50 to 1500 ms) that typically occur at each word; sometimes more stops are needed, depending on words length and difficulty. Saccades are fast (ranging from 10 to 100 ms) movements between each two fixations, used in repositioning the point of focus. Based on these underpinnings we explored a novel approach to assess text documents readability: in essence, this approach relies on the following intuition. Reading times can be employed as a proxy for different significant stages in linguistic processing. In particular, the total reading time (TRT) —the overall duration of eye fixations for each word, including the backward regression movements— is largely acknowledged to grasp the time taken by the overall semantic integration [22]. Moreover, two partial and finer-grained measures have been also proposed: the duration of the first fixation (FFD) that allows estimating the effort underlying lexical access [23], and the number of fixations (NF), that is typically associated to the integration of words in the frame of what has been read so far [24]. In this setting higher reading times are a function of a higher cognitive load, and report about less readable text excerpt.

A model was trained and tested on eye-tracking data collected from 60 subjects reading a Regional Regulation from the Aosta Valley; it basically relies on a LightGBM regressor that incorporates word-related statistics known to influence sentence and word processing (such as word frequency, word length, word position within the sentence, previous word frequency, and previous word length). This model was also refined through surprisal scores, computed based on a fine-tuned version of an Italian GPT-2 model [25]. This fine-tuning step was performed by exposing the language model to 2,950 Regional laws and 131 Regional regulations from the Aosta Valley Region. The LightGBM regressor is based on the gradient boosting framework, an ensemble learning technique that utilizes a pool of weak learners (decision trees), and its algorithm is featured by a leaf-wise tree growth strategy, implying that the algorithm grows the tree by expanding the leaf with the maximum delta loss instead of growing it level by level, in depth-wise fashion, to find optimal split points more quickly. A comprehensive search for optimal hyperparameters was performed using a grid search technique. The hyperparameters

considered for optimization are: the maximum number of leaves in each tree; the learning rate; the number of estimators (trees) to be built to tune the balance between under- and over-fitting; the maximum depth of each tree. The optimization process targeted the mean absolute error (MAE). The evaluation of different parameter combinations was performed through a 5-fold cross-validation strategy during the grid search.

From a methodological standpoint we are of course aware of the differences between the text properties of the Italian Constitution compared to a Regional Regulation,[10] but since there are no available datasets that include Eye-Tracking data associated with the reading of the Constitution, we resorted to data originally conceived to predict the reading times associated to Regional norms from the Aosta Valley [26]. The adopted model implements an approach successfully employed for the CMCL 2021 Shared Task on Eye-Tracking Prediction [27, 28].

As mentioned, a key element in our model is the adoption of surprisal scores. We briefly recall this feature, which is illustrated in more detail in [26]. Further details on the application of the closely related metrics of perplexity may be found in [29, 30, 31]. In the last few years neural language models gained a central role in analyzing reading as well, since they are able to acquire conditional probability distributions over the lexicon that are to a good extent predictive of human processing times. Probabilistic language modeling, as a device to describe the incremental mechanisms underlying language processing, is acknowledged as helpful to account (at a high level) for basic cognitive strategies [32, 33]. Such strategies are primarily concerned with planning and handling expectations on what follows, and on evaluating how these match with actual stimuli [34]. One chief assumption is that words predictability should be intended as a function of the probability of a word given the context, and the probability of that word may work, in turn, as a main predictor of reading times [35]. In essence, the less likely the emission of a word, the higher the *surprisal* associated to that word, and —what counts more for our present concerns— the longer the time it requires for readers to process it,

$$\text{effort}(t) \propto \text{surprisal}(w_t) = -\log(P(w_t|w_1, \ldots, w_{t-1})).$$

Surprisal scores were thus plugged into our model to support the prediction of reading times by also accounting for the difficulty of predicting words.

Results The average predicted reading times for the articles, measured in milliseconds, do not show significant differences. Narrowing the analysis to tokens without stop words, the amended version of Article 92 exhibits slightly shorter reading times, while in Article 94 the proposed amendments result in longer reading times. In Table 2 ('average' section: top of Table) we report the average total reading times (TRTs) associated to tokens in the original and in the amended version of the Article 92, as well as in the original, amended, and simplified version of the Article 94. In the 'sum' section (at the bottom of the same Table) we display the sum of TRTs predicted for tokens in the articles. Different from the above average, this measure is no longer averaged over all tokens in the text, and thus reflects in closer fashion the increased difficulty stemming from lengthy text sequences. As expected, the reading times predicted for the simplified version of Article 94 are slightly lower compared to the proposed reform, both considering the whole text, and by filtering stop words.

In Figure 2 we provide a comparison of the predicted TRTs for the articles 92 and 94 against the rest of the Italian Constitution: in particular we report the average predicted TRTs and the sum of predicted TRTs for the tokens in each article of the Constitution. Articles 92 and 94 (both the in force and proposed versions) show lower average TRT compared to the average over of the entire Constitution. Additionally, the sum of TRTs (reported on the right sub-figure in Figure 2) indicates that, while the in force versions of these two articles align with the rest of the Constitution, the proposed versions are significantly longer, among the longest articles in the Constitution.

---

[10]The language of the Constitution tends to be more formal, precise, and abstract, and operationalizes broad principles and fundamental rights such as 'freedom' and 'democracy'. Regional Legislation, on the other side is more specific and practical, dealing with concrete issues and administrative matters that are the fields in which regions produce norms and regulations. These may use more technical jargon relevant to specific sectors like health, environment, transportation, or education.

Table 2
Predicted reading times for the Articles 92 and 94 ('in force'), and their amended text ('proposed reform' rows); we also report the TRT predicted for the simplified version of the Art. 94, 'simplified prop.ref.'. TRTs associated to the whole text, with no filtering, and to the text after stopwords filtering are reported. In the top subtable ('average') the average TRTs (complemented by their standard deviations) are reported; the bottom subtable illustrates figures computed as the sum of all TRTs predicted for the tokens in the considered article.

average	TRT - whole text	TRT - filtered text
92 in force	260.01 (146.8)	395.14 (45.09)
92 proposed reform	258.32 (156.26)	381.22 (102.75)
94 in force	257.24 (133.03)	351.14 (71.85)
94 proposed reform	261.63 (141.1)	360.19 (87.91)
94 simplified prop.ref.	254.89 (141.37)	357.44 (82.82)

sum	TRT - whole text	TRT - filtered text
92 in force	$9,620.55$	$7,122.48$
92 proposed reform	$41,589.73$	$32,022.29$
94 in force	$22,380.21$	$18,259.32$
94 proposed reform	$62,528.91$	$50,427.21$
94 simplified prop.ref.	$55,566.57$	$44,322.79$

Let us inspect more closely the TRTs predicted for Article 94 at the sentence level, reported in Table 3. These are also computed as averaged figures and as the sum of TRTs of the tokens in the considered sentence, as formerly described. All newly proposed sentences (namely sentences number 4, 5, 8, 9, and 10) are featured by TRTs higher than the mean TRTs of the version presently in force of the Article 94, that is 257.24 ms (please refer to Table 2). The Pearson correlation index between the sum of TRTs predicted for the three versions of Article 94 and the READ-IT scores amounts to 0.69 ($p < 0.2$), 0.60 ($p < 0.068$), and 0.62 ($p < 0.057$) for the current version, its proposed amendment, and the simplified rewrite for the amendment, respectively. This datum shows that reading times predictions are able to capture what has been described at the linguistic and legal level, and through the indices surveyed in Table 1.

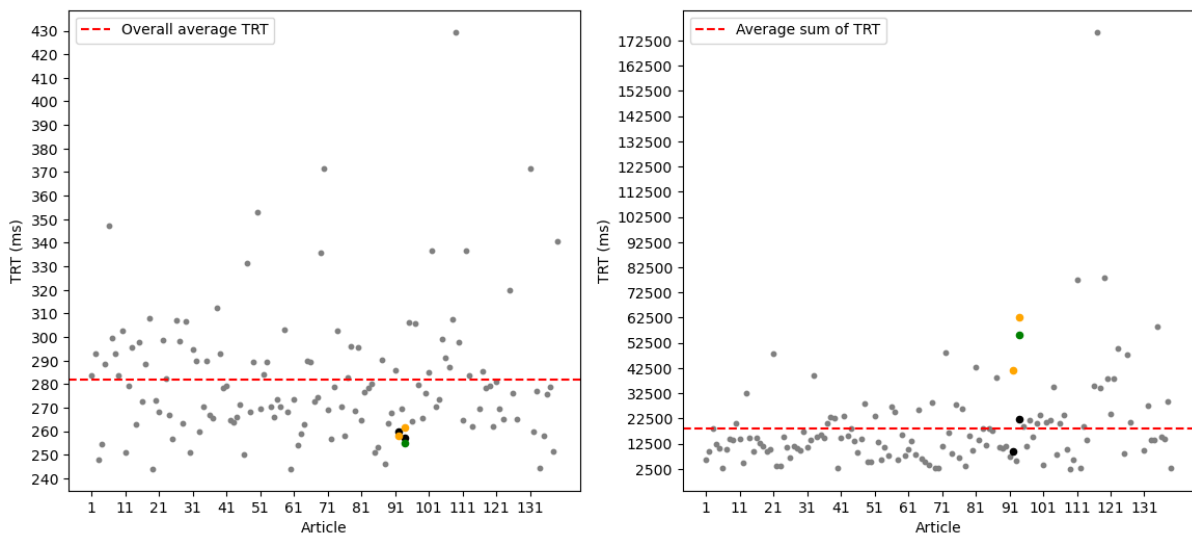To refine our analysis at the word level, we manually annotated the words identified as difficult/poorly

Figure 2: Average total reading times (on the left) and sum of total reading times (on the right) for all the 139 articles in the Constitution. Regards as the Articles 92 and 94, points marked in orange report the values obtained for the proposed modifications, while scores for the in-force versions are marked in black. Green points report the values associated to the simplified version (only available for the Article 94).

Table 3

Predicted total reading times (TRTs) for the Article 94. Reported results refer to values computed at the sentence level for the original formulation, the proposed modifications, and the TRTs predicted for the simplified version: 'in force', 'proposed reform' rows, and 'simplified prop.ref.', respectively. Dashes in the 'in force' column indicate that that specific sentence is not currently present in the Italian Constitution (that is, a new sentence was added in the proposed reform); dashes in the rightmost columns indicate that no TRT was computed for the text of the corresponding sentence, since it was left unaltered in the proposed reform.

Art. 94 Sent. num.	in force average (sum)	proposed reform average (sum)	simplified prop.ref. average (sum)
1	274.03 (2, 192.22)	–	–
2	309.75 (4, 646.2)	–	–
3	259.95 (4, 159.26)	–	–
4	–	257.97 (7, 739.01)	258.06 (5, 161.16)
5	–	271.39 (5, 970.62)	243.78 (4, 631.85)
6	245.29 (4, 660.5)	–	–
7	231.79 (6, 722.03)	–	–
8	–	268.35 (5, 635.35)	294.34 (6, 769.90)
9	–	270.63 (7, 848.38)	242.82 (5, 827.80)
10	–	257.82 (12, 890.76)	244.36 (10, 018.65)

readable or demanding in the proposed reform sentences and calculated their average predicted TRTs.[11] Tokens from the Article 92 and annotated as difficult are associated to predicted TRT that on average reach 412.15 ms (235.25 for tokens not marked as difficult), while for the Article 94 we found that the average over TRTs is 391.80 (and 240.77). Moreover, we partitioned the words in the amended versions into two groups: those whose predicted TRT is above the mean (258.32 ms for Article 92, and 261.32 for Article 94) and those below the mean. We then calculated the correlation with the manually annotated words, finding a Pearson correlation of 0.27 (Article 92) and 0.30 (Article 94) with significance of $p < 0.0006$ and $p < 0.000003$ respectively. Both trials reveal a reasonable fit between the human annotation and the predicted TRTs, thus corroborating the proposed approach as consistent with human annotation. Also, our initial assumption that longer reading times report about an augmented cognitive load seems to be confirmed, based on both correlation with human introspection on what counts as readable (or not), and with more traditional indices.

3.3. GPT-based Analysis

We prompted Chat-GPT4o to analyze and compare the readability of articles in force and their proposed amendment. Specifically, the following prompt was used (here reported in English, but originally used in Italian):[12]

> *I will send you two versions of an article from the Constitution. I would like to know if one is more complicated than the other or not, adding which parts you find more complex and why. Focus on linguistic complexity and not on meaning. Present your comments point by point.*
>
> *Version 1: <article text>*
> *Version 2: <article text>*

Regarding Article 92, GPT answers are in line with our our linguistic analysis. GPT remarks that the proposed amendment has longer and more intricate sentences, packed with information, redundancies and technical language which make it difficult to follow. These evaluations are consistent with the results obtained by applying the Flesch-Vacca, GulpEase and READ-IT indices to the sentences in (old and renewed version of) the Article 92. Such scores are reported in Table 4: we note that the amended

---

[11]A single annotation was collected, performed by an expert linguist.

[12]The transcript of the full interaction with GPT is available at the URL https://github.com/mensae/costituzione-analysis/.

Table 4
Comparison of the Flesch-Vacca, GulpEase, and READ-IT indices, along with the sum of TRTs for the six sentences from the Article 92. Reported figures characterize the version currently in force (only containing two sentences, 1 and 6) and the proposed amendments, where the first sentence is kept unaltered, sentences from 2 to 5 are newly introduced, and sentence 6 modifies the second sentence in the original formulation. The symbol ↑ indicates that a higher score is preferable, while ↓ indicates that a lower score is preferable.

Sent. num	Flesch-Vacca ↑		GulpEase ↑		READ-IT ↓		TRT	
	in force	proposed	in force	proposed	in force	proposed	in force	proposed
1	39.75	39.75	46.5	46.5	77.90%	77.90%	5,094.65	5,094.65
2	–	30.09	–	47.81	–	47.8%	–	10,111.83
3	–	53	–	56.89	–	0.30%	–	3,240.81
4	–	6.35	–	39.93	–	100.00%	–	13,033.98
5	–	52.33	–	57.67	–	98.30%	–	3,510.4
6	53.58	35,4	54.26	47.4	3.20%	97.40%	4,525.89	6,598.06

sentence number 6, corresponding to sentence 2 in the version currently in force, received reduced readability scores with respect to its former wording. We observe that the newly added paragraphs in the proposedly amended version are characterized by reduced readability –or, equivalently, increased reading times–, on average, over all considered indices.

Similarly, for the Article 94 we note that the use of domain-specific language, difficult for non-expert readers, is highlighted for the sentences in the version currently in force, as well. Also in this case the comments collected by prompting GPT, about the overall amount of information, length and complexity of the sentences in the proposed paragraphs treat all such factors as possibly confusing and badly affecting the readability of the text. These notes are corroborated by the indices reported in Tables 1, 2, and 3.

Finally, also consistently with previous results, GPT highlighted how our simplified version for the proposed amendments makes use of more direct and common language, employs shorter sentences, fewer subordinate clauses and incidental propositions, and exhibits less redundancy and repetitions.

4. Conclusions

Recently, there has been a growing interest in the Italian legal field among researchers in computational linguistics, as demonstrated by the works in [36] and [37]. Our work contributes to this field from a different angle. We presented a multi-layered analysis of the Constitutional Reform Bill. This analysis integrates legal and linguistic perspectives, traditional readability indices, and a novel approach employing predictive methods for reading times.

From the viewpoint of linguistic and legal experts, the proposed amendments exhibit defects in terms of the quality of the writing: numerous issues arise from both syntactic and semantic perspectives. These aspects contribute to a text that is not only challenging to follow and comprehend compared to the original, but also susceptible to multiple plausible interpretations, potentially becoming a source of disputes. These observations are substantiated through readability analyses, which also confirm improved readability in our simplified version of the amendment. Furthermore, correlations derived from our analysis using TRTs and READ-IT indices show that predictions of reading times effectively capture the linguistic and legal complexities described. Similar to READ-IT, our method may be computed at the sentence level. In addition, our system may be trained with eye-tracking data from different sorts of reader, and the employed language model may be fine-tuned on various kinds of text, thus targeting a more flexible notion of readability, associated to a specific group of readers and to a specific kind of text. More specifically, provided that eye-tracking data on texts as close as possible to those of interest are needed, the proposed analytical approach may be easily extended to novel domains and different sorts of text —such as, e.g., textbooks, newspapers, code, assembly instructions for general manufactured items, and so forth—. Likewise, since our approach does not depend on specific (possibly arbitrary) parameters, it may be employed to predict reading times of specific target groups such as,

e.g., children, domain experts, laypeople in domain-specific settings. Along this axis, again, it would be necessary to record reading times from such profiled reader groups.

Finally, a brief examination using a modern Large Language Model, GPT-4o, aligns with previous findings, identifying complexities in the proposed amendments, highlighting both general problems like text length and intricate text structure (e.g., due to overuse of subordinate clauses) and adoption of specialist jargon, in accordance with the analysis offered by jurists and linguists. These outcomes indicate that future research can significantly benefit from the automated and combined use of LLMs alongside more specialized tools to identify critical components within texts.

References

[1] G. Bonetta, C. D. Hromei, L. Siciliani, M. A. Stranisci, Preface to the Eighth Workshop on Natural Language for Artificial Intelligence (NL4AI), in: Proceedings of the Eighth Workshop on Natural Language for Artificial Intelligence (NL4AI 2024) co-located with 23th International Conference of the Italian Association for Artificial Intelligence (AI*IA 2024), 2024.

[2] G. Busia, Il percorso di elaborazione del testo costituzionale, Bologna: Il Mulino, 1998, pp. 129–164.

[3] OpenAI, J. Achiam, S. A. et al., Gpt-4 technical report, 2024. URL: https://arxiv.org/abs/2303.08774. arXiv:2303.08774.

[4] OpenAI, Hello GPT4o Web page, 2024. URL: https://openai.com/index/hello-gpt-4o/.

[5] T. De Mauro, Storia linguistica dell'Italia unita, Biblioteca di cultura moderna, Laterza, 1963. URL: https://books.google.it/books?id=1l0mAAAAMAAJ.

[6] B. M. Garavelli, L'italiano della repubblica: caratteri linguistici della costituzione, in: V. Coletti (Ed.), L'italiano dalla nazione allo Stato, Le Lettere, Firenze, 2011, p. 211.

[7] L. Cignetti, Sfondi e rilievi testuali nella costituzione della repubblica italiana, in: Le (Ed.), Rilievi, 2005.

[8] M. A. Cortelazzo, Un elogio linguistico [della costituzione], LID'O. Lingua italiana d'oggi VI (2009) 43–52.

[9] T. De Mauro, Il linguaggio della costituzione, Lid'O: lingua italiana d'oggi: VI, 2009 (2009) 31–42.

[10] E. Corino, La costituzione italiana è ancora un testo facile?, in: A. Ferrari, L. Lala, F. Pecorari (Eds.), L'italiano dei testi costituzionali, Edizioni dell'Orso, Alessandria, 2022, pp. 293–318.

[11] E. Leso, 27 dicembre 1947: Lingua della costituzione e lingua di tutti, in: F. Bambi (Ed.), Un secolo per la Costituzione (1848-1948). Concetti e parole nello svolgersi del lessico costituzionale italiano, Accademia della Crusca, Firenze, 2012, pp. 277–290.

[12] G. Rovere, Annotazioni metodologiche sulla comprensibilità del lessico costituzionale italiano, in: A. Ferrari, L. Lala, F. Pecorari (Eds.), L'italiano dei testi costituzionali, Edizioni dell'Orso, Alessandria, 2022, pp. 271–292.

[13] M. E. Piemontese, M. Piemontese, et al., Capire e farsi capire. teorie e tecniche della scrittura controllata (1996).

[14] Accademia della Crusca in collaborazione con il CLIEO e l'ITTIG, Guida alla redazione degli atti amministrativi, CLIEO e ITTIG, Firenze, 2011. URL: https://www.ittig.cnr.it/Ricerca/Testi/GuidaAttiAmministrativi.pdf, documento online.

[15] L. Revelli, A. Mastropaolo, R. D. Paolo, et al., La sottodeterminazione nei testi giuridici: verso un'analisi linguistico-computazionale, in: Fare linguistica applicata con le digital humanities, volume 14, Officinaventuno, 2022, pp. 131–144.

[16] A. Siddharthan, Syntactic simplification and text cohesion, Research on Language and Computation 4 (2006) 77–109. URL: http://dx.doi.org/10.1007/s11168-006-9011-1. doi:10.1007/s11168-006-9011-1.

[17] R. F. Flesch, A new readability yardstick., The Journal of applied psychology 32 3 (1948) 221–33. URL: https://api.semanticscholar.org/CorpusID:39344661.

[18] R. Flesch, Marks of readable style; a study in adult education., Teachers College Contributions to Education (1943).

[19] V. Franchina, R. Vacca, Adaptation of Flesh readability index on a bilingual text written by the same author both in Italian and English languages, Linguaggi 3 (1986) 47–49.

[20] P. Lucisano, M. E. Piemontese, et al., GulpEase: una formula per la predizione della leggibilità di testi in lingua italiana, Scuola e città (1988) 110–124.

[21] F. Dell'Orletta, S. Montemagni, G. Venturi, READ–IT: Assessing readability of Italian texts with a view to text simplification, in: Proceedings of the second workshop on speech and language processing for assistive technologies, 2011, pp. 73–83.

[22] R. Radach, A. Kennedy, Eye movements in reading: Some theoretical context, The Quarterly journal of experimental psychology 66 (2013) 429–452.

[23] M. J. Hofmann, S. Remus, C. Biemann, R. Radach, L. Kuchinke, Language models explain word reading times better than empirical predictability, Frontiers in Artificial Intelligence 4 (2022) 730570.

[24] L. Frazier, K. Rayner, Making and correcting errors during sentence comprehension: Eye movements in the analysis of structurally ambiguous sentences, Cognitive psychology 14 (1982) 178–210.

[25] W. de Vries, M. Nissim, As Good as New. How to Successfully Recycle English GPT-2 to Make Models for Other Languages, in: Findings of the Association for Computational Linguistics: ACL-IJCNLP 2021, Association for Computational Linguistics, 2021. URL: http://dx.doi.org/10.18653/v1/2021.findings-acl.74. doi:10.18653/v1/2021.findings-acl.74.

[26] C. J. Scozzaro, D. Colla, M. Delsanto, A. Mastropaolo, E. Mensa, L. Revelli, D. P. Radicioni, et al., Legal text reader profiling: Evidences from eye tracking and surprisal based analysis, in: Proceedings of the Workshop on DeTermIt! Evaluating Text Difficulty in a Multilingual Context@ LREC-COLING 2024, ELRA and ICCL, 2024, pp. 114–124.

[27] N. Hollenstein, E. Chersoni, C. L. Jacobs, Y. Oseki, L. Prévot, E. Santus, CMCL 2021 shared task on eye-tracking prediction, in: E. Chersoni, N. Hollenstein, C. Jacobs, Y. Oseki, L. Prévot, E. Santus (Eds.), Proceedings of the Workshop on Cognitive Modeling and Computational Linguistics, Association for Computational Linguistics, Online, 2021, pp. 72–78. URL: https://aclanthology.org/2021.cmcl-1.7. doi:10.18653/v1/2021.cmcl-1.7.

[28] Y. Bestgen, LAST at CMCL 2021 shared task: Predicting gaze data during reading with a gradient boosting decision tree approach, in: E. Chersoni, N. Hollenstein, C. Jacobs, Y. Oseki, L. Prévot, E. Santus (Eds.), Proceedings of the Workshop on Cognitive Modeling and Computational Linguistics, Association for Computational Linguistics, Online, 2021, pp. 90–96. URL: https://aclanthology.org/2021.cmcl-1.10. doi:10.18653/v1/2021.cmcl-1.10.

[29] D. Colla, M. Delsanto, M. Agosto, B. Vitiello, D. P. Radicioni, Semantic coherence markers: The contribution of perplexity metrics, Artificial Intelligence in Medicine 134 (2022) 102393.

[30] D. Colla, M. Delsanto, D. P. Radicioni, Semantic coherence dataset: Speech transcripts, Data in Brief 46 (2023) 108799.

[31] F. Sigona, D. P. Radicioni, B. G. Fivela, D. Colla, M. Delsanto, E. Mensa, A. Bolioli, P. Vigorelli, A computational analysis of transcribed speech of people living with dementia: The anchise 2022 corpus, Computer Speech & Language 89 (2025) 101691.

[32] E. G. Wilcox, J. Gauthier, J. Hu, P. Qian, R. Levy, On the predictive power of neural language models for human real-time comprehension behavior., CoRR abs/2006.01912 (2020). URL: http://dblp.uni-trier.de/db/journals/corr/corr2006.html#abs-2006-01912.

[33] E. G. Wilcox, T. Pimentel, C. Meister, R. Cotterell, R. P. Levy, Testing the predictions of surprisal theory in 11 languages, Transactions of the Association for Computational Linguistics 11 (2023) 1451–1470.

[34] R. Levy, Expectation-based syntactic comprehension, Cognition 106 (2008) 1126–1177.

[35] I. F. Monsalve, S. L. Frank, G. Vigliocco, Lexical surprisal as a general predictor of reading time, in: Proceedings of the 13th Conference of the European Chapter of the Association for Computational Linguistics, 2012, pp. 398–408.

[36] D. Licari, G. Comandè, ITALIAN-LEGAL-BERT: A Pre-trained Transformer Language Model for Italian Law, in: D. Symeonidou, R. Yu, D. Ceolin, M. Poveda-Villalón, D. Audrito, L. D. Caro,

F. Grasso, R. Nai, E. Sulis, F. J. Ekaputra, O. Kutz, N. Troquard (Eds.), Companion Proceedings of the 23rd International Conference on Knowledge Engineering and Knowledge Management, volume 3256 of *CEUR Workshop Proceedings*, CEUR, Bozen-Bolzano, Italy, 2022. URL: https://ceur-ws.org/Vol-3256/#km4law3, iSSN: 1613-0073.

[37] S. Auriemma, M. Madeddu, M. Miliani, A. Bondielli, L. C. Passaro, A. Lenci, BureauBERTo: adapting UmBERTo to the Italian bureaucratic language, in: F. Falchi, F. Giannotti, A. Monreale, C. Boldrini, S. Rinzivillo, S. Colantonio (Eds.), Proceedings of the Italia Intelligenza Artificiale - Thematic Workshops co-located with the 3rd CINI National Lab AIIS Conference on Artificial Intelligence (Ital IA 2023), volume 3486 of *CEUR Workshop Proceedings*, CEUR-WS.org, Pisa, Italy, 2023, pp. 240–248. URL: https://ceur-ws.org/Vol-3486/42.pdf.