Applying Information-theoretic Notions to Measure Effects of the Plain English Movement on English Law Reports and Scientific Articles

Sergei Bagdasarov
Language Science and Technology
Saarland University
sergeiba@lst.uni-saarland.de

Stefania Degaetano-Ortlieb
Language Science and Technology
Saarland University
s.degaetano@mx.uni-saarland.de

Abstract

We investigate the impact of the Plain English Movement (PEM) on the complexity of legal language in UK law reports from the 1950s-2010s, contrasting it with the evolution of scientific language. The PEM, emerging in the late 20th century, advocated for clear and understandable legal language. We define complexity through the concept of surprisal – an information-theoretic measure correlating with cognitive processing difficulty. Our research contrasts surprisal with traditional readability measures, which often overlook content. We hypothesize that, if the PEM has influenced legal language, there would be a reduction in complexity over time and a shift from a nominal to a more verbal style. We analyze text complexity and lexico-grammatical changes in line with PEM recommendations. Results indicate minimal impact of the PEM on both legal and scientific domains. This finding suggests future research should consider processing effort when advocating for linguistic norms to enhance accessibility.

1 Introduction

Legal language has been notorious for its intricate syntax and specialized jargon, making it challenging for non-experts to comprehend. This complexity has not only been a barrier to understanding for the general public but has also posed significant challenges for computational analysis. In response to this, the Plain English Movement (PEM) emerged as a pivotal initiative in the latter half of the 20th century, advocating for clear, concise, and understandable legal language (cf. Mazur (2000)).

Our study is corpus-based and explores whether this movement has led to a measurable change in the complexity of legal language, specifically UK law reports, from the 1950s to the 2010s. We contrast this with scientific language, which we hypothesize has not been similarly influenced by the PEM. By comparing the evolution of language in these two domains, we work towards uncovering the unique trajectories of language complexity in response to professional and societal pressures.

We define complexity based on the concept of surprisal – an information-theoretic measure of unpredictability in language, which has shown to be proportional to cognitive effort and thus processing difficulty (Hale, 2001; Levy, 2008). Thus, high surprisal in a text, indicative of less predictable content, is associated with increased cognitive processing effort and serves as a reliable indicator of increased language complexity (Smith and Levy, 2013). We contrast surprisal with readability measures usually used to measure language complexity but often criticized for not taking into account the content of the text being evaluated (cf. Schriver (1997); Mazur (2000)). In fact, while recently, in the computational linguistic community text simplification systems have been applied to simplify legal language (Garimella et al., 2022), many open questions remain, for example, which information should be retained. Here, a measure of informativity could enhance approaches on the matter.

We introduce the PEM and work on the diachronic tendency of phrasal compression as a densification strategy in specialized discourse, which is an opposite trend to what the PEM advocates. We present our rationale putting forward two main hypotheses regarding law reports: If the PEM has an impact, then we assume (H1) reduced complexity over time and (H2) change from a heavy nominal towards a verbal involved style. For scientific articles, we hypothesize no PEM impact. We then present the corpora used and the methodology applied to measure text complexity and to analyze lexico-grammatical changes related to the suggestions of the PEM. Finally, our result section dives into a macro-analytical diachronic perspective and a micro-analysis of the linguistic features typical of contemporary law reports and scientific texts. Results show that the PEM had little to no impact on

Proceedings of LaTeCH-CLfL 2024, pages 101–110

language use in these two domains. From this, we derive implications for future research that should account for processing effort when implementing linguistic norms towards increased accessibility.

2 Background and Rationale

2.1 Plain English Movement

The Plain English Movement (PEM) originated in the second half of the 20th century in response to the writing style of legal documents, incomprehensible for a general audience. While there had been some critical voices already in the 1960s, it was not until the 1970s when the PEM began to gain momentum and resulted in the adoption of the first governmental regulations, mainly in the U.S., imposing the use of a more reader-friendly language (Mazur, 2000; Williams, 2005). In subsequent decades, many other laws, regulations and initiatives followed, which testifies to a broad acceptance of the PEM in the legal community.

Although the PEM is mainly concerned with legal documents, there have also been important efforts to influence the writing of scientific articles, another domain that can be hardly understood by a lay audience. Plain language can be highly advantageous to the scientific community, with its benefits ranging from the general popularization of scientific research to the better ability to obtain funding. Despite the lack of compelling legislation on this matter, many important journals encourage researchers to use plain language in their papers (Locke, 2003; Sedgwick et al., 2021).

In the last decades, many style guides and guidelines have appeared that present the main principles of plain writing (Garner, 2001; European Commission, 2016; Federal Government of the United States, 2011). These principles are mainly driven by reducing processing cost and can be summarized into the following suggestions:

S1: Use short sentences. As processing cost is proportional to sentence length, the PEM recommends a max. of 15-20 words per sentence on average.

S2: Use 1st and 2nd person pronouns for a more personal connection with the reader.

S3: Avoid nominalizations and use verbal style instead (e.g. *apply* instead of *submit an application*) to promote a verbal style of writing that should enhance clarity and reduce sentence length.

S4: Avoid compounds as they leave implicit the semantic relations between nouns.

S5: Avoid unnecessary jargon and terminology where these can be replaced with general language without semantic loss.

S6: Avoid unnecessary abbreviations for the sake of clarity.

S7: Use active voice. Active voice allows shorter and generally easier-to-process sentences.

S8: Avoid *shall* because of its semantic ambiguity resulting from a generalized overuse in legal texts.[1]

We will investigate, whether these recommendations are somehow reflected as possible tendencies over time for legal and scientific texts.

2.2 Phrasal compression in specialized discourse

Both law reports and scientific articles are considered to be rather complex registers that can be hardly understood by non-experts. The most distinctive feature of both is a style of writing that favors nominal phrases – a preference illustrated in a number of synchronic studies (Breeze, 2019; Gotti, 2012).

Diachronically, the shift towards increasing phrasal complexity has been especially notorious in the case of scientific articles. While they used to rely on a more verbal style of writing with long subordinate clauses in the 17th and 18th centuries, the 19th century saw a sharp increase of prepositional phrases functioning as postmodifiers at the cost of clausal elements (Biber and Gray, 2016; Degaetano-Ortlieb and Teich, 2019). The shift towards a major phrasal complexity consolidated in the 20th century, with compound nouns adopting an increasingly important role in scientific articles (Biber and Gray, 2016; Degaetano-Ortlieb, 2021).

Although there was no such a dramatic transformation in the history of law reports, this register did evolve to include more nominal elements in the last 300 years, with its most prominent features being prepositional postmodifiers, compounds (albeit to a lesser extent than in scientific articles), and nominalizations (Biber and Gray, 2019).

2.3 Rationale

Considering the PEM suggestions, we put forward the following hypotheses: (H1) *Reduced complexity:* If the PEM suggestions influenced language use in these two domains, results should show shorter sentences and a lower degree of complexity

[1] Consider the following example: *The applicant shall be notified by registered mail in all cases where ...* (Federal Government of the United States, 2011). Here, ***shall*** can denote an obligation or be just describing a future action.

Decade	CoCELD		RSC	
	Number of Tokens	Number of Texts	Number of Tokens	Number of Texts
1950s	101,770	40	23,760,143	3,656
1960s	102,093	40	28.695,408	4,168
1970s	101,621	40	40,611,994	5,231
1980s	101,707	40	44,035,328	5,488
1990s	102,083	40	34,915,666	4,925
2000s	101,629	40	-	-
2010s	122,324	48	-	-

Table 1: Number of tokens and texts in CoCELD and RSC

over time, which we aim to capture by readability formulas and surprisal. (H2) *Nominal vs. involved verbal style:* If suggestions S2-S8 have an impact, distinctive features of more contemporary periods would be 1st and 2nd person pronouns and verbal style, while distinctive of earlier periods would be nominalizations and a heavy nominal style with abbreviations as well as the use of *shall*. In general, we assume the impact of the PEM to be more pronounced for law reports than for scientific articles.

3 Data and Methods

3.1 Corpora

For law reports, the Corpus of Contemporary English Legal Decisions (CoCELD) is used (Rodríguez-Puente and Hernández-Coalla, 2023). It contains legal decisions produced by the Privy Council, the House of Lords and the UK Supreme Court between 1950 and 2021. For research articles, we used the 6.0 version of the Royal Society Corpus (RSC), comprising the Proceedings and Transactions of the Royal Society of London (Kermes et al., 2016; Fischer et al., 2020). We selected a subcorpus from the RSC including texts from 1950 to 1996, which partly corresponds with the time span of CoCELD. The distribution of texts and tokens across decades in both corpora is summarized in Table 1.

Both corpora were annotated with TreeTagger using the Penn Treebank Tagset (Schmid, 1995)[2]. The corpora feature metadata (publication date; for RSC also authors, titles, journal series, etc), linguistic annotation (word, lemma and part of speech), and surprisal annotation (see Section 3.2.1).

3.2 Methods

We analyze the possible impact of the PEM by considering (a) text complexity using readability measures and surprisal to address H1 (reduced complexity), and (b) changes in the use of lexicogrammatical features to address H2 (nominal vs. involved verbal style).

3.2.1 Measuring text complexity

For text complexity, three metrics are employed: sentence length, Dale-Chall readability formula[3], and sentence-based surprisal.

Sentence length is a parameter directly addressed by the PEM. It should go down if any significant PEM influence exists. In case of CoCELD, we calculated median sentence length values for each text (40 texts per decade in total) and then calculated a single median value for each decade. Since RSC is substantially larger than CoCELD, we randomly selected 40 texts[4] for each decade to ensure better comparability between the corpora. The values were subsequently computed following the same procedure. All calculations were performed using a Python script.

Dale-Chall readability formula (Dale and Chall, 1948) is a commonly used readability metric that attempts to capture both syntactic and lexical complexity.[5] The score ranges from 4.9 or lower (level of <=4th-graders) to 9.9 (level of an average college student), and is calculated as follows:

$$0.1579 \times \left(\frac{\text{difficult words}}{\text{total words}} \times 100 \right) + 0.0496 \times \frac{\text{total words}}{\text{total sentences}}$$

Surprisal (Shannon, 1948) is an information-theoretic measure proportional to processing effort

[2]The RSC has been parsed using Universal Dependencies (UD) syntax; however, it is important to note that, at the time this paper was published, we are still in the process of evaluating this parsed version.

[3]Flesh Reading Ease and Gunning Fog Index were much less accurate, indicating law reports to being on par with high-school knowledge. Thus, we excluded them from the analysis.

[4]Dale-Chall readability scores and sentence-based surprisal for RSC are based on the same sample.

[5]For calculation we used the *Textatistic* Python package (Hengel, 2022), which contains an extended version of the original word list used in the formula (e.g., verb tense forms and plural noun forms).

(Hale, 2001; Levy, 2008). It measures the amount of information (in bits) transmitted by a word in context: $S(word) = -\log_2 p(\text{word}|\text{context})$. As context, we use a trigram of the preceding three words of the given word. Similarly to sentence length, we calculate one median sentence-based surprisal score for each decade, estimating the overall processing cost at the text level. High surprisal indicates higher processing effort, indicating a more complex text. To calculate surprisal values for texts within each decade, we first establish a reference corpus for that decade. This reference corpus is composed of all the texts (of the RSC or CoCELD respectively) from the decade, excluding the specific text for which we are calculating surprisal. We then use this reference corpus to generate probabilities for each word in our target text. These probabilities form the basis for calculating surprisal. This method of using a decade-specific reference corpus is advantageous because it provides a contextually relevant baseline for understanding linguistic patterns and changes over time. By comparing the language in a specific text against the broader linguistic trends of its time period, we can more accurately assess the relative novelty or commonality of its usage, thereby gaining deeper insights into the evolving dynamics of language use within that historical context.

3.2.2 Analyzing lexico-grammatical changes

At the lexico-grammatical level, we are interested in linguistic features distinctive of law reports and scientific texts over time. For this, we use Kullback-Leibler Divergence, which is commonly applied to compare two probability distributions of linguistic features (see Klingenstein et al. (2014); Fankhauser et al. (2014); Degaetano-Ortlieb and Teich (2018); Barron et al. (2018) for application across the digital humanities). KLD indicates the number of additional bits of information needed to encode one distribution (here of a decade) using another (a previous decade), and is formalized as:

$$D(A||B) = \sum_i p(\text{feature}_i|A) \log_2 \frac{p(\text{feature}_i|A)}{p(\text{feature}_i|B)} \quad (1)$$

where A stands for a decade and B for a previous decade. Advantageous for interpretability is that KLD calculates the contributions of individual features to a divergence, allowing us to generate feature rankings with the most distinctive features of a decade.

For the lexical level, we apply a unigram model (all words), and for the grammatical level a trigram model (sequences of three parts of speech) to analyze diachronic changes. Given the KLD scores, we subsequently identify those decade pairs that showed the most noticeable differences (high overall divergence) and analyze high-ranking features distinctive of the comparison.

To measure changes related to processing effort, we again use surprisal, but here calculated as the average amount of information that a word or part-of-speech trigram transmits across the whole time period (rather than in a single text). The average surprisal of individual words is calculated by summing the surprisal values of all occurrences of a word and dividing them by the total number of occurrences in a decade:

$$\text{AvSrp}(\text{word}) = \frac{1}{|\text{word}|} \sum_i -\log_2 p(\text{word}_i|\text{context}_i) \quad (2)$$

For part-of-speech trigrams, we first calculate the average surprisal of each of their individual words, sum all resulting values, and divide them by the number of occurrences N of the part-of-speech trigrams in a decade:

$$\text{AvSrp}(\text{postrigram}) = \frac{1}{|\text{N}|} \sum_i \left(\frac{\text{AvS}(\text{word}) + \text{AvS}(\text{word}) + \text{AvS}(\text{word})}{3} \right)_i \quad (3)$$

4 Results and Analysis

Considering the suggestions put forward by the PEM, first, we test whether the overall text complexity is reduced especially for law reports (H1). Second, an in-depth analysis of lexico-grammatical features will show whether the heavy nominal style changes towards a more involved verbal style (H2).

4.1 Overall text complexity

Considering sentence length (see Table 2), suggested to be kept short by the PEM, for law reports it stays relatively stable. For scientific articles, sentence length goes slightly down. However, both remain above the limit of 20 words recommended by the PEM. Based on the Dale-Chall formula, law reports show a continuous increase reaching 9.55 in the 2010s, matching almost the highest possible score and corresponding to language use at the level of a college student. The values for scientific articles are even higher (e.g., 10.08 in 1950s and 10.52

Decade	Sentence Length	
	Law	Science
1950s	31.0	27.0
1960s	30.0	27.0
1970s	29.0	26.5
1980s	30.0	27.8
1990s	29.5	25.0
2000s	27.0	-
2010s	32.0	-

Table 2: Median sentence length in law reports and scientific articles (as measured by the number of tokens)

in 1990s), indicating a major degree of complexity. These results are confirmed by sentence-based surprisal, showing a slight continuous increase for both domains (law: from 6.61 in the 1950s to 6.97 bits in the 2010s; science: from 6.41 to 6.85 bits).

This indicates no shift towards reduced complexity, on the contrary, both registers have become even more challenging to process over time.

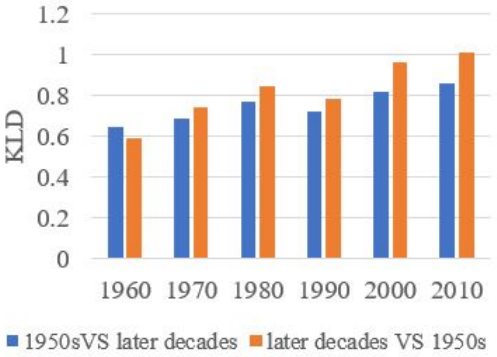

Figure 1: KLD comparison for the 1950s given the other periods and vice versa for law reports.

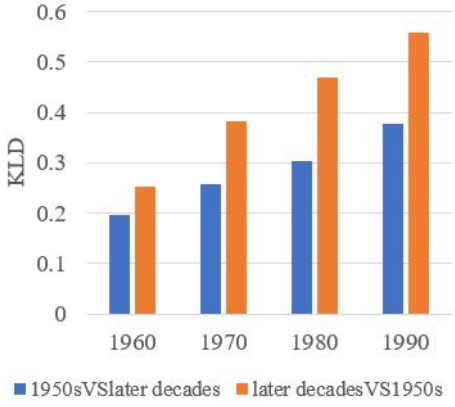

Figure 2: KLD comparison for the 1950s given the other periods and vice versa for scientific articles.

4.2 Changes at the lexico-grammatical level

4.2.1 General diachronic trends

We analyze changes at the lexical level by first asking (a) how much language use of the 1950s can be modeled by a more contemporary model, and (b) how well a language model of the 1950s can capture contemporary language use. Here, we make advantage of KLD's asymmetry (see Section 3), which allows us to model this directionality.

Figure 1 shows a comparison for law reports. For both directions, divergence increases, but in particular for the more contemporary models, which increasingly diverge from the 1950s model over time (orange bars). Thus, more contemporary language use is not well modeled by past language use.

Figure 2 shows a comparison for scientific articles. Again, KLD rises over time, with an even more pronounced tendency, showing how contemporary models are increasingly less well modeled by the 1950s model.

To better understand what drives an increase in divergence, we consider lexical features which are distinctive (i.e. have a major contribution to the increase in divergence). As depicted in Figure 3, law reports from the 1950s are characterized by the presence of words forming part of formulaic expressions (*lordship, noble*), auxiliary *be*, and pronouns (e.g., *I, my, me, their*, etc.). To a lesser extent also modal verbs (*must, can, shall*), mental verbs such as *think* and *agree*, and the relative pronoun *which* are distinctive.

On the contrary, the 2010s (see Figure 4) are characterized by the use of honorifics (*mr, ms*), abbreviations (*ltd, wlr, ukpc*, etc.) and a more pronounced use of nominalizations, especially those ending with *-tion* (*conviction, application, constitution*, etc.).

For scientific articles, the most distinctive feature in the later decades (see Figure 6) is *et* proba-

Figure 3: Lexical features distinctive of the 1950s when modeled by the 2010s for law reports

bly indicating an increase in multi-authored papers. Other relevant features include the conjunction *and*, prepositions *in* and *for* as well as the 1st person pronouns *I* and *we*. Both earlier and later decades include many terms of art, some of which are nominalizations (*absorption, bifurcation, selection*, etc). These tendencies seem to reflect the trend towards phrasal compression indicated by Biber and Gray (2016, 207), which is also depicted by the obvious amount of nominal lexis in the 1990s indicating an increased compound use (cf. Degaetano-Ortlieb (2021)). In contrast, we can observe a more varied use of word classes in the 1950s (see Figures 5) with the determiner *the*, the verb *be*, various prepositions (*of, about, at*), post-modification patterns (*by, which, to*), general verbs (*give, make, obtain, show*), and conjunctions (*but, so*).

4.2.2 Inspecting PEM influence

Let us now have a closer look at those features that are of particular interest to us in the context of the PEM. Here, we specifically address hypothesis H2, i.e. most of the PEM suggestions S2-S8 above for the use of a more involved less nominal style.

Use 1st and 2nd personal pronouns As illustrated in Figure 7, for law reports the overall frequency of personal pronouns[6] decreased, mainly due to the decline in 1st person pronouns, while 3rd person pronouns remained relatively stable over time. This is an indicator of a distant and objective style of writing as opposed to a more involved and subjective one recommended by the PEM (Rodríguez-Puente, 2019).

In scientific articles, 1st person pronouns, in contrast, became more common (see Figure 8). This trend has been observed in various studies on scientific writing (most prominently Hyland (2005)'s work) highlighting that by using personal pronouns, authors can create a sense of dialogue and interaction, making their writing more accessible and reader-friendly, which is in line with the PEM.

Avoid nominalizations Nominalizations[7] rise in frequency in both law reports (6,000 to 7,200 per million) and scientific articles (32,000 to 32,600 per million). This goes clearly against the PEM. We also consider the surprisal of nominalizations to observe tendencies in terms of processing effort. A rise in the use of conventionalized nominalizations would lead to a decrease in surprisal (enhanced predictability and lower processing effort). However, surprisal stays quite stable in both registers (see Figures 9 and 10), which seems to indicate a constant varied use of nominalizations.

Avoid compounds To inspect diachronic trends in compound use, we run KLD at the level of part-of-speech trigrams which allows us to determine distinctive grammatical features of variation for the more contemporary periods against the 1950s.

By inspecting the top ranking phrase and clause types[8], both law reports and scientific articles evolved quite similarly shifting even more towards a denser style of writing with a high proportion of nominal elements (see Figure 11). Importantly, most nominal trigrams characteristic of the later decades are either compounds (gray, NP (comp)) or complex noun phrases (orange, NP+)[9], which contradicts the suggestions of the PEM.

Compare also the top 5 most distinctive trigrams for law reports showing a varied set of trigrams distinctive of the 1960s, while the 2010s are marked by nominal trigrams (see Tables 3 and 4).

We also observe an increase in the frequency of two-noun and three-noun compounds in both registers (from 1,241 to 2,271 per million in law

[6]Following Rodríguez-Puente (2019), we focused on the nominative forms of personal pronouns (*I, you, he, she, we, they*), their accusative and genitive forms and the corresponding reflexive forms.

Figure 4: Lexical features distinctive of the 2010s when modeled by the 1950s for law reports

[7]We consider nouns with the following suffixes: $-ion$, $-ment$, $-al$, $-ibility$, $-ty$, $-ness$.

[8]We grouped part-of-speech trigrams into phrase and clause types as follows: NP: nominal phrase including premodification by adjective and use of determiner, NP+: nominal phrase with a postmodifier or conjunction, NP (comp): compound phrase, VP: clause, VP (modal/ing/ed): clause with modal verb, -ing or -ed forms, to-inf: to-infinitive clause, wh-clause, AdvP: adverbial phrase, CC: conjunction, PrepP: prepositional phrase, that-clause, AdjP: adjectival phrase

[9]Complex noun phrases are grouped under the NP+ label, which includes noun phrases with prepositional or non-finite postmodifiers or noun clauses with more than one head joined by a conjunction.

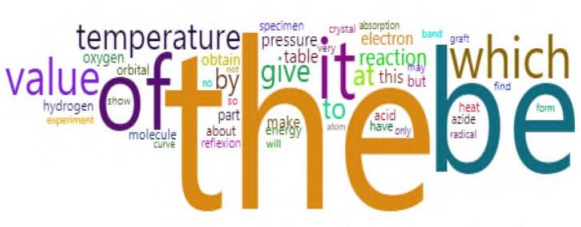

Figure 5: Lexical features distinctive of the 1950s when modeled by the 1990s for scientific articles

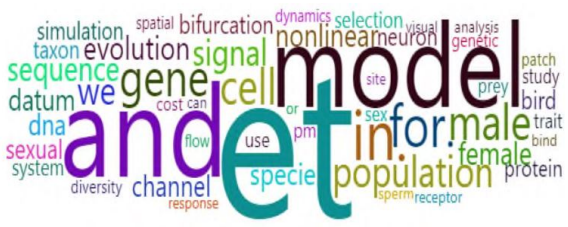

Figure 6: Lexical features distinctive of the 1990s when modeled by the 1950s for scientific articles

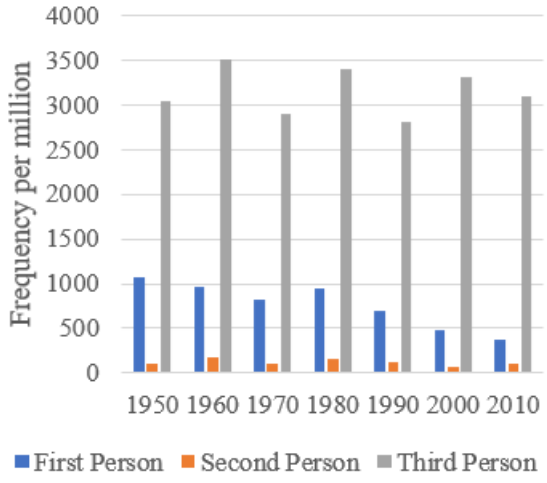

Figure 7: Evolution of personal pronouns in law reports

PoS	Example	KLD
VV.IN.DT	agree with the	0.0026
PP.MD.VV	I must regard	0.0023
DT.NN.WDT	this incident which	0.0022
CC.IN.DT	or by the	0.0019
PP.NNS.MD	their Lordships may	0.0019

Table 3: Top 5 trigrams characteristic of law reports drafted in 1960s

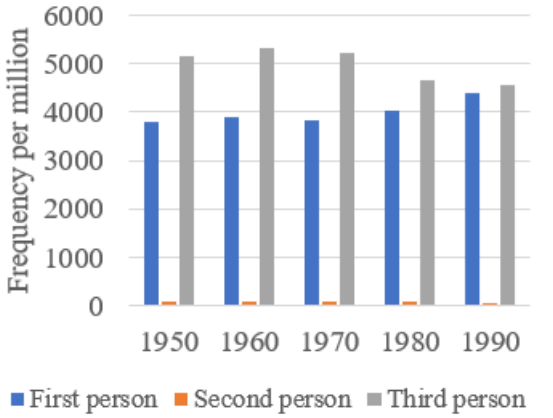

Figure 8: Evolution of personal pronouns in scientific articles

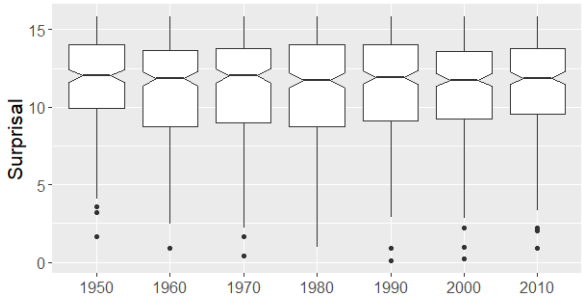

Figure 9: Average surprisal of nominalizations in law reports

reports and from 28,188 to 35,034 per million in scientific articles, X^2 p-value < 0.01), indicating a higher reliance on compact syntactic structures. To link this back to processing effort, we again consider surprisal. For illustration, we compare compound patterns distinctive of the recent decades with simple nominal phrases characteristic of the earlier decades (see Figure 12). There is a considerable difference in surprisal between compounds with three lexical words (e.g JJ.NN.NN: adjective+noun+noun) and simple nominal phrases. Even compound patterns with one function word (e.g., preposition (IN) or conjunction (CC)), which are lower in surprisal, have slightly higher surprisal values than simple nominal phrases.

Interestingly, this trend holds also for the NP+ category (i.e. complex nominal phrases such as NNS.IN.JJ): Patterns distinctive of the later periods tend to be more informationally loaded (higher surprisal) than those distinctive of the earlier periods. This is illustrated in the following examples, where example (1) shows one possible lexical realization of the DT.NNS.VVN trigram (characteristic of the earlier periods) and example (2) shows one possible realization of the NNS.IN.JJ trigram (characteristic of the later periods). The numbers indicate the average surprisal calculated on the decade basis for each element of the trigrams (average values over the trigram are provided in square brackets).

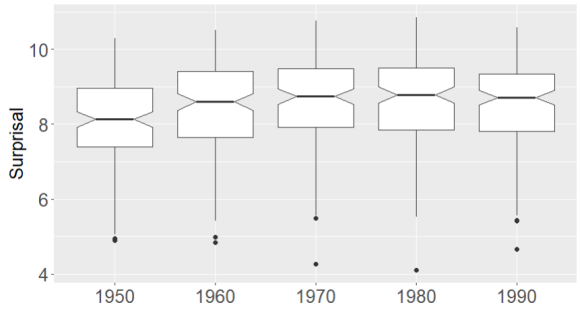

Figure 10: Average surprisal of nominalizations in scientific articles

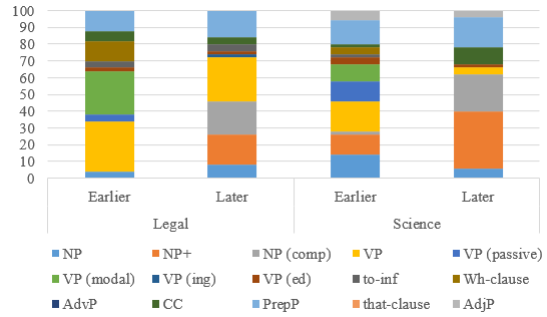

Figure 11: Proportion of the PoS trigrams in law reports and scientific articles

(1) *In the wheat grain a supplementary effect was demonstrated between* ***the/1.841 proteins/8.530 situated/11.447*** *[7.27] in the outer layers of the grain (bran) and those contained in the endosperm.* (RSC)

(2) *The importance of CMT was that it provided the first really practical means of detecting* ***bodies/8.902 at/6.034 normal/9.555*** *[8.16] room temperature.* (RSC)

In summary, compounds are not only more frequently used, which goes against the PEM suggestion but are also heavy in their informational content for both law reports and scientific articles.

Use active voice As already shown, verbal patterns become less distinctive of both domains over time at the expense of a pronounced nominal style. This applies also to passive constructions which dropped significantly in frequency (from 2,923 to 2,611 per million in law reports and from 17,574 to 12,720 per million in scientific articles, X^2 p-value < 0.01 for both registers). Although this is in line with the PEM, an inspection of general English as depicted by the LOB and FLOB corpora shows that this is more likely a general trend in the evolution of the English language, with a significant decrease

PoS	Example	KLD
NP.CC.NP	Regulations and Guidance	0.0072
NP.NP.NP	Land Registration Act	0.0062
IN.NP.NP.	by Theresa Henry	0.0058
NN.IN.NN	disclosure of information	0.0054
DT.NN.NN	the anonymity order	0.0046

Table 4: Top 5 trigrams characteristic of law reports drafted in 2010s

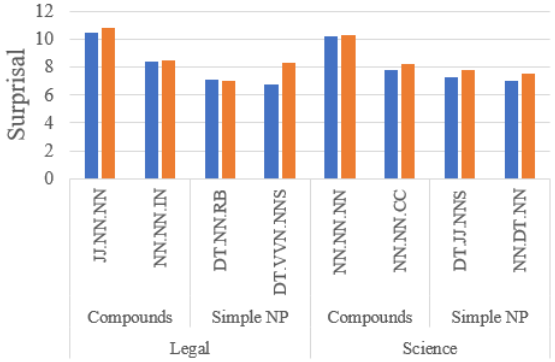

Figure 12: Surprisal of compound patterns and simple nominal phrases in law reports and scientific articles. The blue bars show the average surprisal for the first decade (1950s), the blue ones the average surprisal for the last decade (2010s for law reports and 1990s for scientific articles).

in the use of passives (from 11,324 to 10,541 per million, X^2 p-value < 0.01).

Avoid *shall* As already suggested by the word clouds at the lexical level (see again Figure 3), *shall* has decreased in frequency over time and is distinctive (t-test with p-value < 0.01) for the 1950s. It primarily occurs in direct quotations from other legal documents (see Example (3)).[10]

(3) *Section 5 provided that "no owner* ***shall*** *... convey [or] agree to convey ... any land in a new subdivision . . . "* (CoCELD)

Thus, even though the decrease of *shall* might have been triggered by the PEM, the influence on law reports seems to be rather indirect.

5 Conclusion

We investigated the impact of the Plain English Movement (PEM) on the complexity of legal language in UK law reports from the 1950s to the 2010s, contrasting this with the evolution of scientific language. The study was grounded in the hy-

[10] Evaluated on a random sample of 50 occurrences of *shall* from the 1950s and 2010s each. The analysis yielded 82% and 88% of direct quotations in 1950s and 2010s, respectively.

pothesis that if the PEM had a significant influence, we should see a reduction in language complexity (H1) and a shift from a nominal to a more verbal style (H2) in legal texts. Conversely, we anticipated that scientific language, not being a direct target of the PEM, would not demonstrate similar changes.

Our findings, however, reveal that the impact of the PEM on the complexity of legal language has been minimal. Despite the efforts of the PEM, legal language has largely maintained its traditional complexity and style. This suggests that professional norms and the inherent nature of legal discourse may resist simplification efforts, even in the face of concerted campaigns like the PEM.

Surprisal and more elaborated notions of it (cf. Futrell (2023)) can serve as a robust indicator of the cognitive load imposed on readers, thereby guiding the development of more accessible yet accurate renditions of complex information, often not captured by readability measures. We have employed it to quantify complexity to account for the unpredictability and processing effort associated with language comprehension. However, its application would not only enhance readability but also ensure that critical nuances and technical accuracies are not lost in the process of simplification. For example, endeavors to produce simpler legal texts (cf. Garimella et al. (2022)) would profit from using approaches which reflect processing effort and measure its reduction. The approach is general in nature and can be applied across various fields.

References

Alexander T. J. Barron, Jenny Huang, Rebecca L. Spang, and Simon DeDeo. 2018. Individuals, institutions, and innovation in the debates of the French Revolution. *Proceedings of the National Academy of Sciences*, 115(18):4607–4612.

Douglas Biber and Bethany Gray. 2016. *Grammatical complexity in academic English: Linguistic change in writing*. Studies in English Language. Cambridge University Press, Cambridge.

Douglas Biber and Bethany Gray. 2019. Are law reports an 'agile' or an 'uptight' register? tracking patterns of historical change in the use of colloquial and complexity features. In Teresa Fanego and Paula Rodríguez-Puente, editors, *Corpus-based Research on Variation in English Legal Discourse*, number 91 in Studies in Corpus Linguistics, pages 147–170. John Benjamins Publishing Company.

Ruth Breeze. 2019. Part-of-speech patterns in legal genres: Text-internal dynamics from a corpus-based perspective. In Teresa Fanego and Paula Rodríguez-Puente, editors, *Corpus-based Research on Variation in English Legal Discourse*, number 91 in Studies in Corpus Linguistics, pages 79–104. John Benjamins Publishing Company.

Edgar Dale and Jeanne S. Chall. 1948. A formula for predicting readability: Instructions. *Educational Research Bulletin*, 27(2):37–54.

Stefania Degaetano-Ortlieb. 2021. Measuring informativity: The rise of compounds as informationally dense structures in 20th century scientific english. In Elena Soave and Douglas Biber, editors, *Corpus Approaches to Register Variation*, chapter 11, pages 291–312. John Benjamins Publishing Company.

Stefania Degaetano-Ortlieb and Elke Teich. 2018. Using relative entropy for detection and analysis of periods of diachronic linguistic change. In *Proceedings of the 2nd Joint SIGHUM Workshop on Computational Linguistics for Cultural Heritage, Social Sciences, Humanities and Literature at COLING*, pages 22–33, Santa Fe, NM. Association for Computational Linguistics.

Stefania Degaetano-Ortlieb and Elke Teich. 2019. Toward an optimal code for communication: The case of scientific English. *Corpus Linguistics and Linguistic Theory*, 0(0):1–33. Ahead of print.

European Commission. 2016. *English Style Guide: A Handbook for Authors and Translators in the European Commission*. European Union.

Peter Fankhauser, Jörg Knappen, and Elke Teich. 2014. Exploring and visualizing variation in language resources. In *Proceedings of the 9th Language Resources and Evaluation Conference (LREC)*, pages 4125–4128, Reykjavik, Iceland. European Language Resources Association.

Federal Government of the United States. 2011. *Federal Plain Language Guidelines*.

Stefan Fischer, Jörg Knappen, Katrin Menzel, and Elke Teich. 2020. The Royal Society Corpus 6.0: Providing 300+ years of scientific writing for humanistic study. In *Proceedings of the Language Resources and Evaluation Conference (LREC)*, pages 794–802, Marseille, France. European Language Resources Association.

Richard Futrell. 2023. Information-theoretic principles in incremental language production. *Proceedings of the National Academy of Sciences*, 120(39):e2220593120.

Aparna Garimella, Abhilasha Sancheti, Vinay Aggarwal, Ananya Ganesh, Niyati Chhaya, and Nandakishore Kambhatla. 2022. Text simplification for legal domain: Insights and challenges. In *Proceedings of the Natural Legal Language Processing Workshop 2022*, pages 296–304, Abu Dhabi, United Arab Emirates (Hybrid). Association for Computational Linguistics.

Bryan A. Garner. 2001. *Legal Writing in Plain English: A Text with Exercises*. University of Chicago Press.

Maurizio Gotti. 2012. Text And Genre. In *The Oxford Handbook of Language and Law*. Oxford University Press.

John Hale. 2001. A probabilistic Earley parser as a psycholinguistic model. In *Proceedings of the 2nd meeting of the North American Chapter of the Association for Computational Linguistics on Language technologies*, pages 1–8, Pittsburgh, PA.

Erin Hengel. 2022. Publishing While Female: are Women Held to Higher Standards? Evidence from Peer Review. *The Economic Journal*, 132(648):2951–2991.

Ken Hyland. 2005. *Metadiscourse: Exploring Interaction in Writing*. Continuum.

Hannah Kermes, Stefania Degaetano-Ortlieb, Ashraf Khamis, Jörg Knappen, and Elke Teich. 2016. The Royal Society Corpus: From uncharted data to corpus. In *Proceedings of the 10th International Conference on Language Resources and Evaluation (LREC)*, pages 1928–1931, Portorož, Slovenia. European Language Resources Association.

Sara Klingenstein, Tim Hitchcock, and Simon DeDeo. 2014. The civilizing process in London's Old Bailey. *Proceedings of the National Academy of Sciences*, 111(26):9419–9424.

Roger P. Levy. 2008. Expectation-based syntactic comprehension. *Cognition*, 106(3):1126–1177.

Joanne Locke. 2003. The plain language movement. *AMWA Journal*, 18(1):5–8.

Beth Mazur. 2000. Revisiting plain language. *Technical Communication*, 47:205–211.

Paula Rodríguez-Puente. 2019. Interpersonality in legal written discourse: A diachronic analysis of personal pronouns in law reports, 1535 to present. In Teresa Fanego and Paula Rodríguez-Puente, editors, *Corpus-based Research on Variation in English Legal Discourse*, number 91 in Studies in Corpus Linguistics, pages 171–200. John Benjamins Publishing Company.

Paula Rodríguez-Puente and David Hernández-Coalla. 2023. The : A new tool for analysing recent changes in english legal discourse. *ICAME Journal*, 47(1):109–117.

Helmut Schmid. 1995. Improvements in Part-of-Speech Tagging with an application to German. In *Proceedings of the ACL SIGDAT-Workshop*, Cambridge, MA.

Karen A. Schriver. 1997. *Dynamics in Document Design: Creating Text for Readers*. John Wiley Sons, New York, NY.

Cassie Sedgwick, Laura Belmonte, Amanda Margolis, Patricia Osborn Shafer, Jennifer Pitterle, and Barry E. Gidal. 2021. Extending the reach of science – talk in plain language. *Epilepsy Behavior Reports*, 16:100493.

Claude E. Shannon. 1948. A mathematical theory of communication. *Bell System Technical Journal*, 27:379–423, 623–656.

Nathaniel J. Smith and Roger Levy. 2013. The effect of word predictability on reading time is logarithmic. *Cognition*, 128(3):302–319.

Christopher Williams. 2005. *Tradition and Change in Legal English*. Peter Lang Verlag, Lausanne, Schweiz.