Poetry, Songs, Literature, Legalese and Translationese: Automated Sentence Complexity Perspective

Vilém Zouhar
2022 project report ETH Zurich
vzouhar@inf.ethz.ch

Abstract

Although non-trivial to measure, natural texts come in varying complexities. As a result, multiple domains and genres can be compared based on their complexities. In this study, focused on measuring sentence complexity, I use automated methods of complexity estimation to compare poetry, natural prose, literary prose and machine and human translation. The conclusion is that old poetry and old literature is more complex than their modern counterparts, as measured by language model complexity, Flesch Reading Ease and syntactic depth. Furthermore, we observe that machine translations are faithful to human references in terms of sentence complexity, which is a positive result for the translation industry. Most importantly, this paper discusses the reason for different complexities across varying text domains, which is framed as “form (complexity) follows function and aesthetics with least effort.”

github.com/zouharvi/genre-complexity

1 Introduction

Most speakers would be able to quickly recognize poetry from natural prose and, after a few sentences, possibly also natural prose from literary prose.¹ There is a plethora of aspects which determine the differences between these three genres, especially focused on defining poetry [Stevenson, 1957, Collins, 1988, Jakobson, 2010, Fabb, 2015]. I take an inductive stance and examine these differences through the lenses of sentence complexity. It is important to note that all of these genres encompass a vast realm of various language usages. For example, poetry does not include solely rhymes, but also poems which are written in a prosaic manner. On the other hand of the spectrum of poetry, I also include songs under the umbrella genre of poetry. The reason for this is that lyrics are sometimes considered an extension of poetry because they have to follow an external form given by the music [Galvez, 2012, Singhi and Brown, 2014].

This paper heavily utilizes the hypothesis of Hale [2016] in which complexity is linked to surprisal. However, approaches from the research of text simplification [Alva-Manchego et al., 2020, Al-Thanyyan and Azmi, 2021, Štajner, 2021] are also used. The overarching thesis of this paper is that languages, as observed e.g. in legal and encyclopedic texts, are vastly different. I explain the observed empirical differences between the distribution using different optimization objectives of authors of said texts which then leads to an intuitive generalization of form following function and aesthetic requirements with minimum effort.

In a different field of translatology, the translation of source text $s$ can be modelled as a process either by a human: $s \rightarrow_{\text{hum.}} t$ or a machine translation system $s \rightarrow_{\text{mach.}} t$. Human and machine translations

¹For the purpose of this paper, natural prose is used for natural usage of language outside of literature and literary prose stands prosaic language as used in literature (fiction books).
are known to be of different distributions [Roberts et al., 2020]. With a similar setup, I evaluate their complexity and pose it as one additional way to understand possible differences between their distributions.

First, I formulate the following motivating questions for the experimental part of this paper:

<table>
<thead>
<tr>
<th>MQ1:</th>
<th>Does text complexity distinguish poetry, literary prose and natural prose?</th>
</tr>
</thead>
<tbody>
<tr>
<td>MQ2:</td>
<td>Is human translation equally complex to machine translation?</td>
</tr>
</tbody>
</table>

I first review what text complexity is and how it can be measured in Section 2. Then I quantitatively and qualitatively compare different genres of texts, most importantly poetry and prose in Section 3 (MQ1). Unsurprisingly, I find that legal texts are by far the most complex. Within prose, the less prepared and curated it is, the lower its complexity (law > Wikipedia > presentation > natural speech). Within poetry, which is constrained by the formal requirements of its genre, songs are lexically and syntactically the least complex. In this section, I also sketch a theory, which legitimizes the link between authors’ optimization goals and sentence complexity.

In Section 4 I briefly verify that machine translation preserves the same level of complexity as human translations (MQ2). Finally, in Section 5, I discuss important limitations and the results in a broader context.

2 What is complexity?

In this section, I will first refine the intuition of what text complexity is and then later discuss how it can be measured in practice.

2.1 Examples of Complexity

Consider the following two descriptive poems (Texts 1 and 2).² While the poems describe the same scene using similar concepts, most readers would agree, that the poem on the left is more difficult to read and therefore more complex.³

| The silvery moon above |
| and the clouds that drift and shake |
| their shadows dance and mourn |
| across the burbling waters below. |

Text 1: Complex poem.

| The moon is shining bright |
| Its light falls on the lake tonight |
| The clouds drift across the sky |
| Their shadows shake, oh very dark |
| And the rippling water, silver and stark |

Text 2: Simple poem.

There are multiple reasons why the first poem could be considered more complex. I divide the reasons into the following categories:

- **Lexical**: The complex poem uses either words which have lower frequency in the English language, such as *burbling* instead of *rippling* or words which require some further processing, such as *silvery moon*. This requires the association of *silver color* with the property of being *shiny*.
- **Syntactic**: The complex poem is written as a single sentence, while the simple poem contains five individual short sentences. The maximum syntactic tree depth of the simple poem is therefore bounded by the maximum length of a single line.
- **Conceptual**: The complex poem attributes agency to an inanimate object *moon* which *mandates*. Similarly, it anthropomorphizes *shadows* which *dance and mourn*.

Next, consider the following two prose texts in Texts 3 and 4. The simple version, which makes a tradeoff between literary merit and readability, is less complex for similar reasons as with the poems. Additionally, it contains greater coherence. The complex prose introduces a bracket with additional information that is not well incorporated into the rest of the sentence, which breaks the coherence.

²Crafted by the author based on *Summer’s Splendor By The Sea*, Patricia L. Cisco, 2018.
³Later I will revise this assumption that reading difficulty means complexity.
Mr. Utterson the lawyer was a man of a rugged countenance, that was never lighted by a smile; cold, scanty and embarrassed in discourse; backward in sentiment (he was born old, as Mr. Guest, his clerk, who was later interviewed, would recount); lean, long, dusty, dreary, and yet somehow lovable.

Mr. Utterson was a lawyer with a rugged appearance and a personality that was cold, reserved, and unemotional. He was not a very talkative person and was not very affectionate, but somehow people still found him lovable. His accountant, Mr. Guest, was interviewed later and would say that Mr. Utterson was born old.

Text 3: Complex prose adapted from Strange Case of Dr Jekyll and Mr Hyde, Robert Louis (1886).

Text 4: Simple prose (simplified by the author from the original).

We are able to easily determine which of the two poems or which of the two prose texts is more complex. However, at a first glance, it seems impossible to answer a hypothetical question: “Is the poem in Text 1 more or less complex than the text in Text 3?”

To complicate matters, texts do not exist in an imaginary vacuum and the context and the reader need to be taken into account [Štajner, 2021]. Consider the love letters of Abelard and Heloïse (Text 5). The original in Old French would be too complex, to the point of being incomprehensible because of the language barrier. Furthermore, most modern readers lack the knowledge of both general 12th-century context and also of the relationship between the sender and the recipient. In contrast, relevant scholars and the recipient of this letter find it less complex than the general modern readers. The second, more contemporary, example in Text 6 contains ungrammatical sentences which are unexpected and increase complexity. However, when an avid listener hears this song multiple times, the specific lines may be memorized and expected, therefore having much lower complexity.

But oh! where is that happy time fled? I now lament my lover, and of all my joys there remains nothing but the painful remembrance that they are past. Now learn, all you my rivals who once viewed my happiness with such jealous eyes, that he you once envied me can never more be yours or mine.

We don’t need no education
I don’t need no thought control
No dark sarcasm in the classroom
Teachers leave them kids alone
Hey! Teachers! Leave them kids alone!

Text 5: 12th century love letters [Radice, 1974], translated from Old French to English.

Text 6: Late 20th century English song from Pink Floyd’s rock opera The Wall (1979).

2.2 Quantitative Methods

I define three groups of methods in which text complexity can be evaluated from most to least faithful: human extrinsic, human intrinsic, and machine intrinsic.

**Human extrinsic.** Despite the issues that individual readers (their knowledge and experience) and specific time contexts pose, I attempt to quantify sentence complexity formally. I first use expected human reading effort as a proxy for complexity:

$$ C(t) = \mathbb{E} h \in \text{readers} \; \text{effort}(h, t) $$

Note that this is the expected value across all possible readers. In the spirit of the previous discussion, we would be more interested in answering the question: “How complex is this text for an average person in 2022?” The formula then becomes $ C’(t) = \mathbb{E} h \in \text{readers in 2022} \; \text{effort}(h, t) $. For practical reasons during experiments, we are usually only able to hire a subset of possible readers for which we are able to measure the reading effort. The underlying assumption is that those readers are representative of the readers in 2022 distribution (or any other, e.g. scientific community etc.):

$$ C''(t) = \mathbb{E} h \in \text{participants} \; \text{effort}(h, t) $$

The possible effort function approximations range from MRI scans to performing the cloze task or, in the simplest of scenarios, reading times. This addresses the connections between complexity and reading effort and ungrammaticality (less grammatical texts are less expected, require more reading effort and therefore correspond to higher complexity). It should be noted, that complexity is not a quantity that is always better when low. While the poem in Text 2 has lower complexity and is easier to read, it is also duller and probably contains less artistic value.

---

4I will later address the connection of ungrammaticality and text complexity.
**Human intrinsic.** Many complexity and text simplification studies [Xu et al., 2016, Leroy et al., 2016, Schwarzer and Kauchak, 2018] evaluate complexity by asking participants directly the question “How complex is the following text?” While this may be a more robust approach to evaluation which also offers a degree of explainability (i.e. a follow-up question “Why did you grade the text as «very complex»?”), it also hinges on the researchers’ prior idea of what complexity is, as codified in the guidelines they created, and the participants’ prior conception of complexity.

**Machine intrinsic.** The issue of both human extrinsic and intrinsic approaches is the cost of participants, variance and irreproducibility. For this reason, many methods for automatic complexity estimation have been developed.

The way in which complexity is commonly approached in modern NLP is *language model perplexity* [Chen et al., 1998, Hale, 2016, Meister et al., 2021]. Intuitively, it can be seen as assigning a probability to text $W$ given a model $P$ and its parameters $\theta$.

$$PP(P(W|\theta)) = \prod_{i=0}^{W} P(w_i|w_{<i}, \theta)$$  

$$\log PP(P(W|\theta)) = \sum_{i=1}^{W} \log P(w_i|w_{<i}, \theta)$$  

$$\text{avg. } \log PP(P(W|\theta)) = \frac{1}{W} \sum_{i=1}^{W} \log P(w_i|w_{<i}, \theta)$$

The popular link between LM perplexity and text complexity is that higher perplexity corresponds to higher human surprisal and text complexity [Hale, 2016, Goodkind and Bicknell, 2018, Arps et al., 2022]. However, there are serious issues with using this method to evaluate complexity across domains. Specifically, the given language model’s parameters $\theta$ are trained on a specific corpus, with a specific data distribution. (Negative) perplexity to some extent then indicates how well the presented text corresponds to the training domain. For the purposes of experiments in this paper, I use GPT-2-distil for performance reasons. See examples of average measured PPL in Appendices B and C.

The *Fleisch-Kincaid Grade Level* [Kincaid et al., 1975] is a well-known *readability* metric, which I repurpose for measuring text complexity on word level. It is computed as

$$\text{FKGL-READABILITY} = -A \cdot \frac{\#\text{words}}{\#\text{sentences}} - B \cdot \frac{\#\text{syllables}}{\#\text{words}} + C$$

$$\text{FKGL-GRADE}(x) = \text{GRADE}($$FKGL-READABILITY(x)$$))$$

with specific positive coefficients $A$, $B$, $C$ and monotone non-increasing function grade. From the formula, it is clear that shorter sentences and words produce higher readability and lower grade level.

Finally, the maximum *syntactic depth* measures the structural complexity of the sentence. Although it can be easily understood intuitively, I define it formally. Given a particular parse of the sentence, which provides the CHILDREN function:

$$\text{MAXDEPTH}(x) = \begin{cases} 0 & \text{if } \text{CHILDREN}(x) = \emptyset \\ 1 + \max_{c \in \text{CHILDREN}(x)} \text{MAXDEPTH}(c) & \text{otherwise} \end{cases}$$

At first, given two sentences, one may be tempted to make a universal claim, that longer sentences are more complex. However, it is possible to create simple and very long sentences by e.g. enumeration: “I went to Zürich, Zapopan, Zaria, Zoliborz, ….” Despite its length, this sentence would have very low syntactic depth and would correspond to our notion of not being complex. The usage of syntactic depth as a measure of complexity is not novel [Brunato et al., 2018, Martin et al., 2020, Arps et al., 2022].
3 Complexity of Poetry and Prose

In this section, I attempt to compare the genre of poetry and prose in terms of complexity. To this end, I use multiple domains which, especially if separated by creation time, allow us to make conclusions regarding their differences within one genre. Their overview is shown in Table 1.

<table>
<thead>
<tr>
<th>Genre</th>
<th>Domain</th>
<th>Dataset</th>
</tr>
</thead>
<tbody>
<tr>
<td>Natural prose</td>
<td>natural speech</td>
<td>Blended Skill Talk [Smith et al., 2020]</td>
</tr>
<tr>
<td></td>
<td>presentation</td>
<td>TED Talks [Tiedemann, 2012]</td>
</tr>
<tr>
<td></td>
<td>legal</td>
<td>joelito/Multi_Legal_Pile</td>
</tr>
<tr>
<td></td>
<td>encyclopedic</td>
<td>Simple English Wikipedia [Jiang et al., 2020]</td>
</tr>
<tr>
<td>Poetry</td>
<td>modern poetry</td>
<td>merve/poetry, Gutenberg poetry [Jacobs, 2018]</td>
</tr>
<tr>
<td></td>
<td>old poetry</td>
<td>merve/poetry, Gutenberg poetry [Jacobs, 2018]</td>
</tr>
<tr>
<td></td>
<td>songs</td>
<td>Annanay/aml_song_lyrics_balanced</td>
</tr>
<tr>
<td>Literary prose</td>
<td>modern literary prose</td>
<td>10 books from Gutenberg</td>
</tr>
<tr>
<td></td>
<td>old literary prose</td>
<td>10 books from Gutenberg</td>
</tr>
</tbody>
</table>

Table 1: Overview of datasets and domains used for comparison. From each domain, at least 1k (data availability limitation) and at most 50k sentences were used. Old poetry and literature are considered until the 18th century. See Appendix A for an overview of selected books and examples.

The average complexities are shown in Figure 1. An immediate observation is the dominance in the complexity of the legal texts. A second observation is that the ordering of complexity, as given by FKGL and tree depth, is identical. Furthermore, universally old literary prose seems to be more complex than modern literary prose, as measured by the three metrics. As expected, Simple Wikipedia is simpler than its counterpart but more complex than less curated texts, such as presentations or chats. The results for language model complexity seem to be misaligned with our intuition and evaluation of individual samples confirms, that it is hard to consider this quantity to be a meaningful and interpretable indicator of sentence complexity. A possible explanation is that, to some degree, perplexity measures the difference between training and evaluation text distributions.
3.1 Text as Artefact of Optimization

In this section, I provide a speculative legitimization of the observed differences in complexities among genres. It can be summarized as “form follows function and aesthetic with minimum effort”. 5

Consider the examples in Appendix B (Texts 8 to 10), chosen to be prototypical representatives of old poetry, modern poetry and modern song lyrics. The song (Text 10) contains very short phrases which are not even sentences. This is because the lyrics need to follow a specific meter in order to fit the rhythm of a modern pop song (i.e. follow function) and the artist chose shorter segments as, possibly, easier to work with. The old poem (Text 8) also contains a strict meter. However, the specific meter choice (iambic hexameter), which also influences the length of one line, that of the sentence and ultimately the complexity, is made not because of functional necessity, but out of aesthetic norms of the 17th-century poetry. The complexity of the modern poem (Text 9) is guided by both contemporary poetry aesthetics and serving the author’s artistic intent. It would be possible to write the old poem (Text 8) more simplistically, as shown in the introductory example in Texts 1 and 2, though it would be highly unusual for the specific time period and would lessen its artistic value.

In contrast to the overall simplicity of presented poems, the legal text in Appendix C (Text 11) is highly complex and requires higher effort to read. While it would be possible to create a lyrical version of the text (Text 12), making it easier to read (lower complexity score), it omits certain information (e.g. company, offices, “Buyer”) and adds new, fabulated, information (e.g. mug) in order to fit the poetic aesthetic. A much more realistic scenario follows a rhetorical question that layman readers of legal texts are posing: “Why didn’t they write it in a simpler manner?” Example in Text 13 shows a simplified counterpart of the original text. While the complexity is again lower, new ambiguities arise (e.g. address not having to be registered)

While considerable effort could be exerted by the author to fix these problems, it is unnecessary. The text with the lowest effort to craft that satisfies the constraints is chosen. This is in line with formalisms of conversational paradigms [Barasa, 2010, Plantadosi et al., 2012, Buşu, 2022]. As discussed in Section 2, the target readership needs to be taken into account. In the case of many legal texts, the readers are trained lawyers which are not slowed down by the text’s complexity.

Overall, it seems to be the case that language complexity closely follows functional and aesthetic requirements. Furthermore, it is optimized to minimize speaker effort within constraints of acceptability.

4 Complexity in Translation

In contrast to evaluating differences among genres, similar evaluation for translation requires parallel corpora, which are less readily available, especially in literary genres. Because of this, I limit the data in this section to 10k sentences from the Web [Arteche and Schwenk, 2019]. I translate texts from German to English using 3 publicly available strong machine translation models [Ng et al., 2020, Rothe et al., 2020, Tiedemann and Thottingal, 2020] and compute the complexities also for human references.

![Figure 2: Average of sentence complexity metrics for human translations and corresponding machine translations. Lower values mean simpler texts.](image)

5In our case, form being the complexity level.
A comparison in Figure 2 shows very few differences between various translations. Furthermore, there is no distinct pattern between the complexities of machine systems and human translations. This is a positive result for the MT community because it shows, that the translation faithfully transfers the complexity as a human would. Indeed, the average correlation between the machine-translated outputs and human ones is $\rho_p = 0.74$ ($p < 0.0001$, Pearson) and $\rho_s = 0.85$ ($p < 0.0001$, Spearman). Nevertheless, I include an example in which machine translation systems created too complex of a translation Text 7. The fact that the translation was too literal provides an idea of evaluating complexity comparison as part of a machine translation evaluation metric.

Source: Noch nie ist etwas Großes erreicht worden, indem man dem Risiko aus dem Weg ging.
Human reference: Nothing great has been achieved by playing it safe.
(System 1: 'Never before has anything great been achieved by taking the risk out of it.
(FKGL: 6.8, tree depth: 6.0, PPL: 11.3)

Text 7: Mismatch in sentence complexity between a hypothesis and reference translation.

5 Discussion

In the context of related work, various genres have been framed under the “form follows function” paradigm, such as poetry [Schwemer, 2014], literary prose [Glenn, 2004] and historical texts [Kootz, 2020]. With respect to those, this essay shows a specific feature (complexity) which can be interpreted in a similar manner.

5.1 Limitations

There are multiple limitations to this work. The most notable one is the assumption of homogeneity of genre and domains. While I attempted to choose representative and balanced samples from each domain, it is impossible to do so because of the great variance in writing styles. For example, contrast the complex Arthur Rimbaud’s poem Season in Hell (1873) with the simple and playful poetry of Edward Lear Nonsense Songs, stories, Botany, and Alphabets (1870). The other limitation is the use of standard sentence complexity metrics, which produce some number given a piece of text. However, the complexity of poetry is certainly different from the complexity of legal texts. I focused on lexical and syntactic complexities but it may be the case that the most fundamental part of the complexity is conceptual, which is harder to measure automatically. It would be fallacious to therefore make statements in the form of “poetry is less complex than legal texts” even though the metric numbers suggest this.

5.2 Future Work

Through this work, I identified the following two directions for future work. On a practical side, the use of complexity could be explored as an additional feature for machine translation metrics. This is justified by the example where the deviation from the complexity of the reference leads to worse translation quality. More ambitiously, the idea of a unified system for text complexity across genres could be explored either by proposing such a system or by providing criticism on why such a system could not exist.

6 Conclusion

In this paper, I provided an introduction to how complexity across literary genres and domains could be studied using standard complexity metrics. This line of study offers one possible explanation for answering the question of what motivates the specific writing style of a particular text genre and domain. The proposed hypothesis is that the complexity aspect of the form is also given by the function and aesthetical requirements of the text together with the principle of least effort.
References


A Overview of Selected Literary Prose

All books are translated into contemporary English.

Old literary prose: The Middle Class Gentleman (Moliere), The Life and Adventures of Robinson Crusoe (Daniel Defoe), The Odyssey (Homer), Grimms’ Fairy Tales (Jacob Grimm and Wilhelm Grimm), Pride and prejudice (Jane Austen), The Expedition of Humphry Clinker (Tobias Smollett), The Adventures of Ferdinand Count Fathom (Tobias Smollett), The Complete Works of William Shakespeare (William Shakespeare), Middlemarch (George Eliot), A Room With A View (E. M. Forster).

Modern literary prose: The Adventures of Tom Sawyer (Mark Twain), Moby-Dick or The Whale (Herman Melville), Alice’s Adventures in Wonderland (Lewis Carroll), The Brothers Karamazov (Fyodor Dostoyevsky), Ulysses (James Joyce), The Strange Case Of Dr. Jekyll And Mr. Hyde (Robert Louis Stevenson), Metamorphosis (Franz Kafka), The Time Machine (H. G. Wells), The Hound of the Baskervilles (Arthur Conan Doyle), The Sun Also Rises (Ernest Hemingway).

B Poetry Genre Examples

| The silver swan, who living had no note, / When death approached, unlocked her silent throat; |
| Leaning her breast against the reedy shore, / Thus sung her first and last, and sung no more; |
| Farewell, all joys; Oh death, come close mine eyes; / More geese than swans now live, more fools than wise. |

Text 8: Old poem Silver Swan, Orlando Gibbons, 1612. FKGL: 3.1, tree depth: 3.2, PPL: 11.5
The sea-wash never ends.
The sea-wash repeats, repeats.
Only old songs? Is that all the sea knows?
Only the old strong songs?
Is that all?
The sea-wash repeats, repeats.


**C Legal Prose Genre Example**

This sales contract, effective as of [DATE], is made and concluded between [BUYER’S NAME], a company organized and existing in [STATE], with offices located at [ADDRESS] (hereinafter the “Buyer”), and [NAME OF SELLER], a company [STATE], with a registered address located at [ADDRESS] (hereinafter, the “Seller”). Considering that, the Seller is the manufacturer and/or distributor of [PRODUCT DESCRIPTION], and considering that, the Buyer wishes to buy from the Seller, and the Seller wishes to sell said products to the Buyer only under the terms and conditions contained in this Contract of sale.

Text 11: Excerpt from a legal text. FKGL: 24.3, tree depth: 8.5, PPL: 8.8

This is a sales contract that goes into effect on [DATE]. It is between the [BUYER’S NAME] (“Buyer”), a company located in a [STATE] at [ADDRESS], and the [NAME OF SELLER] (“Seller”), a company located in [STATE] at [ADDRESS]. The Seller manufactures or distributes [PRODUCT DESCRIPTION], and the Buyer wants to buy this product from the Seller under the terms and conditions stated in this contract. The Seller wants to sell the product to the Buyer only under these terms and conditions.

Text 13: Excerpt from a legal text. FKGL: 8.3, tree depth: 5.3, PPL: 9.8