Monetary systems implement perhaps the most common form of economic interactions among people in a society. In a similar manner, the instrument of language provides individuals with the ability to exchange ideas and information. While it is not difficult to see shadows of hierarchical organization in language in general, these features are most prominent in the well-defined structure of written language.

The occurrences of elements at the levels of organization represented by letters, syllables, words, sentences, paragraphs, and chapters are relatively easy to determine. However, unlike the lifeless structure of a monetary system, written language demonstrates a wealth of complexity between the levels of words and sentences. This complexity is so prevalent at these scales that hierarchical organization ceases to be a distinctive feature.

The most obvious candidates for additional levels of organization are perhaps phrases and clauses. However, it is not clear how to organize these structures hierarchically. Adjective phrases and clauses modify noun phrases and clauses, whereas adverb phrases and clauses modify verb phrases. However, clauses consist of a subject, which can be a noun phrase or clause, composed with a verb phrase. We see that our description is plagued by a nesting property. This nesting or recursion of the various levels of structure is responsible for the complexity\(^1\), which is not present in the system at the extreme scales where hierarchical organization dominates.

The breakdown of the hierarchy from a simple tree-graph into the complex cycles that have become familiar trademarks of self-organized complex systems makes the study of hierarchical structure difficult at this scale. For simplicity, we approximate hierarchical structure by concentrating on noun phrases (and clauses) by treating them as a single level below the level of verb phrases. Adjective and adverb phrases and clauses are ignored, since it is not clear how these structures could relate to one another hierarchically.

In our analysis of the first chapter of the English translation of *The Little Prince* (de Saint-Exupéry 1943), we denote noun phrases as text within curly braces \{\cdot\}, and verb phrases as text within the square brackets [\cdot]. The following sentence demonstrates the complexity that can arise due to the nesting of noun phrases:

\[
\{\text{The grown-ups' response, \{this time\}\]}, [\text{was \{to advise \{me\\} \{to lay aside my drawings of \{boa constrictors\}, whether from \{the inside\} or \{the outside\}\}, and \{devote \{myself\} instead to \{geography\}, \{history\}, \{arithmetic\} and \{grammar\}\}}]. \text{ (de Saint-Exupéry 1943, p. 1)}
\]

The following sentence in the text provides an example of recursive nesting of verb phrases through the vehicle of the noun clause, which is enclosed within less-than and greater-than symbols <\cdot>:

\[
\text{An entertaining recursive constructive model of noun phrases and clauses is presented by Hofstadter (1979, pp. 130-4) in his discussion on ornate and fancy nouns.}
\]

\(^1\) An entertaining recursive constructive model of noun phrases and clauses is presented by Hofstadter (1979, pp. 130-4) in his discussion on ornate and fancy nouns.
That is why, at the age of six, I gave up what might have been a magnificent career as a painter. (de Saint-Exupéry 1943, p. 1)

The nested verb phrases are used to construct noun clauses, which act as nouns in the higher-order verb phrases.

There are sentences in the text, which contain quotations consisting of several sentences themselves. Hence, nested sentences can also be observed. Other subtleties such as the understood noun 'years' after the adjective 'six' in the example above, in addition to the more creative absence of verb phrases can occur. Table 1 lists the number of structural elements in the first chapter of The Little Prince.

TABLE 1 and FIGURE 2 ABOUT HERE

We find that the results follow a power law (Fig. 2). However, since there is no known way to assign a value to any level and hence a modulus to this system, we cannot prove that grammatical structure follows the power law expected for a HMS. Instead, we assume that the HMS model holds (Eqn. 4) and use the data to solve for the best estimate of $M$ and $n_0$, ($M = 6.15 \pm 1.01$ and $n_0 = 2144 \pm 87$). Using data acquired in a more comprehensive study by Negro et al. (1987) of an Italian newspaper containing 14 articles (49751 letters) we estimated a modulus of $M = 5.37 \pm 0.12$. The similarity between the moduli of two extremely different samples of English and Italian text is quite remarkable and speaks to the importance of hierarchical structure as a concept in the study of written language. Future research may elucidate what principles or constraints determine optimal values of moduli and how these vary with language and the type of text.

It is interesting that the system follows the power law over all scales, especially since hierarchical organization gives way to recursive structures in the intermediate levels. This suggests that the complex recursive structures are built upon a hierarchical backbone and retain some features of the hierarchical substrate. One might wonder why recursive complexity is limited to the intermediate levels of organization. There are two possible explanations for the lack of recursion in the lower levels. First, it is well known that correlations in messages are important for error checking. The hierarchical nature of language at these levels greatly simplifies this task. Second, the letters and syllables are used to define words; significant subsets of which correspond directly to objects and
actions in our environment. These low-level hierarchical structures are used to build the basic ‘useful’ element of written language – the word.

There is often a complex interplay among objects and actions, which necessitates complex descriptive structures when one begins combining words into phrases, clauses, and sentences. However, biological constraints on memory place constraints on our ability to resolve recursive structures. The limitation of being able to keep no more than 7 ± 2 items in short term memory simultaneously limits both the depth and the range of the recursive complexity that occurs in written language. Thus higher levels of organization are constrained to remain hierarchical. This does not mean that distant parts of text do not refer to one another\(^2\); only that recursion cannot occur.

\(^2\) In fact, the first word in our second example sentence from *The Little Prince* is the pronoun “That”, which is used to represent the entire preceding sentence in the text, our first example. By using the mechanism of the pronoun, recursion in the form of nested phrases or sentences is avoided.
Table 1. Occurrences of elements of written text in the first chapter of the English translation of *The Little Prince*. Numbers in parentheses include counts of noun clauses and nested sentences.

<table>
<thead>
<tr>
<th>Level of Structure</th>
<th>Occurrences</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>Letters: 2177</td>
</tr>
<tr>
<td>1</td>
<td>Syllables: 711</td>
</tr>
<tr>
<td>2</td>
<td>Words: 513</td>
</tr>
<tr>
<td>3</td>
<td>Noun Phrases: 172 (178)</td>
</tr>
<tr>
<td>4</td>
<td>Verb Phrases: 59</td>
</tr>
<tr>
<td>5</td>
<td>Sentences: 34 (39)</td>
</tr>
<tr>
<td>6</td>
<td>Paragraphs: 11</td>
</tr>
<tr>
<td>7</td>
<td>Chapters: 1</td>
</tr>
</tbody>
</table>
Hierarchical Structure of Written Text

Log Level Value (h \log_{10} M)

Number of Occurrences

0 2 4 6

1
10
100
1000
10000
Figure 2. The occurrence of elements of written text (Table 1) follows a power law behavior. The theoretical slope of -1/2 was used to estimate a modulus $M = 6.15 \pm 1.01$. 