

## **Textometry: A Method for Numerical Representation of a Text**

**Assistant Prof. Dr. İlker Aydın**  
Yüzüncü Yıl Üniversitesi  
Van, Turkey

**Emrullah Şeker**  
Muş Alparslan Üniversitesi  
Muş, Turkey

### **Abstract**

*This study aims to suggest a systematic text linguistic analysis method. It includes numerical representation of any text not only to illustrate the textual features of any text in concrete numerical terms but also to set a standard of classification via these numerical data. In order to achieve this aim, we initially set our variables upon the fundamental characteristics of what makes a piece of writing a text. The resulting variables were grouped into non-textual, textual and metatextual categories, as to which four different types of sample texts were analyzed. The results were illustrated in tables and represented in numerical values and then compared. The different types of texts produced different textometric values, which were interpreted as the level of these texts in terms of textual features. The outcome values obtained from the administration of the textometry on any text are suggested to be used as a method of labeling the texts for text linguistic or educational purposes.*

**Key Words:** textometry, text analysis, text linguistics, metatextuality, non-textuality

### **1. Introduction**

Text linguistics is the study of a text as a linguistic product. This focuses on a text in terms of the linguistic criteria which constitute the fundamentals of any text making it meaningful and concrete message. Discourse analysis, text linguistics and pragmatics are the methods used for this purpose. They have differences in their scope of the analysis. Structuralism, outlined by Ferdinand de Saussure (1983), describes language as an analyzable structure, composed of parts that can be defined in relation to others. In the early stages of the linguistic analysis, therefore, it was grammar and structure that determined what to be analysed in a given text. Pragmatics, on the other hand, as Mey (1993) states, differs from structural linguistics in text analysis in that it studies the ways in which context contributes to meaning. Without considering the context, reading any text may cause deviation in meaning, which leads the reader to understand less than the author's intention or to misunderstand. According to Blommaert (2005), moreover, discourse analysis differs from text linguistics in that they take characteristics of persons into account rather than text structure. However, all these approaches contribute to understanding a linguistic material.

What we do in this study is to consider the factors contributing to written language beginning from the author's initial intention to the addressee's final recognition of pragmatic competence which is defined by Chomsky (1980) as the knowledge of how language is related to the situation in which it is used. Halliday and Hassan (1985) regards text linguistics as an analysis of a text in its semio-socio-cultural environment since text and context are so intimately related that neither concept can be comprehended in the absence of the other. We, therefore, will set our study on multi approaches to administrate the linguistic analysis of a textual material. We start the text analysis from the author's possible initial aim before producing the text by referring to his / her biography, world view and life experiences; continue with textual characteristics defined by Beaugrande and Dressler (1981) such as lexical and structural cohesion and coherence, and finally complete with the reader's (the addressee's) possible perception as a final product.

This multi approach is called ‘textometry’ since the text used as data source is finally represented in a numerical figures so that the output results can be listed in a more concrete and meaningful form. Textometer as a method of numerical representation of a text is described and explained with all its components.

Consequently, we will look into the textual material in two principle perspectives in this study, one of which is the surface structure, that is, the analysis of the text in word, phrase or sentence level with cohesive frequencies of referential elements that make individual sentences a text by referring to the following or preceding sentence and the other of which is the deep structure, that is, the analysis of the text in word, phrase or sentence level with cohesive frequencies of symbolic, idiomatic and intertextual elements that take the text beyond the surface meaning.

Our suggestions were administrated on four sample texts of different types and sources. The texts were carefully chosen among different fields, levels and contents. One is *Araby*, a well-known short story in *Dubliners* by James Joyce (1914). It is a well-known work and highly symbolic, which makes the text a good sample to be compared and contrasted with the other the texts under the study. Therefore, the original findings of this material were expected to lead better recognition of the suggestions claimed by this study. Then an academic text and a simple children’s story were administered by textometry to see the comparative results. In addition, composing the randomly produced sentences, we produced a sample text particularly for this study, in order to see how the textometry administration would respond to this artificial text in contrast to the other authentic materials. The findings obtained from the administrations were listed in tables and illustrated in graphs so that they could easily be interpreted. This study comes up with a concrete text value for any written language product, resulting in better classification of texts in common descriptive linguistic features for educational or literary purposes.

## 2. Methodology

The idea of textometer is basically set on Saussure’s theory (1983) based on the paradigmatic and syntagmatic relations of language, which was later represented with two axes by Jakobson (1980): selection and combination. The former is the selective axis, on which we determine which word to use from the lexicon or by which morphemes or auxiliaries they are inflected for person or tense in order to forward the message we intend to address. At this stage, the lexical and morphological preferences are discussed. The latter, on the other hand, is the linear axis, on which we organize the order of words or decide which one is followed by another in a syntactic order. This stage of production reveals the mechanization of the language and the relation between the constituents and the contexts in or out of the text. Figure (Fig. 1) illustrates the two axis of a language as illustrated by Jakobson (1980).

Next, we build up our methodology on Chomsky’s (1972) suggestion that we match sounds and meanings via a computational system present in the human mind that relates meanings to sounds and sounds to meaning. Principles and parameters theory represents the relation between sound and meaning through Phonetic Form (PF) as a sound level and Logical Form (LF) as a meaning level. These two levels of representations are connected to each other by the syntactic structure as shown in Figure 2, the sound-meaning bridge illustrated by V.J. Cook and M. Newson (1996).

However, from now on, we prefer using TF to PF since we deal with the textual form of the language in this study and represent it as shown in Figure 3, adapted from the sound-meaning bridge (V.J. Cook and M. Newson 1996) above (Fig.2).

To understand the relation between the TF and LF, we illustrate the situation with the word *Araby* which is consisted of the letters ‘a’, ‘r’, ‘a’, ‘b’, and ‘y’ and together make up the word *araby* in a linear order, thus, reproducing the image that appears in the addressee’s mind, called “meaning”. However, this image appearing in the mind of the reader cannot be restricted to a single appearance. It may result in non-textual meaning (invalid or an empty image), ambiguous meaning (more than one images), meta-textual meaning (the image beyond the visible - symbolism), or text-dependent meaning (images likely to change depending on the context). If there is a question like “*How are you?*” to a close friend of the character in a text and the friend replies “*Araby*”, then non-textual meaning occurs. If the same question is directed to a friend of the same character, who is described in the text as an intellectual person interested in literature, at this time the same reply will be meaningful for the reader, which means “*he feels as how the protagonist of the Araby feels during the situation he was in*”.

Further, supposing that the question above is directed to an Arabian friend who is in love as “*How are you, Araby?*”, then the addressee can infer more than one meaning, one of which is a racial manner of address and another is a suggestion that it is not worth running after her, resulting in an ambiguity. Consequently, the word *Araby* with the same letters and spelling has different meanings, depending on the contexts, in which it is produced.

Now, what we understand from the examples above is that any text produced in different contexts but with the same content may be decoded in various meanings depending on the intention of the producer, the position of the addressee, the setting, the time and the content of the production. These contexts at TF and LF react with each other to produce a compound meaning, which is sometimes narrower or wider than the *initial intention (II)* of the producer of the text. We will use these language contexts as textual variables in this study. In order to explain the relation between these variables and the text produced, we can illustrate them like the elements undergoing reaction and thus resulting in a multi-bound molecular compound. Each bond represents an individual variable (V), the central element of which is the TF and the LF. Then the illustration of the TF and the LF will be a molecular branching diagram as in Figure 4 and 5. The variables in Fig.4 and Fig.5 are introduced as the elements contributing to the meaning of any text. The number of the variables is illustrated as  $V_n$  since the variables cannot be restricted. The variables to be analyzed are under the limited headings but with unlimited number, depending from one text to another.

## 2. 1. Textometry

*Textometry* is the evaluation of any text in terms of concrete figures in order to categorize it as to the complexity of its texture, considering possible textual and metatextual contexts both in favor of the author and the addressee. In order to achieve this purpose, we should be acquainted with the linguistic approaches not only to the production process of language but also to the description of text. A text is described as an extended structure of syntactic units such as words, phrases, clauses and textual units that are marked by coherence. It is also defined by Halliday and Hassan (1976) as a unit of language in use. Accordingly, the text is not regarded as a grammatical unit, like a clause or a sentence; and it is not defined by its size. It is best regarded as a semantic unit; a unit not of form but of meaning. Fowler, similarly, states that text is made up of sentences, but there exist separate principles of text-construction, beyond the rules for making sentences (Fowler, 1991). According to these opinions on text and language, we tried to develop a systematic text linguistic analysis looking for textual and linguistic features from or even before the production of the material in the author’s mind (II) to the addressee’s *final image (FI)*.

The FI occurring at LF acts as the complementary hypotenuse of the vertical and horizontal dimensions of language suggested by Saussure (1916). Accordingly, the language returns where it originated and the communication is broken off unless the message is received by the addressee. In this case, the hypotenuse, which represents the FI, is not different from the II in that both forms lack communication and they are still in the producer’s mind. If the message is conveyed, however, then the communication occurs and the hypotenuse will be of a certain value, which ranges from the II of the producer to the FI of the addressee. This range may be less or more than intended, or parallel to the II of the producer, or the author for this study. These results are the purpose of the linguistic analysis called as *textometry* in this study. This method of text linguistic analysis includes variables to be assessed and *textometric output* to be obtained. *Variables* are the contexts influencing the meaning and message of the text. They include the textual elements about where and when the text was produced, the author’s background and life experiences, why and how the text was written and to whom it is/was addressed, the structural and lexical cohesion and coherence of the symbolic or biographic expressions in the text, on which we set the variables of textometry. The variables are examined in three categories in the study, each of which is used to illustrate the textual value of the final linguistic product as *non-textual*, *textual* or *metatextual*.

### 2. 1. 1. Non-textual Variables

*Non-textual variables* are the author’s personal background during or before the production period. They are the author’s biography and the author’s world view, which in turn reveals the II. They are non-textual since they are not the lexical component of the text, but they are variables to be considered in the text analysis since they provide the II, or image in the author’s mind. One handicap in this illustration is the question whether it is possible to understand what the author had in his/her mind when he was writing the text. However, it should not be forgotten that we only try to understand the possible II, referring to the author’s biography and world view.

Although it is not possible to know exactly the II, we only try to predict why the author intended to write this text. Whether we know something or nothing does not affect the result of the textometric results since the II is regarded as ineffective variable in the administration of textometry. Non-textual variables are of no frequency value in textometry.

### 2. 1. 2. Textual Variables

*Textual variables* are the structures composing the text. They make up the texture of a text. According to Beaugrande and Dressler (1981), a text will not be communicative without assuring seven standards of textuality: cohesion, coherence, intentionality, acceptability, informativity, situationality and intertextuality. In this study, cohesion and coherent are given priority and the textual variables are based on them. Accordingly, we analyse textual variables by using lexical and structural cohesions in the text. The cohesive elements to play a key role in the textometry should be those having referential matching in the text. Lexical or structural cohesion components of referential value in the textometry include reiterations, collocations, anaphora and cataphora, conjunctions, transitions, subordination, ellipsis, substitution, all of which have referential matching in following or preceding sentences and construct a network between the constituents of text.

### 2. 1. 3. Metatextual Variables

*Metatextual variables* are lexically cohesive components such as symbolic and ambiguous expressions, idioms, the author favor traces in the text, matching with the author's biography, world view and life experiences, the addressee's world view, the addressee's intellectual level or intertextuality. These variables are metatextual since they are not the physical part of the text. However, they influence the meaning of the sentence by taking it beyond the surface meaning. The variables to be analyzed are listed on a *textometric chart* (see Table: 1) in order to make the process easy:

## 2. 2. Textometric Output

The textometry is administered to achieve three outputs from the text. One is the *density*, which roughly shows the number of cohesive elements per sentence in a given text and is used to determine the simplicity or complexity of the text. Another is the *textual value*, which describes the text as non-textual, textual or metatextual and can be used to classify the genre, level and appropriateness of the text. The other is the *semantic deviation*, which aims to reveal the difference or similarity between the II and the FI caused by the text. We will look into the sample expressions or sentences from the textual material according to the variables in the textometer in both structural and lexical point of view, that is, the analysis of the text at TF (on vertical and horizontal axis) and the other is the analysis of the text at LF, at which the author finally aims to forward the message, depending on different variables. Following the analysis of the variables and counting the lexical items under each heading, the data is illustrated in proportional frequencies on the textometric chart shown in **Table 1**. While counting the cohesive elements, it is important not to count the same word repeatedly under different variables. After determining the kind of variable to which it belongs, each word should be counted once. Otherwise, the data to be collected may be misleading. By means of the data processed, we can determine the *textual density*, *textometric value* and *semantic deviation* of the text.

### 2. 2. 1. Textual Density

Roughly, *the frequency of referential components* per a sentence will provide the *textual density* ( $d_t$ ) of any text. The number of sentences is determined by the finite verbs except for the subordinate clauses. So the definition comes out as the frequency of *referential cohesion* ( $c_r$ ) per a finite verb. In order to achieve this value, *the total frequency of referential cohesion* is divided by *the total finite verbs* ( $c_f$ ) except for the finite verbs in the subordinate clauses. The reason why we formulated such an operation is based on the aforementioned linguistic theories (Beaugrande and Dressler 1981 and Halliday and Hassan 1976) on what constitutes a text. Thanks to the textual density, we try to obtain a concrete value showing the frequency of referential elements in the text. This datum provides the researcher to interpret the structural complexity or simplicity of a given text.

$$\text{Textual density} = \frac{\text{Total number of Referential components}}{\text{Net Finite Verbs}}$$

$$d_t = \frac{c_r}{c_f}$$

$d_t$  = textual density,  $c_r$  = referential cohesion,  $c_f$  = cohesive finite verbs (or per sentence)

If the textual density is found between zero and one ( $0 < d_t \leq 1$ ), then the material is non-textual.

### 2. 2. 2. Textometric Value

*Textometric value* is the final target of the textometry administration to find out how much any given written material is text. It is the projection of the FI indicator on the semantic *deviation scale* of the textometer. Since non-textual variables are imaginary and paratextual components, they are of zero frequency in the text. They have author favor effect in textometry and make up the II or 0 on the deviation scale. On the other hand, textual and metatextual variables have the addressee favor effect, which are interpreted as positive values on the scale. In order to find the textometric value of any text, both *the sum of the frequencies of textual and metatextual variables* ( $c_t$ ) and *the number of total words* ( $w$ ) of the text are multiplied by *the textual density* ( $d_t$ ). The reason why we multiply both the dividend and the divider by the textual density is to avoid calculating *textometric value* for a non-textual material having textual density of 0-1. Then, the textometric value will be undefined ( $0 / 0 = \infty$ ), or non-textual. If the textual density is over 0-1, then the equation is simplified and the operation will not affect the textometric value. The proportion of *the total variables* to the *total number of words* is then multiplied by 100 to find the exact percentage of the textometric value:

$$\text{Textometric Value} = \frac{(\text{Textual Variables} + \text{Metatextual Variables}) \cdot \text{Textual Density}}{\text{The number of total words of the text} \cdot \text{Textual Density}} \cdot 100$$

If the textual density is zero (0), then the textual value of the material will be *undefined* ( $\infty$ ), or non-textual. Unless  $d_t$  is not 0, the result will be:

$$t = \frac{c_t \cdot d_t}{w \cdot d_t} \cdot 100 \rightarrow t = \frac{c_t}{w} \cdot 100$$

$t$  = textometric value,  $c_t$  = total textual and metatextual cohesion,  
 $w$  = total words in the text

### 2. 2. 3. Semantic Deviation

The textual relations between the variables make up the meaning, being bound not only to TF but also to LF since there are similar factors contributing the meaning in both forms. Therefore, what the author intends to write may not match with what the reader understands. This difference is called *Semantic Deviation*. Figure 6 demonstrates us that the text analysis starts with the author's intention as the II, which constitutes the fixed vertical axis of the text, follows lexical selection (lexical cohesion) at TF, appearing as syntactic structure (structural cohesion) which constitutes the horizontal axis of the text, and ends at LF as the FI under the effect of several contexts such as the author's biography ( $C_1$ ) and life experiences ( $C_2$ ) and the addressee's intellectual level ( $C_3$ ) or metatextual variables ( $C_4$ ), resulting in the FI, which is the moving part of the scale.

At LF stage, a specific language, say English in this study, is transformed into a universal message or image by a narrative text. This transformation occurs under the influence of several contexts. While the author is active in shaping the TF along with  $Va$  and  $Ha$ , it is the reader who interprets the concrete visual message into an abstract mental one. As Chomsky (1972, p.17) states, language is a particular relationship between sounds (letters) and meaning. Letters are the author's job, whereas meaning is the reader's. The textual and metatextual contents make up the textual body to form the FI, while the non-textual ones make the text author's favor (as shown on Fig.7):

The textometric value is represented on *Semantic Deviation Scale*. If there is an amount of distance between the II and the FI, then it is understood that the written material has a certain degree of textual value. This value is represented on the scale, ranging between 0 and 100. If the distance between the II and the FI overlaps or nearly overlaps (Fig.8), then this means the final image indicator shows the initial place where only the author knows what it is and *non-textual* for the addressee. The message is meaningless and failure. So it is called *non-textual*. It is not a text, but an image in the author's mind. Only does the author know what it is, or it is only a group of words which are not composed purposefully.

In contrast, if there is an amount of distance between the initial intention of the author and the final image of the reader, then it is understood that the written material has a certain degree of textual value, ranging between 0 and 100, whereas non-textuality is represented as 0 (Fig.9). When the projection of FI ranges between 0 and 50 on the semantic deviation scale, then the text is called *textual* and the value which the FI demonstrates between 0 and 50 will be interpreted as the textual value, depending on the simplicity of the textual density, cohesion, coherence, structure and genre of the text. However, because it is almost impossible to learn the II of the author, or the FI of an individual reader, the FI cannot be interpreted exactly at 50 or at 100, which represent the full interpretation of the text. Textual value starts after 0, and extends up to 50 representing the ideal parallelism between II and FI. As the FI becomes parallel or almost parallel to II ( $FI < 50$ ), then it becomes similar to II, that is, the text can be understood by the addressee as much close as the author's initial intention (Fig.9).

When the projection of FI ranges between 50 and 100 on the deviation scale, then the text is called *metatextual* and the value to which the FI correspond between 50 and 100 will be interpreted as the metatextual value. The textual value is represented on the deviation scale from 50 up to 100, representing the ideal understanding of metatextual meanings such as symbolism, metaphors, irony and other literary purposes. As the FI gets nearer to 100 ( $50 < FI < 100$ ), then it means that the text exposes the reader to many metatextual interferences, taking the meaning from the surface to deeper. The text is interpreted by the researcher to be highly metatextual, that is, it cannot be understood by the addressee unless the necessary intellectual level is achieved and metatextual contexts are considered by the addressee (Fig.10).

### 3. Findings and Discussions

In this part of the study, four different types of texts are analyzed by textometry and their textometric results are illustrated with tables and figures to make the data more comprehensible and concrete. Then, the findings are compared and contrasted in order to discuss the utility of the textometry.

#### 3. 1. Textometric Output of “Araby”

Considered one of James Joyce's best known short stories, “Araby” is the third story in his short fiction collection, *Dubliners*, which was published in 1914. Critical interest in the story has remained intense in recent decades as each story in *Dubliners* has been closely examined within the context of the volume and as an individual narrative. The story is composed of 144 sentences and 2332 individual words. Below are results of the textometric analysis of Araby. Initially, the results are shown on the Textometric Chart (Table 2). Accordingly, the textual density, the textometric value and the semantic deviation are listed respectively.

$$3. 1. 1. \text{ Textual density of Araby} = \frac{\text{Total Referential Cohesions}}{\text{Net Finite Verbs}}$$

$$d_t = \frac{c_r}{c_f} = \frac{969}{220} = 4,4$$

#### 3. 1. 2. Textometric Value of Araby

$$\text{Textometric Value} = \frac{(\text{Textual Variables} + \text{Metatextual Variables})}{\text{The number of total words of the text}} \cdot 100$$

$$t = \frac{c_t}{w} \cdot 100 = \frac{969+411}{2332} \cdot 100 = \frac{1380}{2332} \cdot 100 = 59$$

#### 3. 1. 3. Semantic Deviation of Araby

Then the resulting value is illustrated on the deviation scale of the textometer as in Figure 11. The text is found to be metatextual. The addressee should know the author's biography and other intertextual materials in order to understand the text in parallel to the author's intention.

#### 3. 2. Textometric Output of “Fear of Flying”

This text is cited from an English proficiency exam (UDS, 2012). It is an academic text on medicine. It is composed of 10 sentences and 143 individual words (Table 3).

### 3. 2. 1. Textual Density

$$\text{Textual density} = \frac{\text{Total Referential cohesions}}{\text{Net Finite Verbs}}$$

$$d_t = \frac{c_r}{c_f} = \frac{50}{12} = 4,1$$

The textual density of this text is seen 4,1. The textual elements are less intense than in *Araby* consistent with the expectations from a literary short story.

### 3. 2. 2. Textometric Value

$$\text{Textometric Value} = \frac{(\text{Textual Variables} + \text{Metatextual Variables})}{\text{The number of total words of the text}} \cdot 100$$

$$t = \frac{c_t}{w} \cdot 100 = \frac{50}{143} \cdot 100 = 34$$

The textometric value of “Fear of Flying” is relatively much less than *Araby* since it is a scientific text and does not involve metatextual expressions. Therefore, it has an expected textual value of 34, which is near ideal textual value of 50.

### 3. 2. 3. Semantic Deviation

Then the resulting value is illustrated on the deviation scale of the textometer as shown in Figure 12. According to the semantic deviation scale, the text can almost be understood as much as the initial intention of medical description providing the addressee is interested in medical or psychological subjects.

## 3. 3. Textometric Output of “The Ant and the Grasshopper”

This text is a narrative tale. It is a fable for children and composed of 11 sentences and 134 individual words (Table 4).

### 3. 3. 1. Textual Density

$$\text{Textual density} = \frac{\text{Total Referential cohesions}}{\text{Net Finite Verbs}}$$

$$d_t = \frac{c_r}{c_f} = \frac{38}{12} = 3,1$$

The textual density of the text is seen 3,1. The texture is less intense than *Araby* and the *Fear of Flying*. That the structure of the children’s tale is not as dense as the other two texts is thought to result from the addressee of the text since it is written for young children. The textometric results confirm our suggestion.

### 3. 3. 2. Textometric Value

$$\text{Textometric Value} = \frac{(\text{Textual Variables} + \text{Metatextual Variables})}{\text{The number of total words of the text}} \cdot 100$$

$$t = \frac{c_t}{w} \cdot 100 = \frac{44}{134} \cdot 100 = 32$$

The textometric value of *The ant and the grasshopper* is relatively less than the other texts since it is fable for teaching moral values to children. Although structural elements are simple, it involves metatextual symbolic expressions. Therefore, it has metatextual value of 32, which shows that the text is simple and can be understood in parallel to the author’s intention.

### 3. 3. 3. Semantic Deviation

Then the resulting value is illustrated on the deviation scale of the textometer as shown in Figure 13.

According to the semantic deviation scale, the text can almost be understood as much as the initial intention of teaching moral values for young children.

### 3. 4. Textometric Output of “Disharmony”

This text is our own production, composed of randomly selected sentences, each of which is about different subjects from different academic texts and titled as “disharmony”. It is composed of 11 sentences and 139 individual words (Table 5). The text is analyzed by the administration of textometry and then compared and contrasted with the other texts studied by the method of textometry to observe the differences between the results.

#### 3. 4. 1. Textual Density

$$\text{Textual density} = \frac{\text{Total Referential cohesions}}{\text{Net Finite Verbs}}$$

$$d_t = \frac{c_r}{c_f} = \frac{1}{11} = 0.09 \cong 0$$

The textual density of the text is seen 0, 09. The result shows that there is no texture in the text. The textometric results confirm our suggestion that referential cohesion determines the texture of any text and thus its value.

#### 3. 4. 2. Textometric Value

According to the textometry, if the textual density is less than “0,1”, then the material is non-textual and it does not have any textual value because:

$$\text{Textometric Value} = \frac{(\text{Textual Variables} + \text{Metatextual Variables}) \cdot d_t}{\text{The number of total words of the text} \cdot d_t} \cdot 100$$

$$t = \frac{c_t \cdot d_t}{w \cdot d_t} \cdot 100 = \frac{20.0}{139.0} \cdot 100 = \infty$$

Since the textual density is almost 0, the textual value of the text is undefined, which means non-textual.

#### 3. 4. 3. Semantic Deviation

Then the resulting value is illustrated on the deviation scale of the textometer as shown in Figure 14. According to the semantic deviation scale, the text is non-textual. That is, the message cannot be interpreted by the addressee. The message of the author is still an II and cannot be transformed into FI. Final image does not occur since the message of the author is not understood by the addressee. The findings revealed that the cohesion of the referential structures and collocations are the most important factors to determine whether any written material is a text. The cohesive frequencies of the structural or lexical items which are not matching with the following or preceding sentences in the text are unnecessary to be considered in text linguistics analysis. Therefore, the quality of the elements to be numbered is more important than the quantity.

### 4. Conclusion

In this study, we tried to administrate a systematic linguistic text analysis method based on the principle theories of Saussure (1983), Jakobson (1980) and Chomsky (1993), using an evaluation scale (textometer) for textual products. The textometer is developed to achieve three outputs from the text. One is the *textual density*, which roughly shows the number of cohesive elements per sentence in a given text and is used to determine the simplicity or complexicity of the text. Another is the *textometric value*, which describes the text as non-textual, textual or metatextual and can be used to classify the genre, level and appropriateness of the text. The other is the *semantic deviation*, which aims to reveal the possible difference or similarity between the possible initial image of the author and the final image of the addressee invoked by the text. We demonstrated the theoretical approach on four different types of texts in order to see how the textometry would response to different types of texts. One of the materials analyzed in this study was a well-known short story, *Araby* by Joyce (1914). The text was found metatextual, with relatively high density and textual value. The semantic deviation was meaningful, since the FI deviates from the II in that it is highly symbolic. Another was an academic text, *Feeling of Fear (OSYM, 2012)*, prepared for the language proficiency exams. It was found to be textual as expected since it is a plain academic text. The density of the text was relatively high although the textual value was found relatively low.



The inverse proportion was explained by the genre of the text. Since it is a well-organized academic text, it includes structural cohesive elements such as conjunctions or transitions. This increased the density of the text. However, since the scarcity of the other textual and metatextual cohesive variables, the text was found relatively simple. Next, a fable called *The ant and the grasshopper* was for young children; and had a simple structure, which interested us in terms of textometry. The textometric density and the value were relatively less than the other texts since it was a fable addressed to young children. The semantic deviation in those two texts made us confused since the former is a plain explanatory text and the latter is a simple children's tale. The results show that there is a slight deviation from the II, but this deviation is not the one the author initially intended since the ideal parallelism is around 50. The textometer lacks the ability to analyze the semantic deviation. The other one, *Disharmony*, was an artificial text made up of randomly selected sentences from different texts. The sentences of the text were irrelevant, which was deliberately organized to see the reaction of the textometry when compared with the reaction to other original texts. It was found to be a non-textual according to the textometric results. The principle reason was that the text does not involve any cohesive or coherent elements, making up the fundamentals of any text. The textometric value was undefined, which shows that the text is non-textual and since there is no any FI occurring in favor of the addressee, there is also no semantic deviation between the II and the FI. The findings obtained during the study were also introduced and illustrated in tables and figures, followed by the discussion of each datum severally.

*The textometer* provided a practical method and more objective and concrete point of view for text linguistic analysis. Thanks to the textometry, the texture of any text could be determined and interpreted in figures, which may result in the classification of texts according to their complexity and structure. It was also found out that the density or complexity of a text may not be direct proportional to the textual value. While one is relatively high, the other may be relatively low. However, while the textometer is successful in determining the simplicity or complexity and non-textuality, textuality or metatextuality of a text, it fails in interpreting the semantic deviation from the initial intention of the author to the final image of the addressee. A text may be simple but successful in conveying the message. Accordingly, all simple texts should be ranged close to the textual value of 50, which represents the ideal parallelism between those two images. Consequently, *semantic deviation* concept cannot be relied on until developed by further studies focusing on the problem.

The outcome values obtained from the administration of the textometry on any text are suggested to be used as a method of labelling the texts in literary textbooks or course books of different classes. They may also be used as more concrete labeling criteria for the novels or simplified story books prepared for language learners. Moreover, the textometric value may also be used as a text linguistic criterion to label the texts under study.

## References

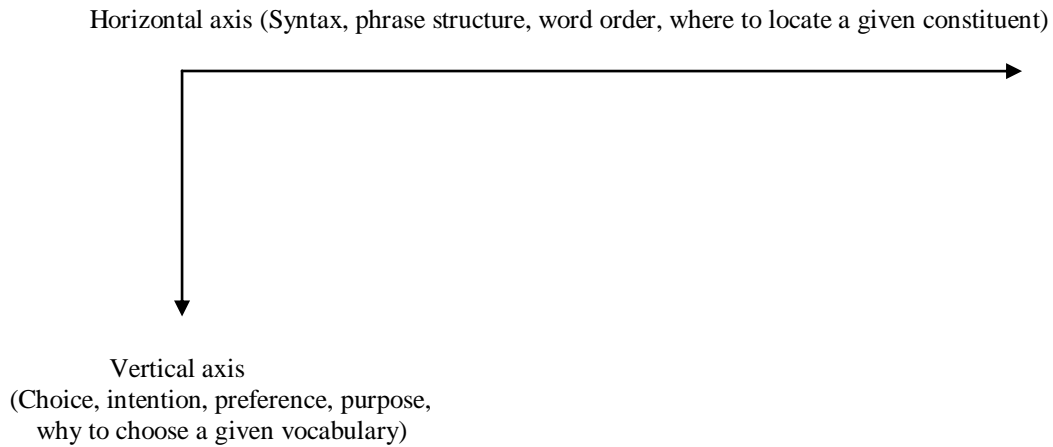
- Beaugrande, Robert and Dressler, Wolfgang U. 1981. *Introduction to text linguistics*. Robert-Alain De Beaugrande, Wolfgang Ulrich Dressler. London .New York : Longman.
- Blommaert, Jan. 2005 .*Discourse*. Cambridge: Cambridge University Press.
- Cook, Vivian J. and Newson, Mark. 1996. *Chomsky's Universal Grammar: An Introduction*. Oxford: Blackwell.
- Chomsky, Noam. 1972. *Language and mind*. New York: Harcourt Brace Jovanovich.
- Chomsky, Noam. 1980. *Rules and Representations*. Oxford: Basil Blackwell.
- Fowler, Roger. 1991. *Language in the News: Discourse and Ideology in the Press*. London/ New York: Routledge.
- Halliday, Michael A. K. and Hassan, Ruqaiya. 1976. *Cohesion in English*. London: Longman.
- Halliday, Michael A. K. and Hassan, Ruqaiya. 1985. *Language, Context, and Text: Aspects of Language in a Social-Semiotic Perspective*. Geelong: Deakin University.
- Jakobson, R. (1980). "Two Aspects of Language and Two Types of Aphasic Disturbances." In Jakobson and Halle, *Fundamentals of Language*. 1956. 4th ed. The Hague: Mouton.
- Joyce, James. 1914. *Dubliners*. London: Grant Richards.
- Mey, Jacob L. 1993. *Pragmatics: An Introduction*. Oxford: Blackwell.
- Saussure, F. De. ([1916] 1983): *Course in General Linguistics* (trans. Roy Harris). London: Duckworth.
- [http://www.kidsgen.com/fables\\_and\\_fairytales/fables.htm](http://www.kidsgen.com/fables_and_fairytales/fables.htm)
- <http://www.osym.gov.tr/belge/1-14127/2011/ UDS/. sinav-arsivi.html>

**Abbreviations**

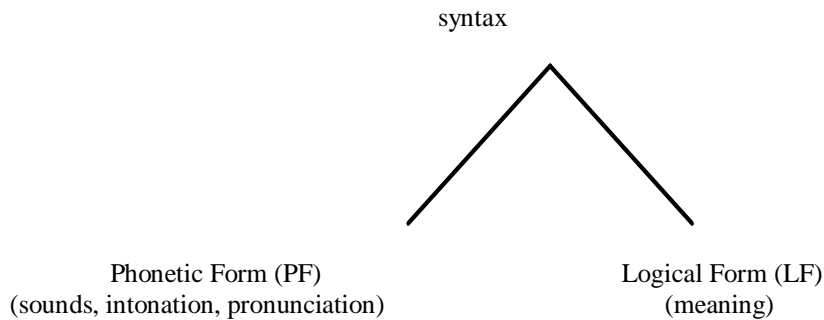
- $c_f$**  Cohesive finite verbs (or per sentence)
- $c_r$**  Referential cohesion
- $c_t$**  Total textual and metatextual cohesion
- $d_t$**  Textual density
- FI** Final Image
- II** Initial Image
- LF** Logical Form
- PF** Phonetic Form
- $t$**  Textometric value
- TF** Textual Form
- V** Variable
- $w$**  Total words in the text

**Figures**

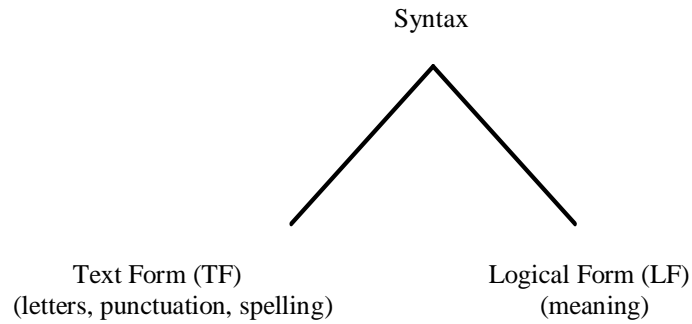
**Figure 1:** Jakobson’s two axis of language



**Figure 2:** The bridge between Phonetic Form and Logical Form

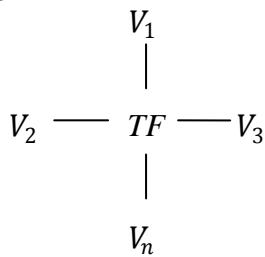


**Figure 3:** The relation between Text Form and Logical Form

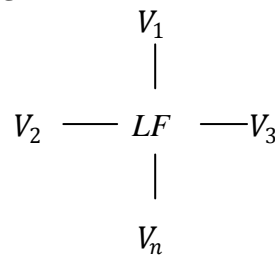


**Figure 4 and 5:** Molecular Models

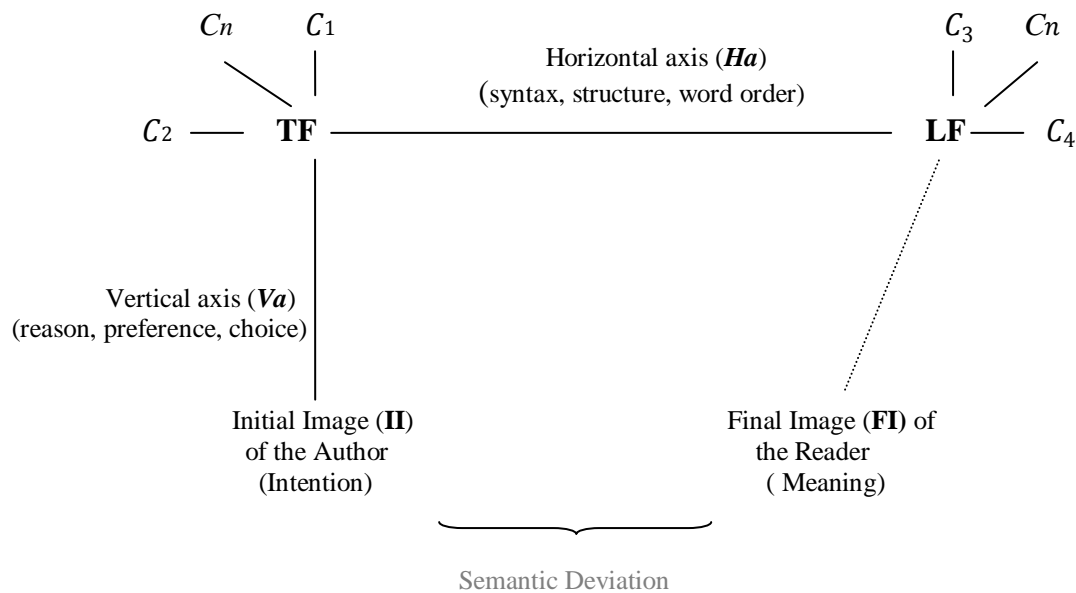
**Figure 4** Molecular Model for TF



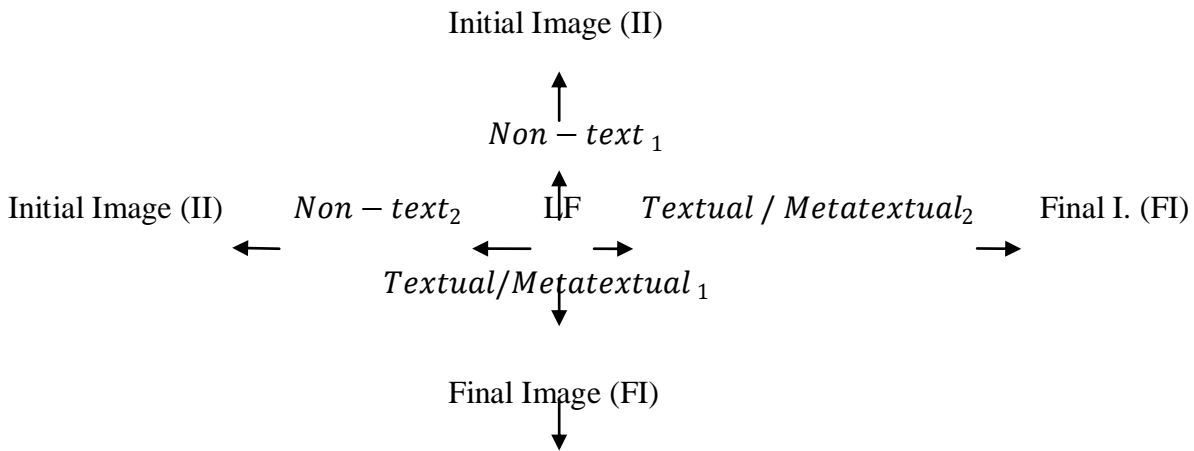
**Figure 5** Molecular Model for LF



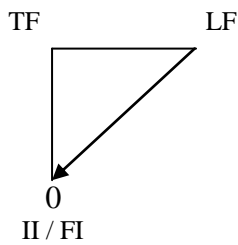
**Figure 6:** Semantic Deviation



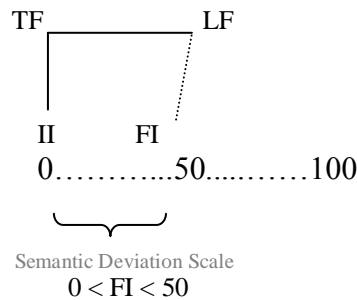
**Figure 7: Model for LF**



**Figure 8: II and FI overlap (Non-textual ≤ 0)**



**Figure 9: The textual semantic deviation between II and FI (0 < Textual < 50)**



**Figure 10: The metatextual semantic deviation between II and FI (50 < Metatextual < 100)**

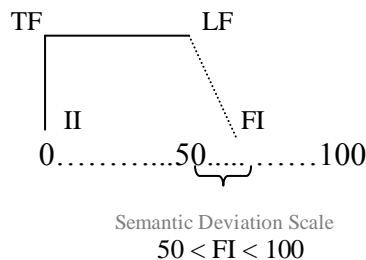


Figure 11: The final Textometric illustration and the Semantic Deviation of Araby

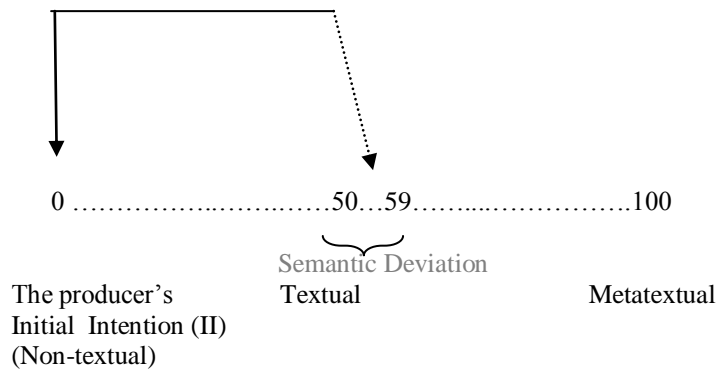


Figure 12: The final Textometric illustration and the Semantic Deviation of “Fear of Flying”

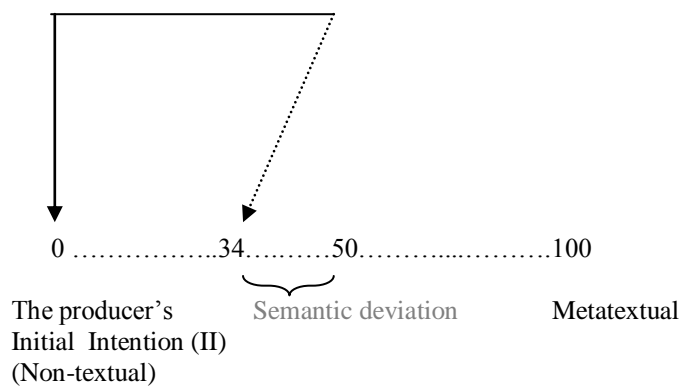
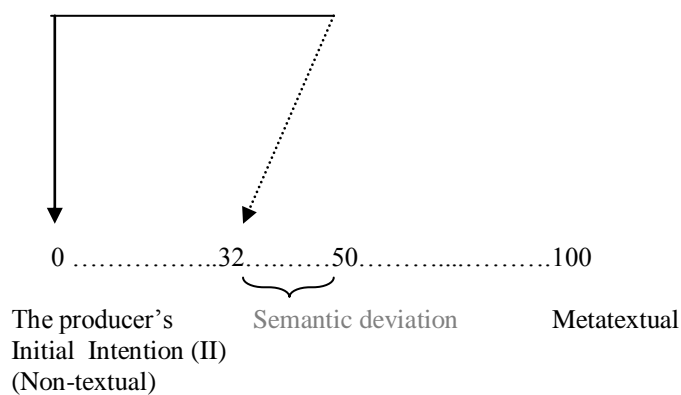
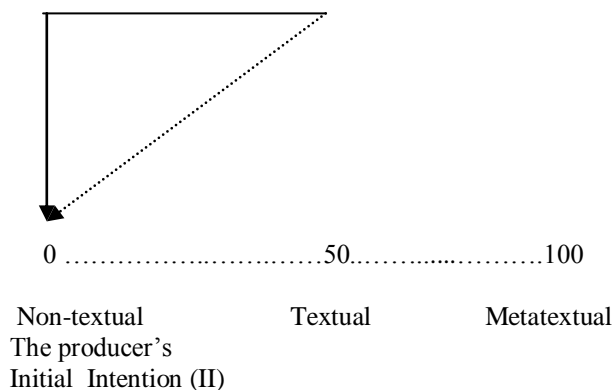


Figure 13: The final Textometric illustration and the Semantic Deviation of “The ant and the grasshopper”



**Figure 14:** The final Textometric illustration and the Semantic Deviation of “Disharmony”



**Tables**

**Table 1:** Textometric Chart

<b>Overall lexical and structural components (Total Number of Words)</b>	
<b>Net Finite Verbs</b>	
<b>Non-textual Variables:</b>	<b>Frequency</b>
The producer's effect: - The author's initial intention - The author's biography - The author's world view	<b>0</b>
- The addressee's effect - The addressee's intellectual/educational level - The addressee's world view	<b>0</b>
<b>Textual Variables:</b>	<b>Frequency</b>
Referential Cohesion:	
- Conjunctions / Transitions:	
- Subordination:	
- Ellipsis:	
- Substitution:	
- Anaphora:	
- Cataphora:	
- Reiteration:	
- Collocation	
<b>Metatextual Variables:</b>	<b>Frequency</b>
Symbolism	
Idioms	
Intertextuality	

**Table 2:** Textometric Chart of Araby

<b>Overall lexical and structural componenets (Total Number of Words)</b>	<b>2332</b>
<b>Net Finite Verbs</b>	<b>220</b>
<b>Non-textual Variables:</b>	<b>Frequency</b>
The producer's effect: - The author's initial intention - The author's biography - The author's world view	<b>0</b>
- The addressee's effect - The addressee's intellectual/educational level - The addressee's world view	<b>0</b>
<b>Textual Variables:</b>	<b>Frequency</b>
Referential Cohesion:	
- Conjunctions / Transitions:	15
- Subordination:	55
- Ellipsis:	9
- Substitution:	3
- Anaphora / Cataphora & Pronouns	570
- Reiteration:	street :14 bazaar /Araby :14
- Collocation	dark,blind, quiet : 58 Religion/ Christianity : 24 the beloved : 47 (she, daughter, sister) Pessimism : 160
<b>Metatextual Variables:</b>	<b>Frequency</b>
Symbolism	397
Idioms	9
Intertextuality	5

**Table 3:** Textometric Chart of "Fear of Flying"

<b>Overall lexical and structural componenets (Total Number of Words)</b>	<b>143</b>
<b>Net Finite Verbs</b>	<b>12</b>
<b>Non-textual Variables:</b>	<b>0</b>
The producer's effect: - The author's initial intention - The author's biography - The author's world view	<i>The text was written for an academic purpose to describe a psychological condition.</i>
- The addressee's effect - The addressee's intellectual/educational level - The addressee's world view	<b>0</b>
<b>Textual Variables:</b>	<b>Frequency</b>
Referential Cohesion:	
- Conjunctions / Transitions:	2
- Subordination:	5
- Ellipsis:	0
- Substitution:	0
- Anaphora:	6
- Cataphora:	3
- Reiteration:	<b>fear :6</b> <b>flying :4</b>
- Collocation	<b>fear : 15</b> <b>flying : 9</b>
<b>Metatextual Variables:</b>	<b>Frequency</b>
Symbolism	0
Idioms	0
Intertextuality	0

**Table 4:** Textometric Chart of “*The ant and the grasshopper*”

<b>Overall lexical and structural components (Total Number of Words)</b>	<b>134</b>
<b>Net Finite Verbs</b>	<b>12</b>
<b>Non-textual Variables:</b>	<b>0</b>
The producer’s effect: - The author’s initial intention - The author’s biography - The author’s world view	<i>The text was written for elementary school children of the age 6 to teach expected educational behaviours and moral.</i>
- The addressee’s effect - The addressee’s intellectual/educational level - The addressee’s world view	<b>0</b>
<b>Textual Variables:</b>	<b>Frequency</b>
Referential Cohesion:	
- Conjunctions / Transitions:	<b>3</b>
- Subordination:	<b>4</b>
- Ellipsis:	<b>1</b>
- Substitution:	<b>0</b>
- Anaphora:	<b>7</b>
- Cataphora:	<b>1</b>
- Reiteration:	<b>winter :3 food :2</b>
- Collocation	<b>winter :6 food :11</b>
<b>Metatextual Variables:</b>	<b>Frequency</b>
Symbolism	<b>6</b>
Idioms	<b>0</b>
Intertextuality	<b>0</b>

**Table 5:** Textometric Chart of “*Disharmony*”

<b>Overall lexical and structural components (Total Number of Words)</b>	<b>139</b>
<b>Net Finite Verbs</b>	<b>11</b>
<b>Non-textual Variables:</b>	<b>0</b>
The producer’s effect: - The author’s initial intention - The author’s biography - The author’s world view	<i>The text was written for elementary school children of the age 6 to teach expected educational behaviours and moral.</i>
- The addressee’s effect - The addressee’s intellectual/educational level -The addressee’s world view	<b>0</b>
<b>Textual Variables:</b>	<b>Frequency</b>
Referential Cohesion:	
- Conjunctions / Transitions:	<b>0</b>
- Subordination:	<b>1</b>
- Ellipsis:	<b>0</b>
- Substitution:	<b>0</b>
- Anaphora:	<b>0</b>
- Cataphora:	<b>0</b>
- Reiteration:	<b>0</b>
- Collocation	<b>0</b>
<b>Metatextual Variables:</b>	<b>Frequency</b>
Symbolism	<b>0</b>
Idioms	<b>0</b>
Intertextuality	<b>0</b>