A Study on Human Perception of Document Paragraph Layout using Surveys*

Benjamin Aziz† and Aysha Bukhelli‡

†School of Creative and Digital Industries
Buckinghamshire New University
High Wycombe HP11 2JZ
United Kingdom
‡Office of the Prime Minister
Bahrain

Abstract

We identify a list of characteristics of a text document that we use as the basis for a study on human perception of what the structure of a document should look like. Our study reveals that the native language dimension has a significant impact on the human perception of some such characteristics, including the average number of paragraphs starting with a verb, as well as other characteristics describing average paragraph proportions. The study results also show what the mean values are, for the different document characteristics, and their distribution across the native language dimension, therefore providing some idea of what a normalised structure of a document should look like. The results of the study have a direct application in a new method for embedding secret messages in text documents that has been recently proposed by the authors and which uses manipulations in the paragraph layout of a document.

1   Introduction

The activity of reading a textual document, for example, a scientific paper, a novel, a prose, a newspaper column or a magazine article is often accompanied subconsciously by a perception the human mind forms in relation to the manner in which the document should be presented in terms of its structure and visual characteristics. In fact, it was shown in [1] that text characteristics like word frequency, syntactic complexity and signals aiding comprehension have direct impact on the cognitive capacity of the human mind to process information

---

*A preliminary version of the research reported herein was presented at the 14th International Conference on Information, Intelligence, Systems and Applications (IISA2023), Volos, Greece, July 10-12, 2023.

1

while reading, and therefore, aid in the learning process itself. One form of such learning processes we are interested in here, is relevant to the task of inspecting text in search of hidden secret messages that could have been embedded in that text using the well-known field of textual steganography.

Our study presented in this paper, therefore, is motivated by the need to understand human perception of the structure of a document, particularly in the context of paragraph characteristics (e.g. size and number of paragraphs and their lexical elements) in English language texts. In [2], we had presented a steganographic embedding method that relies on the changing of paragraph sizes as a means of encoding secret bits of information. However, missing from [2] was the understanding of whether such changes have any impact on human perception of the document, since one way of attacking the embedding method in [2] is through human eye examination of the structure of the document and its characteristics. For example, if the human perception was that a single page text excerpt should have the majority of words centered in the middle of the text (i.e. middle paragraphs), then manipulating this distribution of words (e.g. by shifting words towards the opening or the closing paragraphs) for the purpose of encoding secret bits, would certainly be vulnerable to human eye detection. The purpose of this study, therefore, is to understand if such "perception" indeed does exist in the human mind, and whether it depends on common factors related to human native language, age and level of education. We identify the characteristics of the textual document in terms of a number of *data points*. Our study reveals what the mean values of such data points could be, based on the results of a survey conducted for over two hundred participants, and therefore providing some understanding of what *normal* text structure characteristics might be. An earlier version of this paper was published in [3], and this paper adds more detail on and analysis of the results of the experiment carried out in the study, that the original paper did not present due to lack of space.

The rest of the paper is structured as follows. In Section 2, we discuss some of the most related works in literature. In Section 3, we discuss the design and implementation of our study survey. In Section 4, we present some of the results of the survey, mainly, along the native language dimension. In Section 5, we analyse the results and interpret their meaning. Finally, in Section 6, we conclude the paper giving directions for future work.

2   Related Work

The most obvious and noticeable errors in documents that a human eye can detect are usually grammatical, punctuation-related or spelling mistakes in written documents. If used with any typical text editor, the majority of such errors are detected and highlighted with relative ease, rendering the document suspicious if these are used as a means of hiding secret messages. Paragraph size and structure manipulation, on the other hand, is not so easily detectable by the human eye or any standard text editing application that highlights errors, and hence such manipulation is not usually detectable. Nonetheless, the genre of

the document and the author's writing style might suggest some traits. For instance, romantic novels harness long descriptive paragraphs because of the minimal presence of dialogue between the novel's characters and the extensive presence of the description of feelings and atmosphere.

We review in this section a few works that demonstrate how the problem of paragraph and text structure manipulation is perceived through the human mind. We also cover linguistic error analysis in general, and discuss the most recognisable errors in written English, highlighting literature that attempts to pinpoint the sources of such errors.

2.1 Paragraph Strucutre

Padučeva [4] states that: "Just as not every combination of words constitutes a well-formed sentence, not every combination of well-formed sentences constitutes a well-formed paragraph". Limitations on the cohesiveness of the sentences of a paragraph may be associated with the distinction between a coherent text and a meaningless accumulation of phrases, or with the difference between good and bad writing styles or they may simply be of an entirely different nature. Padučeva observes that the beginning of a new paragraph must have a phrase in which the primary name is either not dominated at all or is dominated by a word, which is not from the last two or three phrases but from some earlier one.

Braddock [5] did not believe that all paragraphs 'must' have a topic sentence, and that topic sentences, if they exist, are at the beginning of the paragraph. Contrary to that, Williams and Stevens [6] define the structure of a paragraph to contain four main key parts: topic sentence, development, example and summary. On the other hand, Christensen [7] assumes the paragraph structure to consist of a sequence of structurally related sentences. The top sentence of the sequence is the topic sentence. The topic sentence is nearly always the first sentence of the sequence. Although Christensen subsequently allows for exceptions, stating that some paragraphs have no topic sentences, there is no mistaking that Christensen's second and third "rules" are essentially those which Braddock found to be false. Unlike Braddock, Christensen seems to believe that the term topic sentence is self-explanatory, requiring no precise definition. In support of his claims, Christensen cites that "many scores of paragraphs I have analysed for this study" [7, 5]. He does not state though how these paragraphs were selected or from where. He states only that in the paragraphs he analysed "the topic sentence occurs almost invariably at the beginning." Had he detailed his procedures as he did in his study of sentence openers, we would have reason to be more confident of his conclusions. But he fails in that respect.

Similarly to Christensen, Becker in [8] applies to the paragraph the instruments of sentence analysis, with the purpose of "extending grammatical theories now used in analysing and describing sentence structure to the description of paragraphs". Becker states that "bad" paragraphs, contain poorly constructed and confusing sentences while "good" paragraphs begin with a topic sentence and develop the idea stated by that sentence. Williams [6] explains that an expository writer generally supplies an introductory paragraph stating general

content of the piece, a number of center paragraphs supplying details or examples to support the general idea and a final paragraph generally summarising the material. Within each paragraph, a specific structure is recognisable driven by the main idea and supported by detail.

Hinds [9] discussed that professional writing is controlled by paragraph structure. His analysis centred primarily on a particular type of discourse newspaper and magazine articles. This type of discourse may be characterised by several significant features. First, it is informative; its primary purpose is to convey unknown information to the reader. Second, a relatively short newspaper or magazine article generally consists of only one paragraph. That is, the article has a single theme, and this theme is often, although not always, stated in the headline that accompanies the article. Third, newspaper and magazine articles differ significantly from unmonitored spontaneous speech in that they are written and revised, and then checked by a copy editor. This ensures a more formal structure than in spontaneous conversations, which may be modified immediately according to the reactions, or lack thereof, on the part of the audience. According to Hinds, a paragraph consists of an indeterminate number of sentences or sentence fragments. Although a paragraph is 'about' only one topic, it may be divided into a number of segments, the sentences of which are related more closely to one another than to other sentences in the paragraph. These segments are developments of the paragraph topic and exist in a specified number of relationships to one another. Specifically, the initial segment of a paragraph, which is the most important segment, is termed the introductory segment. The other types of segments have the purpose of offering a motivation, a highlight or an unexpected twist. Within each one of these, there will be one and only one, sentence of particular importance, termed the *peak* sentence. It is within this peak sentence that a full noun phrase occurs, while it is within non-peak sentences that pronouns occur. One method of achieving this organisation involves pronominalization. The choice of full noun phrase or pronoun thus contributes to the organisation of paragraphs. In essence, a full noun phrase is used to indicate semantically prominent information, while a pronoun is used to indicate information that is less prominent semantically. Such pronormalisation encourages paragraph boundaries. Hinds however concludes that the author is free to organise information in any way they feel will highlight or dramatise the points they wish to convey to the reader.

2.2 Error Analysis in English

There is plenty of literature that covers error analysis in written English [10]. According to a study by Darus and Subramaniam [11], after examining errors in a corpus of 72 essays written by students (37 male and 35 female) studying English as a second language at level 4 in a secondary school in Malaysia, they found that the six most common errors made were singular/plural form, verb tense, word choice, preposition, subject-verb agreement and word order [11].

Sermsook [12] observed 17 types of errors found in English sentences written by Thai students studying English as a Foreign Language (EFL). The partic-

ipants in the study were, 26 second year English major students in a Thai university. They were two males and 24 females, whose ages ranged between 20 and 22 years old. All of whom had learned EFL for at least seven years. The errors at the sentence level comprised of; punctuation, subject-verb agreement, capitalisation, fragments, tenses and word order. Errors at the word level included wrong articles, nouns, pronouns, verbs, prepositions, adjectives, literal translation from the Thai language, parts of speech, word choices, spelling and transition words. Punctuation errors ranked as the highest occurrence at the sentence level and Word order errors ranked with the lowest level of occurrence. On a word level, errors in articles ranked the highest, whilst errors in transition words ranked the lowest.

The perception of errors differs between native speakers and foreign speakers, even people who studied the language. For instance, Hughes and Lascaratou [13] submitted 32 sentences containing an error and 4 without any error to 30 judges, 10 Greek native speaking teachers of English, 10 English native speaking teachers of English and 10 native speaking non-teachers. One of the error-free test sentences was, *"neither of us feels happy"*, and this was judged to be erroneous by two of the Greek teachers, 3 of the English native speaking teachers and by no fewer than 5 of the 10 non-teachers. Noting the wide variation in detection rates, the opposite can also happen. People can overlook errors where they should be obvious. For example, Hughes and Lascaratou's test included the sentence *"the boy went off in a faint"* (which is mentioned in the Oxford advanced learner's dictionary of current English) notwithstanding, 9 of the English native speaking teacher and non-teacher group members found this sentence to be erroneous [13]. Another relevant study is by Lennon [14] after having transcribed the second language (L2) English speech for German advanced learners of English, he was able to identify 568 clear errors and there were also 208 doubtful ones. These he submitted to a panel of 6 educated natural speakers of English and found little agreement, for all 6 rejected as ungrammatical 103 of the 208, 2 rejected 53 and 4 rejected 22.

In error detection, as we pointed out earlier, no more than reasonably firm yes or no decisions are called for. Studies use sentences as their unit of analysis and ask their participants to report their intuition. Error analysis has to be more demanding than this, however, as additional questions are asked about the putative detected error. Not all errors are easily detectable in this way, if someone diffused an error throughout the sentence or larger unit of text then this would be known as global errors [15]. The sentence does not simply contain an error, it is erroneous or flawed as a sentence. Secondly, it is often difficult to consistently locate the error and what the learners have said or written and what they should have written.

Burt and Kiparsky [15] suggested that we should identify errors by reference to the Target Language, according to what the person who says has to learn about English, that is, according to a rule, which he has been violated. In other words, identification of errors depends on what the person has learnt previously or how they were educated about the Target Language. This will the reference of errors against which they compare new texts.

5

2.3 Classification of Errors

As stated by Dulay, Burt, and Krashen [16], errors in written English are classified according to their features, which can be divided to six different categories: omission of grammatical morphemes, double marking of semantic features, use of irregular rules, use of wrong word forms, alternating use of two or more forms and disordering. In the late 1990s, James [17] proposed five categories of errors, which include grammatical errors (adjectives, adverbs, articles, nouns, possession, pronouns, prepositions and verbs), substance errors (capitalization, punctuation and spelling), lexical errors (word formation and word selection), syntactic errors (coordination/ subordination, sentence structure and ordering), and semantic errors (ambiguous communication and miscommunication).

In another study by Hengwichitkul [18], errors were analysed at the sentential level. All the errors were classified as subject-verb agreement, tenses, parts of speech, participial phrases, relative clauses, passive voice, parallel structure, punctuation, run-ons and fragments. Likewise, Runkati [19] categorised the errors found in their study into two main types. The first type dealt with errors at the sentential level, which were fragments, such as: run-ons, subject-verb agreement, word order, tenses, capital letters and punctuation. The second one was errors at the word level, such as: articles, prepositions, word choices, nouns and numbers.

Sermsook's [12] study focused on errors in English sentences and the analysis of errors found at a sentential level and word level. Sentential level errors included fragments, subject-verb agreement, word orders, tenses, capitalisation and punctuation. Errors in the word level were articles, prepositions, word choices, nouns, pronouns and verbs. Other kinds of analysis, for example, addition and omission were also referred to as sub-categories of the sentential level errors and the word level errors.

2.4 Sources of Errors

Richards [20], for instance, states that the two major sources of errors in written English are inter-lingual errors and intra-lingual errors. The former refers to errors caused when learners wrongly use the rules of their first language when they produce sentences of the target language. The latter, on the other hand, contains errors caused during the learner's language learning process. Such errors include over-generalisation, false analogy and so on.

James [21] proposes four sources of errors: inter-lingual errors, intra-lingual errors, communication strategy-based errors and induced errors. Based on their study, Penny [22] concludes that there are two major sources of errors: inter-lingual transfer and intra-lingual transfer errors. Likewise, Heydari and Bagheri [23] also state that inter-lingual interference and intra-lingual interference are the two sources of errors committed by EFL learners.

In Thailand, a considerable number of scholars have carried out studies to explore sources of errors in English documents. Kaweera [24], Runkati [19] and Phuket and Othman [25] all conclude that there are two main sources of errors,

namely inter-lingual interference and intra-lingual interference. The former represents negative transfer of learners' first language and the latter involves errors caused by learners' incomplete knowledge of the target language. After studying and analysing research on errors committed by Thai EFL learners, another Thai scholar, Hinnon [26], proposed a slightly different view that there are three sources of errors: negative transfer of the mother tongue, limited knowledge of the target language and the difference between words and sentence structures of the mother tongue and those of the target language.

In summary, two major sources leading to errors made by EFL learners are inter-lingual interference and intra-lingual interference. All the above-mentioned studies do not mention the detection of any erroneously structured paragraphs, or any paragraphs that are too short or too long. However, they do dive deep into the reasoning behind why EFL learners make errors in their writings.

2.5 Summary of Related Work

In summary, from the current literature, we can conclude that the structure of a paragraph is not the main mistake that editors or analysts look for, in general, making it the least significant or detectable error. None of the studies mentioned the detection of any erroneously structured paragraphs, or any paragraphs that were too short or too long. However, they do dive deeply into the reasoning behind why EFL learners make errors in their writing. Other papers and/or techniques in linguistics steganography presumes that or exploit more noticeable or detectable errors because they assume that the human eye or analyser will not doubt or suspect that there is any significance to them and would assume it's a real world style of writing or unedited text.

In error analysis, the studies proved that even native English speakers would assume a sentence or piece of text is erroneous, even if the text has been published by a prestigious entity such as the Oxford Dictionary. This supports the idea that visual detection is not as accurate as computational detection. The studies done regarding paragraph structure are outdated and above 50 years old. There are no recent studies done in the 21st century in relation to the study and perception of paragraph structures.

3 Paragraph Perception Survey

Our main aim is to analyse and understand how the human eye perceives paragraph breaks and their location, and what humans may consider as a *normal* document, on the basis of three factors: age, native language and level of education. Our research used a survey-based approach, in which the survey was answered by a total of 234 participants, between 1 December 2022 and 30 March 2023. Below, we outline the survey design, the document data points of interest and characteristics of the sample.

3.1 Survey Design

The survey conducted in this paper consisted of a number of multiple choice questions to identify the age group, native language and the level of education of the participant. The participants were then asked to complete one assignment of downloading a file containing 5 text excerpts (i.e. documents) from different novels, as shown in Appendix A. The assignment then asked the participant to create paragraph breaks by using the "Enter" command and/or a newline character to create paragraphs where the participant deems suitable. Each text excerpt originally itself contained 5 paragraphs. Then participants were asked to submit the *paragraphed* versions of the text excerpts, along with their attributes.

3.2 Text Data Points

Our analysis defines a number of *data points* that characterise the document in terms of the breaking of its text into paragraphs. These data points are abbreviated and described in Table 1. The points fall into three categories, each describing certain statistical characteristics of the document. Data points (DP1–DP4) describe general significance of differences in a document, data points (DP5–DP10) describe grammatical significance of differences in a document and finally, data points (DP11–DP22) describe average paragraph proportion significance of differences in a document. We consider the special letter for the case of DP14, DP18 and DP22 to be the letter 'e', which has the highest frequency of occurrence in English [27]. Thus its distribution in a text, hypothetically, will hold some significance in analysing the text's structure.

After extracting these data points from the participants' responses, we then calculated the mean and standard deviations of the data according to the three factor dimensions; age, native language and level of education. If any differences are observed between the mean according to the dimension, we apply the one-way (i.e. single factor) Analysis of Variance (ANOVA) test [28] to the data to investigate further. The one-way ANOVA test compares the mean factor of two or more independent groups, via the F-distribution [29], in order to determine whether there is statistical evidence that the associated population mean values are significantly different [30]. If the significance result is greater than 0.05, it means that there are no statistically significant differences in the mean values of the data point in accordance to the dimension (known as the null hypothesis of the ANOVA test). However, if any level of the data points show differences in mean values under 0.05, then the difference is significant according to the dimension at hand (known as the alternative hypothesis.) We then apply Scheffé's test [31] to determine the source of differences. The formula test value for Scheffé's test is as follows:

$$F = \frac{(\overline{X}_i - \overline{X}_j)^2}{MSW[\frac{1}{n_i} + \frac{1}{n_j}]} \tag{1}$$

where:
F is the test statistic (F-ratio)

Table 1: Table of data point abbreviations

Data Point	Abbreviation	Description
DP1	ANPpF	Average of number of paragraphs per file
DP2	AoANSpP	Average of averages of number of sentences per paragraph
DP3	AoANWpP	Average of averages of number of words per paragraph
DP4	AoANCpP	Average of averages of number of characters per paragraph
DP5	ANP(S(V))	Average number of paragraphs starting with a verb
DP6	ANP(S(N))	Average number of paragraphs starting with a noun
DP7	ANP(S(A))	Average number of paragraphs starting with an adjective
DP8	ANP(E(V))	Average number of paragraphs ending with a verb
DP9	ANP(E(N))	Average number of paragraphs ending with a noun
DP10	ANP(E(A))	Average number of paragraphs ending with an adjective
DP11	AFPpp(SC)	Average first paragraph proportion percentage based on sentence count
DP12	AFPpp(WC)	Average first paragraph proportion percentage based on word count
DP13	AFPpp(CC)	Average first paragraph proportion percentage based on character count
DP14	AFPp(SWLp)	Average first paragraph special word or letter percentage
DP15	AMPpp(SC)	Average middle paragraph proportion percentage based on sentence count
DP16	AMPpp(WC)	Average middle paragraph proportion percentage based on word count
DP17	AMPpp(CC)	Average middle paragraph proportion percentage based on character count
DP18	AMPp(SWLp)	Average middle paragraph special word or letter percentage
DP19	ALPpp(SC)	Average last paragraph proportion percentage based on sentence count
DP20	ALPpp(WC)	Average last paragraph proportion percentage based on word count
DP21	ALPpp(CC)	Average last paragraph proportion percentage based on character count
DP22	ALPp(SWLp)	Average last paragraph special word or letter percentage

$\overline{X_i}$ is the mean of the i^{ith} sample
$\overline{X_j}$ is the mean of the j^{ith} sample
n_i is the number of measurements in the i^{ith} sample
n_j is the number of measurements in the j^{ith} sample
MSW is the Mean of Squares Within

Scheffé's test involves computing an F value for each combination of mean values, since in many instances, it is unknown which comparisons of values need to be made for experimenters, and the comparisons of interest are only discovered after a preliminary examination of the data. The Scheffé method computes all possible contrasts between mean values, where the Type 1 error is at most at α level of significance for any of the possible combinations. A Type 1 error is defined as concluding that a difference is significant when it is not, i.e. falsely rejecting the null hypothesis [31, 32]. The test is used as part of an ANOVA test to test multiple comparisons among a group of mean values, in order to discover which pairs have significant differences according to some dimension. A '*' in a Scheffé table indicates a significant difference between two variables of one dimension. The variable with the highest mean is the variable with which significance benefits. Application of this test will show which relation between a dimension and factor, like age and average number of paragraphs per text, has high significance. As such, this would mean that the test has revealed significant existing relation, which needs to be considered when implementing an embedding model, for example, as an application.

3.3 Sample Characteristics

The sample consisted of 242 participants, out of which 234 returned valid answers. Tables 2–4 show the distributions of the sample over the native language, education and age dimensions, respectively.

Table 2: Distribution of the Sample According to Native Language

Native Language	Frequency	%
English	101	43.2
Spanish	36	15.4
Polish	29	12.4
Portuguese	17	7.2
Other	51	21.8
Total	234	100.0

4 Results of the Study

We reported in an earlier version of this paper [3] the survey results for the native language dimension only for our study. Those results are repeated here in Tables

Table 3: Distribution of the Sample According to Education

Level of Education	Frequency	%
Secondary School	12	5.1
High school	59	25.2
Diploma	33	14.1
Bachelors	89	38.0
Postgraduate (Masters, PhD)	41	17.6
Total	234	100.0

Table 4: Distribution of the Sample According to Age

Age	Frequency	%
18-25	118	50.4
26-30	40	17.1
31-49	59	25.2
50+	17	7.3
Total	234	100.0

5 (mean and standard deviation results) and 8 (significance of differences), for each one of our 22 data points identified in Table 1.

In addition to the above, we include, in this extended version of the paper, the results of the other two remaining dimensions; education and age. Tables 6 and 7 show the mean and standard deviation results for the education and the age dimensions, respectively. On the other hand, Tables 9 and 10 show the significance of differences results for the education and age dimensions, respectively. We next discuss these results in more detail.

5 Analysis and Interpretation of Results

From Tables 8–10, we observe two significant differences:

1. First, in Table 8, the difference in the ANP(S(V)) data point values is significant according to the native language dimension. To determine the source of this difference, Scheffé's test is used as shown in Table 11, where '*' indicates that the mean difference is significant at 0.05 or below. From the results of Table 11, we find that the statistically significant differences in the level of the ANP(S(V)) according to the native language dimension is between the Spanish Language and the English and Other languages, for the benefit of the Spanish language, and between the Portuguese language and the English and Other languages, for the benefit of the Portuguese language. Spanish native speakers perceive that, on average, 2 out of 5 paragraphs must start with a verb, corresponding to the mean value of 2.000, with a Standard deviation of 1.171. Similarly, for Portuguese native speakers, we observe a significance of a mean of 2.059, for paragraphs

Table 5: Mean and standard deviation for data points (DP1–DP22) according to the language dimension

Dimensions	Statistic	English	Spanish	Polish	Portuguese	Other	Total
ANPpF	N	101	36	29	17	51	234
	Mean	4.77	5.86	5.17	6.41	5.10	5.18
	Std. Dev.	2.701	3.155	3.506	2.063	2.524	2.832
AoANSpP	N	101	36	29	17	51	234
	Mean	9.921	8.028	9.310	6.706	9.255	9.175
	Std. Dev.	7.3644	7.4928	5.3325	6.6216	6.5080	6.9368
AoANWpP	N	101	36	29	17	51	234
	Mean	162.356	132.139	152.759	110.118	151.196	150.291
	Std. Dev.	124.2071	124.0949	90.5052	109.8658	107.2634	115.9713
AoANCpP	N	101	36	29	17	51	234
	Mean	894.634	726.333	838.897	603.118	832.431	827.098
	Std. Dev.	685.5332	683.6008	498.3605	605.4870	591.4900	639.7054
ANP(S(V))	N	101	36	29	17	51	234
	Mean	1.446	2.000	1.655	2.059	1.412	1.594
	Std. Dev.	0.9325	1.1711	1.2328	0.9663	0.9418	1.0367
ANP(S(N))	N	101	36	29	17	51	234
	Mean	1.188	1.583	1.448	1.529	1.216	1.312
	Std. Dev.	1.1288	1.2956	1.4289	.7998	0.9233	1.1389
ANP(S(A))	N	101	36	29	17	51	234
	Mean	0.535	0.694	0.690	0.588	0.529	0.581
	Std. Dev.	0.5756	0.7491	0.7608	0.5073	0.5780	0.6248
ANP(E(V))	N	101	36	29	17	51	234
	Mean	0.455	0.611	0.621	0.529	0.569	0.530
	Std. Dev.	0.7004	0.5989	9.9416	9.8745	0.6710	0.7245
ANP(E(N))	N	101	36	29	17	51	234
	Mean	1.178	1.139	1.414	1.353	1.216	1.222
	Std. Dev.	.6387	.4245	1.2397	1.1147	0.5025	0.7308
ANP(E(A))	N	101	36	29	17	51	234
	Mean	1.030	1.000	1.138	1.059	1.000	1.034
	Std. Dev.	.1706	.0000	.5809	.2425	0.0000	0.2427
AFPpp(SC)	N	101	36	29	17	51	234
	Mean	29.762	25.278	25.138	20.941	26.804	27.214
	Std. Dev.	23.5704	23.7927	14.9182	20.6563	20.6940	21.8690
AFPpp(WC)	N	101	36	29	17	51	234
	Mean	28.941	24.361	24.483	19.765	26.039	26.385
	Std. Dev.	24.0128	24.0145	15.3661	20.9121	21.0104	22.2345
AFPpp(CC)	N	101	36	29	17	51	234
	Mean	28.792	24.194	24.414	19.706	25.941	26.261
	Std. Dev.	23.9926	24.0592	15.3169	20.9218	21.0109	22.2264
AFPpp(SWLp)	N	101	36	29	17	51	234
	Mean	28.901	24.278	24.655	19.765	26.157	26.402
	Std. Dev.	23.8904	24.0150	15.3792	20.9121	21.0232	22.1788
AMPpp(SC)	N	93	33	26	16	48	216
	Mean	21.097	16.273	20.500	15.375	20.625	19.759
	Std. Dev.	7.7179	4.9767	11.4481	6.4588	8.5019	8.2116
AMPpp(WC)	N	93	33	26	16	48	216
	Mean	23.624	18.061	22.923	17.938	23.604	22.264
	Std. Dev.	7.7850	5.1717	12.3610	6.1043	9.1040	8.5882
AMPpp(CC)	N	93	33	26	16	48	216
	Mean	23.624	18.182	22.962	18.000	23.771	22.329
	Std. Dev.	7.6739	5.0402	12.5299	6.2823	9.2282	8.5952
AMPpp(SWLp)	N	93	33	26	16	48	216
	Mean	23.731	18.273	22.692	18.062	23.771	22.361
	Std. Dev.	8.0260	5.0576	12.3216	6.2553	9.1262	8.6701
ALPpp(SC)	N	96	33	29	16	48	222
	Mean	27.677	21.939	32.379	20.563	28.604	27.126
	Std. Dev.	9.9160	5.5506	18.9009	6.5825	10.7263	11.4351
ALPpp(WC)	N	96	33	29	16	48	222
	Mean	24.771	18.485	29.724	17.312	25.438	24.090
	Std. Dev.	10.4926	6.3004	19.7211	6.6805	11.3342	12.0613
ALPpp(CC)	N	96	33	29	16	48	222
	Mean	25.031	18.848	29.931	17.625	25.521	24.324
	Std. Dev.	10.4524	6.3103	19.6013	6.7909	11.2836	11.9997
ALPpp(SWLp)	N	96	33	29	16	48	222
	Mean	25.167	18.788	30.034	17.500	25.625	24.401
	Std. Dev.	10.5538	6.4312	19.5055	6.9952	11.3674	12.0736

Table 6: Mean and standard deviation for data points (DP1–DP22) according to the education dimension

Dimensions	Statistic	Secondary School	High school	Diploma	Bachelors	Post-graduate	Total
ANPpF	N	12	59	33	89	41	234
	Mean	4.33	5.54	4.82	5.34	4.85	5.18
	Std. Dev.	1.775	3.297	2.378	2.965	2.351	2.832
AoANSpP	N	12	59	33	89	41	234
	Mean	10.667	8.492	10.727	8.551	9.829	9.175
	Std. Dev.	8.6164	5.9520	8.7367	6.0903	7.8227	6.9368
AoANWpP	N	12	59	33	89	41	234
	Mean	179.083	137.542	174.455	140.697	161.585	150.291
	Std. Dev.	148.1808	97.6511	147.9092	101.9109	129.6962	115.9713
AoANCpP	N	12	59	33	89	41	234
	Mean	985.333	757.373	961.515	774.483	887.146	827.098
	Std. Dev.	817.3969	538.2066	816.6442	562.5013	714.7749	639.7054
ANP(S(V))	N	12	59	33	89	41	234
	Mean	1.500	1.661	1.636	1.652	1.366	1.594
	Std. Dev.	.7977	1.2122	1.0845	1.0346	.7667	1.0367
ANP(S(N))	N	12	59	33	89	41	234
	Mean	1.083	1.373	1.182	1.449	1.098	1.312
	Std. Dev.	.5149	1.1876	.9828	1.3900	.5387	1.1389
ANP(S(A))	N	12	59	33	89	41	234
	Mean	.333	.712	.485	.596	.512	.581
	Std. Dev.	.4924	.6446	.5658	.6862	.5061	.6248
ANP(E(V))	N	12	59	33	89	41	234
	Mean	.083	.661	.576	.494	.512	.530
	Std. Dev.	.2887	.7337	.9364	.6763	.6753	.7245
ANP(E(N))	N	12	59	33	89	41	234
	Mean	1.000	1.305	1.364	1.135	1.244	1.222
	Std. Dev.	.0000	.7011	1.2201	.6251	.5376	.7308
ANP(E(A))	N	12	59	33	89	41	234
	Mean	1.000	1.051	1.121	1.011	1.000	1.034
	Std. Dev.	.0000	.2216	.5453	.1060	.0000	.2427
AFPpp(SC)	N	12	59	33	89	41	234
	Mean	30.333	24.814	32.303	25.685	28.976	27.214
	Std. Dev.	26.7593	18.6358	28.1841	19.1220	24.5820	21.8690
AFPpp(WC)	N	12	59	33	89	41	234
	Mean	29.333	24.407	31.606	24.640	27.951	26.385
	Std. Dev.	27.1673	18.8699	28.9244	19.4213	24.9238	22.2345
AFPpp(CC)	N	12	59	33	89	41	234
	Mean	29.250	24.271	31.545	24.472	27.878	26.261
	Std. Dev.	27.1800	18.8010	28.8835	19.4633	24.8950	22.2264
AFPpp(SWLp)	N	12	59	33	89	41	234
	Mean	29.583	24.373	31.697	24.652	27.927	26.402
	Std. Dev.	27.2145	18.7349	28.8731	19.3685	24.8801	22.1788
AMPpp(SC)	N	11	53	29	85	38	216
	Mean	20.727	19.283	18.931	19.635	21.053	19.759
	Std. Dev.	11.7396	8.3098	8.3192	7.4031	8.8224	8.2116
AMPpp(WC)	N	11	53	29	85	38	216
	Mean	22.091	21.962	20.897	22.094	24.158	22.264
	Std. Dev.	12.8176	8.8360	8.5454	7.6713	8.9579	8.5882
AMPpp(CC)	N	11	53	29	85	38	216
	Mean	22.000	22.113	20.931	22.188	24.105	22.329
	Std. Dev.	12.9151	8.9629	8.5604	7.6072	8.9194	8.5952
AMPpp(SWLp)	N	11	53	29	85	38	216
	Mean	21.909	22.075	21.069	22.271	24.079	22.361
	Std. Dev.	13.1564	8.9203	8.7339	7.7466	8.9273	8.6701
ALPpp(SC)	N	11	57	30	86	38	222
	Mean	29.182	28.228	24.167	26.802	27.947	27.126
	Std. Dev.	16.5941	14.4482	10.0998	9.6550	9.2414	11.4351
ALPpp(WC)	N	11	57	30	86	38	222
	Mean	26.455	25.175	21.233	23.686	24.947	24.090
	Std. Dev.	17.7559	15.0085	10.2307	10.3131	10.2560	12.0613
ALPpp(CC)	N	11	57	30	86	38	222
	Mean	26.455	25.333	21.500	23.977	25.211	24.324
	Std. Dev.	17.7503	14.9718	10.1837	10.1784	10.2908	11.9997
ALPpp(SWLp)	N	11	57	30	86	38	222
	Mean	26.636	25.386	21.533	24.070	25.289	24.401
	Std. Dev.	17.3509	15.0473	10.4510	10.2977	10.3295	12.0736

Table 7: Mean and standard deviation for data points (DP1–DP22) according to the age dimension

Dimensions	Statistic	18-25	26-30	31-49	50+	Total
ANPpF	N	118	40	59	17	234
	Mean	5.36	4.68	5.20	5.06	5.18
	Std. Dev.	2.925	2.080	2.987	3.230	2.832
AoANSpP	N	118	40	59	17	234
	Mean	8.746	10.100	9.169	10.000	9.175
	Std. Dev.	6.2322	8.0154	7.1731	8.3292	6.9368
AoANWpP	N	118	40	59	17	234
	Mean	143.686	163.650	151.390	160.882	150.291
	Std. Dev.	104.2561	132.6487	120.6222	140.8918	115.9713
AoANCpP	N	118	40	59	17	234
	Mean	791.110	901.550	831.610	886.059	827.098
	Std. Dev.	575.8609	731.3202	664.6353	776.0401	639.7054
ANP(S(V))	N	118	40	59	17	234
	Mean	1.653	1.425	1.576	1.647	1.594
	Std. Dev.	1.1351	0.8439	0.9136	1.1695	1.0367
ANP(S(N))	N	118	40	59	17	234
	Mean	1.398	1.075	1.339	1.176	1.312
	Std. Dev.	1.1849	0.7299	1.3211	0.8828	1.1389
ANP(S(A))	N	118	40	59	17	234
	Mean	0.644	0.500	0.559	0.412	0.581
	Std. Dev.	0.6734	0.5064	0.5951	.06183	0.6248
ANP(E(V))	N	118	40	59	17	234
	Mean	0.576	0.575	0.424	0.471	0.530
	Std. Dev.	0.7327	0.8738	0.5633	0.7998	0.7245
ANP(E(N))	N	118	40	59	17	234
	Mean	1.212	1.375	1.102	1.353	1.222
	Std. Dev.	0.7379	1.0300	0.4024	0.7019	0.7308
ANP(E(A))	N	118	40	59	17	234
	Mean	1.042	1.050	1.000	1.059	1.034
	Std. Dev.	0.3037	0.2207	0.0000	0.2425	0.2427
AFPpp(SC)	N	118	40	59	17	234
	Mean	25.898	30.475	27.186	28.765	27.214
	Std. Dev.	19.5266	25.4387	22.3255	27.3919	21.8690
AFPpp(WC)	N	118	40	59	17	234
	Mean	25.186	29.275	26.288	28.235	26.385
	Std. Dev.	20.0222	25.8843	22.5702	27.4398	22.2345
AFPpp(CC)	N	118	40	59	17	234
	Mean	25.068	29.125	26.153	28.176	26.261
	Std. Dev.	19.9952	25.9222	22.5876	27.3524	22.2264
AFPpp(SWLp)	N	118	40	59	17	234
	Mean	25.229	29.150	26.339	28.294	26.402
	Std. Dev.	19.9527	25.8641	22.5485	27.3056	22.1788
AMPpp(SC)	N	109	36	55	16	216
	Mean	18.917	20.389	20.545	21.375	19.759
	Std. Dev.	7.4400	8.7154	9.3469	8.0654	8.2116
AMPpp(WC)	N	109	36	55	16	216
	Mean	21.220	23.056	23.345	23.875	22.264
	Std. Dev.	7.9888	8.6782	9.6037	8.5703	8.5882
AMPpp(CC)	N	109	36	55	16	216
	Mean	21.303	23.139	23.364	23.938	22.329
	Std. Dev.	8.0318	8.6360	9.6463	8.3384	8.5952
AMPpp(SC)e	N	109	36	55	16	216
	Mean	21.367	23.111	23.382	23.938	22.361
	Std. Dev.	8.0227	8.8762	9.6963	8.7595	8.6701
ALPpp(SC)	N	114	36	56	16	222
	Mean	27.684	26.528	26.893	25.313	27.126
	Std. Dev.	13.1279	9.6228	9.1208	10.0115	11.4351
ALPpp(WC)	N	114	36	56	16	222
	Mean	24.614	23.361	23.911	22.625	24.090
	Std. Dev.	13.7432	10.1084	10.0187	10.4043	12.0613
ALPpp(CC)	N	114	36	56	16	222
	Mean	24.825	23.639	24.143	22.938	24.324
	Std. Dev.	13.6636	10.1197	9.9168	10.5038	11.9997
ALPpp(SWLp)	N	114	36	56	16	222
	Mean	24.904	23.806	24.143	23.062	24.401
	Std. Dev.	13.7409	10.3505	9.8616	10.6675	12.0736

14

Table 8: Significance of the differences for data points (DP1–DP22) according to the language dimension

Dimensions	Source of Variance	Sum of Squares	df	Mean Squares	F	Significance
ANPpF	Between Groups	59.628	4	14.907	1.887	0.114
	Within Groups	1808.833	229	7.899		
	Total	1868.462	233			
AoANSpP	Between Groups	208.055	4	52.014	1.082	0.366
	Within Groups	11003.761	229	48.051		
	Total	11211.816	233			
AoANWpP	Between Groups	54219.651	4	13554.913	1.008	0.404
	Within Groups	3079478.588	229	13447.505		
	Total	3133698.239	233			
AoANCpP	Between Groups	1684524.330	4	421131.082	1.030	0.393
	Within Groups	93664444.410	229	409015.041		
	Total	95348968.739	233			
ANP(S(V))	Between Groups	13.635	4	3.409	3.297	0.012
	Within Groups	236.796	229	1.034		
	Total	250.432	233			
ANP(S(N))	Between Groups	6.016	4	1.504	1.163	0.328
	Within Groups	296.211	229	1.293		
	Total	302.226	233			
ANP(S(A))	Between Groups	1.159	4	0.290	0.739	0.566
	Within Groups	89.798	229	0.392		
	Total	90.957	233			
ANP(E(V))	Between Groups	1.113	4	0.278	0.526	0.717
	Within Groups	121.178	229	0.529		
	Total	122.291	233			
ANP(E(N))	Between Groups	1.803	4	0.451	0.841	0.500
	Within Groups	122.642	229	0.536		
	Total	124.444	233			
ANP(E(A))	Between Groups	0.426	4	0.107	1.834	0.123
	Within Groups	13.300	229	0.058		
	Total	13.726	233			
AFPpp(SC)	Between Groups	1593.368	4	398.342	0.830	0.507
	Within Groups	109839.948	229	479.650		
	Total	111433.316	233			
AFPpp(WC)	Between Groups	1663.214	4	415.803	0.839	0.502
	Within Groups	113526.171	229	495.747		
	Total	115189.385	233			
AFPpp(CC)	Between Groups	1635.438	4	408.860	0.825	0.510
	Within Groups	113469.660	229	495.501		
	Total	115105.098	233			
AFPpp(SWLp)	Between Groups	1633.652	4	408.413	0.828	0.509
	Within Groups	112978.588	229	493.356		
	Total	114612.239	233			
AMPpp(SC)	Between Groups	925.307	4	231.327	3.596	0.007
	Within Groups	13572.174	211	64.323		
	Total	14497.481	215			
AMPpp(WC)	Between Groups	1151.989	4	287.997	4.132	0.003
	Within Groups	14705.970	211	69.697		
	Total	15857.958	215			
AMPpp(CC)	Between Groups	1133.484	4	283.371	4.054	0.003
	Within Groups	14750.178	211	69.906		
	Total	15883.662	215			
AMPpp(SWLp)	Between Groups	1120.053	4	280.013	3.928	0.004
	Within Groups	15041.780	211	71.288		
	Total	16161.833	215			
ALPpp(SC)	Between Groups	2511.356	4	627.839	5.163	0.001
	Within Groups	26387.113	217	121.600		
	Total	28898.468	221			
ALPpp(WC)	Between Groups	2823.954	4	705.989	5.224	0.000
	Within Groups	29326.244	217	135.144		
	Total	32150.198	221			
ALPpp(CC)	Between Groups	2735.909	4	683.977	5.103	0.001
	Within Groups	29086.740	217	134.040		
	Total	31822.649	221			
ALPpp(SWLp)	Between Groups	2850.256	4	712.564	5.266	0.000
	Within Groups	29365.064	217	135.323		
	Total	32215.320	221			

Table 9: Significance of the differences for data points (DP1–DP22) according to the education dimension

Dimensions	Source of Variance	Sum of Squares	df	Mean Squares	F	Significance
ANPpF	Between Groups	27.232	4	6.808	0.847	0.497
	Within Groups	1841.229	229	8.040		
	Total	1868.462	233			
AoANSpP	Between Groups	186.031	4	46.508	0.966	0.427
	Within Groups	11025.785	229	48.148		
	Total	11211.816	233			
AoANWpP	Between Groups	52227.737	4	13056.934	0.970	0.425
	Within Groups	3081470.503	229	13456.203		
	Total	3133698.239	233			
AoANCpP	Between Groups	1577756.687	4	394439.172	0.963	0.428
	Within Groups	93771212.052	229	409481.275		
	Total	95348968.739	233			
ANP(S(V))	Between Groups	2.860	4	0.715	0.661	0.619
	Within Groups	247.571	229	1.081		
	Total	250.432	233			
ANP(S(N))	Between Groups	4.972	4	1.243	0.958	0.432
	Within Groups	297.255	229	1.298		
	Total	302.226	233			
ANP(S(A))	Between Groups	2.264	4	0.566	1.462	0.215
	Within Groups	88.693	229	0.387		
	Total	90.957	233			
ANP(E(V))	Between Groups	3.602	4	0.900	1.737	0.143
	Within Groups	118.689	229	0.518		
	Total	122.291	233			
ANP(E(N))	Between Groups	2.357	4	0.589	1.105	0.355
	Within Groups	122.088	229	0.533		
	Total	124.444	233			
ANP(E(A))	Between Groups	0.375	4	0.094	1.609	0.173
	Within Groups	13.351	229	.058		
	Total	13.726	233			
AFPpp(SC)	Between Groups	1646.564	4	411.641	0.859	0.490
	Within Groups	109786.752	229	479.418		
	Total	111433.316	233			
AFPpp(WC)	Between Groups	1606.205	4	401.551	0.810	0.520
	Within Groups	113583.180	229	495.996		
	Total	115189.385	233			
AFPpp(CC)	Between Groups	1654.435	4	413.609	0.835	0.504
	Within Groups	113450.663	229	495.418		
	Total	115105.098	233			
AFPpp(SWLp)	Between Groups	1657.574	4	414.393	0.840	0.501
	Within Groups	112954.666	229	493.252		
	Total	114612.239	233			
AMPpp(SC)	Between Groups	107.094	4	26.774	0.393	0.814
	Within Groups	14390.387	211	68.201		
	Total	14497.481	215			
AMPpp(WC)	Between Groups	198.135	4	49.534	0.667	0.615
	Within Groups	15659.823	211	74.217		
	Total	15857.958	215			
AMPpp(CC)	Between Groups	181.912	4	45.478	0.611	0.655
	Within Groups	15701.750	211	74.416		
	Total	15883.662	215			
AMPpp(SWLp)	Between Groups	167.824	4	41.956	0.554	0.697
	Within Groups	15994.009	211	75.801		
	Total	16161.833	215			
ALPpp(SC)	Between Groups	413.096	4	103.274	0.787	0.535
	Within Groups	28485.372	217	131.269		
	Total	28898.468	221			
ALPpp(WC)	Between Groups	415.441	4	103.860	0.710	0.586
	Within Groups	31734.758	217	146.243		
	Total	32150.198	221			
ALPpp(CC)	Between Groups	387.485	4	96.871	0.669	0.614
	Within Groups	31435.163	217	144.863		
	Total	31822.649	221			
ALPpp(SWLp)	Between Groups	396.402	4	99.100	0.676	0.609
	Within Groups	31818.918	217	146.631		
	Total	32215.320	221			

Table 10: Significance of the differences for data points (DP1–DP22) according to the age dimension

Dimensions	Source of Variance	Sum of Squares	df	Mean Squares	F	Significance
ANPpF	Between Groups	14.135	3	4.712	0.584	0.626
	Within Groups	1854.326	230	8.062		
	Total	1868.462	233			
AoANSpP	Between Groups	67.538	3	22.513	0.465	0.707
	Within Groups	11144.278	230	48.453		
	Total	11211.816	233			
AoANWpP	Between Groups	14263.942	3	4754.647	0.351	0.789
	Within Groups	3119434.297	230	13562.758		
	Total	3133698.239	233			
AoANCpP	Between Groups	434848.296	3	144949.432	0.351	0.788
	Within Groups	94914120.443	230	412670.089		
	Total	95348968.739	233			
ANP(S(V))	Between Groups	1.613	3	0.538	0.497	0.685
	Within Groups	248.818	230	1.082		
	Total	250.432	233			
ANP(S(N))	Between Groups	3.481	3	1.160	0.893	0.445
	Within Groups	298.746	230	1.299		
	Total	302.226	233			
ANP(S(A))	Between Groups	1.246	3	0.415	1.065	0.365
	Within Groups	89.711	230	0.390		
	Total	90.957	233			
ANP(E(V))	Between Groups	1.060	3	0.353	0.670	0.571
	Within Groups	121.231	230	0.527		
	Total	122.291	233			
ANP(E(N))	Between Groups	2.094	3	0.698	1.312	0.271
	Within Groups	122.351	230	0.532		
	Total	124.444	233			
ANP(E(A))	Between Groups	.097	3	0.032	0.547	0.651
	Within Groups	13.629	230	0.059		
	Total	13.726	233			
AFPpp(SC)	Between Groups	670.554	3	223.518	0.464	0.708
	Within Groups	110762.763	230	481.577		
	Total	111433.316	233			
AFPpp(WC)	Between Groups	562.351	3	187.450	0.376	0.770
	Within Groups	114627.034	230	498.378		
	Total	115189.385	233			
AFPpp(CC)	Between Groups	559.168	3	186.389	0.374	0.772
	Within Groups	114545.930	230	498.026		
	Total	115105.098	233			
AFPpp(SWLp)	Between Groups	525.568	3	175.189	0.353	0.787
	Within Groups	114086.672	230	496.029		
	Total	114612.239	233			
AMPpp(SC)	Between Groups	167.283	3	55.761	0.825	0.481
	Within Groups	14330.199	212	67.595		
	Total	14497.481	215			
AMPpp(WC)	Between Groups	247.167	3	82.389	1.119	0.342
	Within Groups	15610.791	212	73.636		
	Total	15857.958	215			
AMPpp(CC)	Between Groups	238.683	3	79.561	1.078	0.359
	Within Groups	15644.980	212	73.797		
	Total	15883.662	215			
AMPpp(SWLp)	Between Groups	225.037	3	75.012	0.998	0.395
	Within Groups	15936.796	212	75.174		
	Total	16161.833	215			
ALPpp(SC)	Between Groups	104.070	3	34.690	0.263	0.852
	Within Groups	28794.398	218	132.084		
	Total	28898.468	221			
ALPpp(WC)	Between Groups	86.572	3	28.857	0.196	0.899
	Within Groups	32063.627	218	147.081		
	Total	32150.198	221			
ALPpp(CC)	Between Groups	78.057	3	26.019	0.179	0.911
	Within Groups	31744.591	218	145.617		
	Total	31822.649	221			
ALPpp(SWLp)	Between Groups	73.948	3	24.649	0.167	0.918
	Within Groups	32141.372	218	147.437		
	Total	32215.320	221			

Table 11: Source of differences in the level of the ANP(S(V)) data point, according to the language dimension

Mean	Language	English	Spanish	Polish	Portuguese	Other
1.446	English		*		*	
2.000	**Spanish**					
1.655	Polish					
2.059	**Portuguese**					
1.412	Other		*		*	

starting with a verb, and a standard deviation of 0.966. Hence, whether or not a paragraph starts with a verb is significant with Spanish and Portuguese native speakers.

2. The second significant difference, also as observed from Table 8, was in the average paragraph proportion data points DP15–DP22, also within the native language dimension. To determine the source of these differences, we applied again Scheffé's test, as shown in Table 12. Results in

Table 12: Source of differences in DP15–DP22 data points, according to the language dimension

Dimensions	Mean	Language	English	Spanish	Polish	Portuguese	Other
AMPpp(SC)	21.097	**English**					
	16.273	Spanish	*		*		*
	20.500	**Polish**					
	15.375	Portuguese	*		*		*
	20.625	**Other**					
AMPpp(WC)	23.624	**English**					
	18.061	Spanish	*		*		*
	22.923	**Polish**					
	17.938	Portuguese	*		*		*
	23.604	**Other**					
AMPpp(CC)	23.624	**English**					
	18.182	Spanish	*		*		*
	22.962	**Polish**					
	18.000	Portuguese	*		*		*
	23.771	**Other**					
AMPpp(SWLp)	23.731	**English**					
	18.273	Spanish	*		*		*
	22.692	**Polish**					
	18.062	Portuguese	*		*		*
	23.771	**Other**					
ALPpp(SC)	27.677	**English**					
	21.939	Spanish	*		*		*
	32.379	**Polish**					
	20.563	Portuguese	*		*		*
	28.604	**Other**					
ALPpp(WC)	24.771	**English**					
	18.485	Spanish	*		*		*
	29.724	**Polish**					
	17.312	Portuguese	*		*		*
	25.438	**Other**					
ALPpp(CC)	25.031	**English**					
	18.848	Spanish	*		*		*
	29.931	Polish					
	17.625	Portuguese	*		*		*
	25.521	**Other**					
ALPpp(SWLp)	25.167	**English**					
	18.788	Spanish	*		*		*
	30.034	**Polish**					
	17.500	Portuguese	*		*		*
	25.625	**Other**					

Table 12 show that the statistically significant differences in all of the D15–D22 datapoints according to the language dimension are between the English, Polish and Other languages, on one hand, and the Spanish and Portuguese languages on the other hand, for the benefit of the English, Polish and Other Languages. This meant that for native speakers of English, Polish and Other languages, the values of these data points are important. For English native speakers, DP15–DP18 mean values were (21.097, 23.624, 23.624, 23.731), therefore indicating that this group significantly perceived the middle paragraphs to contain 23.1% of the total number of sentences, words, characters and the special word or letter in a document. This figure was 22.3% for native Polish speakers and 22.9% for the Other languages. As for the DP19–DP22 data points, English native speakers showed mean values of (27.677, 24.771, 25.031, 25.167), which translate into the significant preference that 25.7% of sentences, words, characters and the special word or letter of the text would be in the end part of a text document. That figure was 30.5% for native Polish speakers and 26.3% for the Other languages. Hence, we conclude that middle and last paragraph proportions are significant with English, Polish and some Other language native speakers.

The results also showed that the other two dimensions used in the survey, namely age and level of education of participant, did not hold any significance in terms of the variations in the data points, and therefore, do not demonstrate any preference in the participants' perception of how paragraphs should be structured.

6   Conclusion

We demonstrated through the results of the study presented in this paper that the native language dimension has impact on the readers' perception of some of the characteristics of the document, particularly in relation to the average number of paragraphs starting with verbs and characteristics related to the average paragraph proportions for middle and last paragraphs. Our study also demonstrated, according to the sample surveyed, what the normal values of a number of data points related to the document characteristics. Of particular interest, is the mean value of the number of paragraphs, as this has direct relevance to a recent message embedding method proposed in [2], where this characteristic provides the basis for the human detection of embedded material. Besides forming some understanding of how humans perceive paragraphs and paragraph breaks, we feel that the true significance of the results of this paper lies in the fact that it forms an important step in aiding the visual detection of document structure manipulation in applications that involve the embedding of secret messages, such as in steganography.

In future extensions of this research, we would be interested to know if the actual survey design had an impact on answering the research question, and consequently, their choice of the paragraph structure in a document. Therefore, we plan to redesign the survey in the form of a multiple-choice test where

participants are shown a number of different possible paragraph breaks for a text excerpt, rather than asking directly to break the excerpt into paragraphs. This change would carry impression of the possible answers may be. We also plan to extend the scope of the study to include other categories different from language, level of education and age, to form a higher level "cultural" dimension to the study. The size of the sample surveyed could also be increased.

References

[1] Britton BK, Glynn SM, Meyer BJ, Penland M. Effects of text structure on use of cognitive capacity during reading. Journal of Educational Psychology. 1982;74(1):51.

[2] Aziz B, Bukhelli A, Khusainov R, Mohasseb A. A Novel Method for Embedding and Extracting Secret Messages in Textual Documents based on Paragraph Resizing. In: 19th International Conference on Security and Cryptography. SciTePress; 2022. p. 714-9.

[3] Aziz B, Bukhelli A. A Study on Human Perception of Document Paragraph Layout. In: 2023 14th International Conference on Information, Intelligence, Systems & Applications (IISA 2023). Volos, Greece: IEEE Computer Society; 2023. p. 1-8.

[4] Padučeva E. On the Structure of the Paragraph. Linguistics: An Interdisciplinary Journal of the Language Sciences. 1974.

[5] Braddock R. The Frequency and Placement of Topic Sentences in Expository Prose. Research in the Teaching of English. 1974;8(3):287-302.

[6] Williams M, Stevens VMR. Understanding Paragraph Structure. Journal of Reading. 1972;15(7):513-6.

[7] Irmscher WF, Christensen F, Becker AL, Rodgers PC, Miles J, Karrfalt DH. Symposium on the Paragraph. College Composition and Communication. 1966;17(2):60-87.

[8] Becker AL. A tagmemic approach to paragraph analysis. College Composition and Communication. 1965;16(5):237-42.

[9] Hinds J. Paragraph structure and pronominalization. Paper in Linguistics. 1977;10(1-2):77-99.

[10] Richards JC. Error analysis: Perspectives on second language acquisition. Routledge; 1974.

[11] Darus S, Subramaniam K. Error analysis of the written English essays of secondary school students in Malaysia: A case study. European journal of social sciences. 2009;8(3):483-95.

[12] Sermsook K, Liamnimit J, Pochakorn R. An Analysis of Errors in Written English Sentences: A Case Study of Thai EFL Students. English Language Teaching. 2017;10(3):101-10.

[13] Vann RJ, Meyer DE, Lorenz FO. Error Gravity: A Study of Faculty Opinion of ESL Errors. TESOL Quarterly. 1984;18(3):427-40.

[14] LENNON P. Error: Some Problems of Definition, Identification, and Distinction. Applied Linguistics. 1991 06;12(2):180-96.

[15] Burt MK. Error Analysis in the Adult EFL Classroom. TESOL Quarterly. 1975;9(1):53-63.

[16] Dulay H, Burt MK, Krashen SD. Language two. Oxford University Press; 1982.

[17] James C. Errors in language learning and use: Exploring error analysis. New York: Routledge; 1998.

[18] Hengwichitkul L. An analysis of errors in English abstracts translated by Thai university graduate students. Unpublished master's thesis) Srinakharinwirot University, Bangkok, Thailand. 2006.

[19] Runkati K, Chatupote M. Organizational patterns and common mistakes in English research abstracts. Prince of Songkla University, Songkhla, Thailand. 2013.

[20] Richards JC. Error analysis: Perspectives on second language acquisition. Routledge; 2015.

[21] James C. Errors in language learning and use: Exploring error analysis. Routledge; 2013.

[22] Penny WK. An Analysis of Student Error Patterns in Written English Suggested Teaching Procedures to help. Unpublished master's thesis. 2001 May:1-46.

[23] Heydari P, Bagheri MS. Error analysis: Sources of L2 learners' errors. Theory and practice in language studies. 2012;2(8):1583-9.

[24] Kaweera C. Writing Error: A Review of Interlingual and Intralingual Interference in EFL Context. English language teaching. 2013;6(7):9-18.

[25] Phuket PRN, Othman NB. Understanding EFL Students' Errors in Writing. Journal of Education and Practice. 2015;6(32):99-106.

[26] Hinnon A. Common errors in English writing and suggested solutions of Thai university students. Humanities and Social Sciences. 2015;31(2):165-80.

[27] Ridley DR, Dominguez PS, Walker CB. English Letter Frequencies in Transcribed Speech versus Written Samples. Perceptual and Motor Skills. 1999;88(3 part 2):1181-8.

[28] Fisher RA. XV.—The correlation between relatives on the supposition of Mendelian inheritance. Earth and Environmental Science Transactions of the Royal Society of Edinburgh. 1919;52(2):399-433.

[29] Fisher RA. On the mathematical foundations of theoretical statistics. Philosophical transactions of the Royal Society of London Series A, containing papers of a mathematical or physical character. 1922;222(594-604):309-68.

[30] Tae Kyun K. Understanding one-way ANOVA using conceptual figures. Korean Journal of Anesthesiology. 2017;70(1):22 26.

[31] Scheffé H. A" mixed model" for the analysis of variance. The Annals of Mathematical Statistics. 1956:23-36.

[32] Brown AM. A new software for carrying out one-way ANOVA post hoc tests. Computer Methods and Programs in Biomedicine. 2005;79(1):89-95.

Appendix A: Text Excerpts used in the Survey

"The Jane Austen Book Club" by Karen Joy Fowler
"We sat in a circle on Jocelyn's screened porch at dusk, drinking cold sun tea, surrounded by the smell of her twelve acres of fresh-mowed California grass. There was a very pretty view. The sunset had been a spectacular dash of purple, and now the Berryessa mountains were shadowed in the west. Due south in the springtime, but not the summer, was a stream. "Just listen to the frogs," Jocelyn said. We listened. Apparently, somewhere beneath the clamor of her kennel of barking dogs was a chorus of frogs. She introduced us all to Grigg. He had brought the Gramercy edition of the complete novels, which suggested that Austen was merely a recent whim. We really could not approve of someone who showed up with an obviously new book, of someone who had the complete novels on his lap when only Emma was under discussion. Whenever he first spoke, whatever he said, one of us would have to put him in his place. This person would not be Bernadette. Though she'd been the one to request girls only, she had the best heart in the world; we weren't surprised that she was making Grigg welcome. "It's so lovely to see a man taking an interest in Miss Austen," she told him. "Delightful to get the male perspective. We're so pleased that you're here." Bernadette never said anything once if it could be said three times. Sometimes this was annoying, but mostly it was restful. When she'd arrived, she seemed to have a large bat hanging over her ear. It was just a leaf, and Jocelyn removed it as they hugged. Jocelyn had two portable heaters going, and the porch hummed cozily. There were Indian rugs and Spanish-tile floors of a red that might hide dog hair, depending on the breed. There were porcelain lamps in the shape of

ginger jars, round and Oriental, and with none of the usual dust on the bulbs, because it was Jocelyn's house. The lamps were on timers. When it was sufficiently dark out, at the perfect moment, they would snap on all at once like a choir. This hadn't happened yet, but we were looking forward to it. Maybe someone would be saying something brilliant."

"A Madman's Manuscript" by Charles Dickens
"Yes! - a madman's! How that word would have struck to my heart, many years ago! How it would have roused the terror that used to come upon me sometimes, sending the blood hissing and tingling through my veins, till the cold dew of fear stood in large drops upon my skin, and my knees knocked together with fright! I like it now though. It's a fine name. Show me the monarch whose angry frown was ever feared like the glare of a madman's eye - whose cord and axe were ever half so sure as a madman's gripe. Ho! ho! It's a grand thing to be mad! to be peeped at like a wild lion through the iron bars - to gnash one's teeth and howl, through the long still night, to the merry ring of a heavy chain and to roll and twine among the straw, transported with such brave music. Hurrah for the madhouse! Oh, it's a rare place! 'I remember days when I was afraid of being mad; when I used to start from my sleep, and fall upon my knees, and pray to be spared from the curse of my race; when I rushed from the sight of merriment or happiness, to hide myself in some lonely place, and spend the weary hours in watching the progress of the fever that was to consume my brain. I knew that madness was mixed up with my very blood, and the marrow of my bones! that one generation had passed away without the pestilence appearing among them, and that I was the first in whom it would revive. I knew it must be so: that so it always had been, and so it ever would be: and when I cowered in some obscure corner of a crowded room, and saw men whisper, and point, and turn their eyes towards me, I knew they were telling each other of the doomed madman; and I slunk away again to mope in solitude. 'I did this for years; long, long years they were. The nights here are long sometimes - very long; but they are nothing to the restless nights, and dreadful dreams I had at that time. It makes me cold to remember them. Large dusky forms with sly and jeering faces crouched in the corners of the room, and bent over my bed at night, tempting me to madness. They told me in low whispers, that the floor of the old house in which my father died, was stained with his own blood, shed by his own hand in raging madness. I drove my fingers into my ears, but they screamed into my head till the room rang with it, that in one generation before him the madness slumbered, but that his grandfather had lived for years with his hands fettered to the ground, to prevent his tearing himself to pieces. I knew they told the truth - I knew it well. I had found it out years before, though they had tried to keep it from me. Ha! ha! I was too cunning for them, madman as they thought me. 'At last it came upon me, and I wondered how I could ever have feared it. I could go into the world now, and laugh and shout with the best among them. I knew I was mad, but they did not even suspect it. How I used to hug myself with delight, when I thought of the fine trick I was playing them after their old pointing and leering, when I was not mad, but only dreading that I might one day become so! And

how I used to laugh for joy, when I was alone, and thought how well I kept my secret, and how quickly my kind friends would have fallen from me, if they had known the truth. I could have screamed with ecstasy when I dined alone with some fine roaring fellow, to think how pale he would have turned, and how fast he would have run, if he had known that the dear friend who sat close to him, sharpening a bright, glittering knife, was a madman with all the power, and half the will, to plunge it in his heart. Oh, it was a merry life! 'Riches became mine, wealth poured in upon me, and I rioted in pleasures enhanced a thousandfold to me by the consciousness of my well-kept secret. I inherited an estate. The law - the eagle- eyed law itself - had been deceived, and had handed over disputed thousands to a madman's hands. Where was the wit of the sharp- sighted men of sound mind? Where the dexterity of the lawyers, eager to discover a flaw? The madman's cunning had overreached them all."

"The Lies We Tell" by Jane Corry
"I'd dozed off earlier, despite intending to stay awake, my ears tuned to the sound of our only child tiptoeing or thudding up the stairs, depending on the state of his almost sixteen-year-old hormones. But the neon numbers from my alarm clock on the bedside table now tell me it's 2.53 a.m. A sharp stab of fear pierces the pit of my stomach. Where is he? And why hasn't he texted? I send an Are U OK? Of course, there's no answer. Searching for my slippers in the dark, I edge around the packing boxes marked Main Bedroom and pad across to the wooden sash window. I'm going to miss this old house, despite everything. Outside, in our quiet north London street, the lampposts are spilling their orange light onto the water-filled potholes that the council has promised to repair 'shortly'. It's been the wettest spring for five years, according to the radio. No one is in sight. Not even a car driving past. I crawl back under the duvet, wondering what to do. Freddie's never been this late before. I don't want to wake Tom, but suppose something's happened? I lean over my husband. His back is to me and his shoulders are rising and falling in a steady, solid sleep that matches his character to a T. He's wearing pyjamas, of course, as he always has since I've known him. This pair has blue and white stripes. There's a faint whiff of last night's sex from the sheets; the kind of urgent coupling we have once in a blue moon, as if to prove to ourselves that we're still OK together. We might be if it wasn't for Freddie. Guiltily, I drive the thought out of my mind. No. I won't wake him. It will only cause another argument. Besides, the removal men will be here in the morning to finish off and take us away. It's our clean start. I don't want to mess things up."

"The Storyteller" by Mario Vargas Llosa
"I FIRST became acquainted with the Amazon jungle halfway through 1958, thanks to my friend Rosita Corpancho. Her function at the University of San Marcos was vague; her power unlimited. She prowled among the professors without being one of them, and they all did whatever she asked; thanks to her wiles, doors of officialdom stuck shut were opened and paths of bureaucracy smoothed. "There's a place available for someone on an expedition to the Alto Marañón

that's been organized by the Institute of Linguistics for a Mexican anthropologist," she said to me one day when I ran into her on the campus of the Faculty of Letters. "Would you like to go?" I had finally managed to obtain the fellowship to Europe I'd coveted and was to leave for Spain the following month. But I accepted without a moment's hesitation.Rosita is from Loreto, and if you listen carefully you can still catch in her voice an echo of the delightful singsong accent of eastern Peru. She protected and promoted – as no doubt she still does – the Summer Institute of Linguistics, an organization which, in the forty years of its existence in Peru, has been the object of virulent controversy. I understand that as I write these lines it is packing its bags to leave the country. Not because it has been expelled (though this was on the verge of happening during General Velasco's dictatorship), but on its own initiative, since it considers that it has fulfilled the mission that brought it to Yarinacocha, its base of operations on the banks of the Ucayali, some ten kilometers from Pucallpa, from which it has spread into nearly all the remote folds and corners of Amazonia. What exactly is the purpose of the Institute? According to its enemies, it is a tentacle of American imperialism which, under cover of doing scientific research, has been engaged in gathering intelligence and has taken the first steps toward a neocolonialist penetration of the cultures of the Amazonian Indians. These accusations stem, first and foremost, from the Left. But certain sectors of the Catholic Church – mainly the jungle missionaries – are also hostile to it and accuse it of being nothing more than a phalanx of Protestant evangelists passing themselves off as linguists. Among the anthropologists, there are those who criticize it for perverting the aboriginal cultures, attempting to Westernize them and draw them into a mercantile economy. A number of conservatives disapprove of the presence of the Institute in Peru for nationalist and Hispanist reasons. Among these latter was my professor and academic adviser back in those days, the historian Porras Barrenechea, who, when he heard that I was going on that expedition, solemnly cautioned me: "Be careful. Those gringos will try to buy you." He couldn't bear the thought that, because of the Institute, the jungle Indians would probably learn to speak English before they did Spanish.

"The Garlic Ballads" by Mo Yan
"Emitting one last gasp, the lock snapped, and the gate flew open in the face of a surging tide of people. Poor Gao Yang was swept along, powerless to resist. He hadn't thrown a single bundle of his precious garlic, and was worried that his donkey might get trampled. But he was not even able to look behind him. The crowd carried him along, his feet barely touching the octagonal slabs of cement covering the ground; his face was moistened by an icy spray as he passed the fountain. The crowd surged into the office building, where a grand clatter echoed across the tiled floor, compounded by the crisp tinkle of shattering glass, the thud of splintering cabinets, and the shrieks of terrified women. A sense of ecstasy crept into Gao Yang's mounting anxiety as he saw the destruction of luxurious trappings that induced in him feelings of envy and hatred. As an initial probe, he picked up a flowering cactus in a shallow red-and-pink vase and flung it at a window whose glass was polished until it shone. It parted without a murmur,

allowing the vase and its contents to pass slowly through. He ran to the window in time to see the red-and-pink vase, the green cactus, and shards of window glass dance and skitter across the concrete ground. The vase broke, the detached petals scattered in all directions. A gratifying sight. Then he went back, picked up an oval aquarium, and admired the plump black and orange goldfish for a moment. The sloshing water and filthy debris rising from the bottom alarmed the aquarium's denizens, which began splashing frantically, releasing a fishy odor that he found extremely disagreeable. He flung it against another window, which also disintegrated slowly as he ran up to watch the aquarium float downward, followed by glistening drops of water and sparkling shards of glass. The black and orange goldfish swam in midair. When it hit the concrete below, the aquarium shattered without a sound. Unsettled by the sight of goldfish flapping around on the concrete below, he looked up and saw that the square was alive with people and animals, all in motion. His donkey and wagon were nowhere in sight, he noted with chagrin. Throngs of people poured into the compound, as a phalanx of armed policemen in white uniforms emerged from a lane east of the square and swarmed over it like tigers on a flock of sheep, swinging their batons to clear a path to the compound. He turned away from the window, concentrating on getting out of there as fast as his legs would carry him. But his way was blocked by dozens of people who by then had flocked into the office. He could hardly believe his eyes when he spotted Fourth Aunt Fang, who had hobbled in on tiny bound feet. A youngster in a white vest with an anchor logo shouted, "This is the county administrator's office. Let's hunt him down!" Oh my God! Gao Yang thought, the young man's shout hitting him like a thunderbolt. The county administrator's office! It was his vase, his aquarium, his windows. He would have fled if he could, but there were too many sticks and clubs fanning the air between him and the door. Vases with exotic plants came off the floor and began flying out the windows like so many artillery shells. One of them must have hit someone, if the string of screams and curses below were any indication. Scrolls were ripped off the walls, and one young fellow even smashed a filing cabinet with a dumbbell, sending files, documents, and books tumbling out into a pile. He then used the same dumbbell to smash two telephones on the desk. Meanwhile, Fourth Aunt was grabbing everything in sight, including some green satin curtains, which she pulled down and began tearing to shreds, as if ripping a rival's hair. "Give me back my husband!" she screamed through her tears. "I want my husband back!""