

An Analysis of Written Texts in the Economic Field. Case Study

Alina Buzarna-Tihenea (Gălbează)

“Ovidius” University of Constanta, Faculty of Letters, Romania

alina_buzarna84@yahoo.com

Abstract

This article focuses on the analysis of written texts by text analysis tools as a useful instrument to ESP teachers in the design of their courses and seminars. The first part of this study presents several theoretical aspects related to text analysis, such as the definition of texts and outline of their main characteristics, the roles played by lexical density and cohesiveness in understanding the meanings of texts. The second part represents the practical section and presents the results of three text analysis tools obtained after processing a specialized corpus from the field of economic sciences. These results (in terms of lexical density, readability scores, keywords, most frequently used words) revealed important pieces of information that could assist teachers in developing their students’ reading and writing skills and in choosing their teaching materials.

Key words: text analysis tools, text, lexical density, readability, ESP

J.E.L. classification: Z13

1. Introduction

In their analysis of Halliday and Hasan’s work on cohesion and discourse analysis (which is underlain by the understanding of text structuring mechanisms), Afzaal et al. (2019, p. 79) see language “as a system of meaning accompanied by forms through which the meanings can be expressed”. Texture, which underlies the unity and semantic interdependence of texts, establishes relationships between sentences, and provides meaningfulness, coherence (i.e. the contextual meaning at paragraph level) and cohesion (i.e. the text’s “semantic ties”) (see Eggins, 1994, p. 85). Since “teaching is a challenging endeavor” (Nadrag, 2019, p. 322), understanding how cohesion operates within texts in order to establish semantic connections could be beneficial to the students attending ESP courses and seminars in order to decode meaning.

Furthermore, the analysis of written texts equips teachers with the systematic knowledge necessary in order to describe texts, enabling them to raise their students’ awareness of core characteristics, such as cohesion and coherence. In addition, the understanding of these concepts enhances the students’ writing skills and makes them aware of the main traits that should characterize any well-written text.

2. Theoretical background: written text analysis

It is well-known that the concept of “text” has enriched the system of linguistic levels and has eased the understanding and explanation of certain textual elements, like cohesion and coherence, and their connection with text typology issues. It has also contributed to the clarification of several issues from other related fields, such as translation theory and practice, foreign language teaching, etc. Since the notion of “text” refers to a multidimensional unit (Janaszkiwicz and Różewski, 2019), it cannot be bound to a mere unifying definition; moreover, the elements and the criteria employed in order to define texts differ from one linguist to another, which complicates the situation, expanding the list of available definitions.

In general, linguists agree that this notion makes reference to a stretch of language of some considerable extent that is complete in itself: an invitation letter, a report, a scientific or a news article, etc. Nevertheless, there are also other language units that are perceived as texts as they

fulfill the prerequisite of forming a meaningful whole in their right even if they are made of only one sentence or even of a single word. Such instances are public notices like "NO PARKING" or "NO SMOKING". Therefore, the meaningfulness of texts is not dependent upon their linguistic dimensions, but on their location in certain contexts.

Halliday and Hasan (1976, pp. 1-2) define the "text" in their work *Cohesion in English* as follows:

"Text is used in linguistics to refer to any passage, spoken or written, of whatever length, that does form a unified whole (...). A text is a unit of language in use. It is not a grammatical unit, like a clause or a sentence; and it is not defined by its size. A text is sometimes envisaged to be some kind of super-sentence, a grammatical unit that is larger than a sentence but is related to a sentence in the same way that a sentence is related to a clause, a clause to a group and so on [...]. A text is best regarded as a SEMANTIC unit; a unit not of form but of meaning".

In another work, the two scholars also tackle the text as: "language that is functional (...) Language that is doing some job in some context, as opposed to isolated words or sentences". They also explain that "any instance of living language that is playing some part in a context of situation, we shall call it a text. It may be either spoken or written, or indeed in any other medium of expression that we like to think of" (Halliday and Hasan, 1985, p. 10).

Moreover, Halliday also states that "as a thing in itself, a text is essentially a semantic unit (...). It is not something that can be defined as being just another kind of sentence, only bigger" (Halliday, 1985: 10) and, as such, texts should be approached from a semantic point of view; the semantic significance of a text can be better understood by analysing, among others, lexical and syntactic patterns.

Halliday and Hasan (1985, p. 5) explain that "there is a text and there is other text that accompanies it: text that is 'with', namely the con-text. This notion of what is 'with the text', however, goes beyond what is said and written: it includes other non-verbal signs-on-the total environment in which a text unfolds". These two scholars place emphasis on the part played by language in social interaction, whereby meanings are exchanged; they also tackle the role of situational contexts when approaching texts. Thus, from their perspective, texts are in fact stretches of interconnected sentences whose meanings heavily depend on their contexts.

In their turn, De Beaugrande and Dressler (1981, p. 63) consider that the text is "A naturally occurring manifestation of language, i.e., as a communicative language event in a context. The SURFACE TEXT is the set of expressions actually used; these expressions make some knowledge EXPLICIT, while other knowledge remains IMPLICIT, though still applied during process". From their perspective, there are two text-centered elements that designate operations aimed at text materials: coherence and cohesion. The former deals with "the ways in which the components of the textual world, i.e. the concepts and relations which underlie the surface text are mutually accessible and relevant", while the latter "concerns the ways in which the components of the surface text (the actual words we hear or see) are mutually connected within a sequence" (De Beaugrande and Dressler, 1981, pp. 3-7).

Texts are also characterized by user-centered standards or constitutive principles (Searle, 1965), which outline and establish textual communication and communication rules: informativity; intentionality, acceptability, intertextuality, situationality (De Beaugrande and Dressler, 1981, p. 11). These two scholars also explain that textual communication is controlled by three main regulative principles, i.e. efficiency (its usefulness to the participants with minimum effort); effectiveness (making strong impressions; good potential for achieving certain aims); and appropriateness (the setting should be in accordance with the above-mentioned standards).

Written language has stirred up not only the linguists' interest but also that of language teachers and literary scholars, as written texts are distinct from one another in terms of genre and function, but also as far as their structure and form are concerned; in addition, knowledge of writing organization and variety affects the readers' understanding, memory of conveyed messages and perception speed. Written discourse analysts have also focused on the relationships between neighboring sentences and especially on the elements highlighting that texts are more than the mere sum of their components. By performing written language analysis, certain characteristics of communicative products began to be depicted in adequate ways (McCarthy, 1991, p. 37).

Lexical density, defined by Halliday (1985, p. 67) as "the number of lexical items as a ratio of the number of clauses" focuses mainly of the role played by lexical items (i.e. content words) within a text. Following his analysis, the scholar has reached the conclusion that written texts are endowed with higher lexical density levels; this is due to the fact that the number of non-lexical items and the one of clauses decrease; in addition, he notices that the sentence structure of written texts is less complex (compared to spoken texts, which are characterized by lower lexical density but also by higher levels of grammatical complexity): "The more natural, un-self-monitored the discourse, the more intricate the grammatical patterns that can be woven. Usually, this kind of discourse will be spoken, because writing is in essence a more conscious process than speaking". He also adds that "spoken and written discourse are the outward forms that are typically associated with the critical variable, which is that of consciousness" (Halliday 1987, p. 66).

Thus, from the scholar's viewpoint, spoken language is usually more intricate in terms of grammar and its clauses have fewer lexical items, while written language abounds in lexical items, thus favoring increased lexical density and being less complex in terms of grammar.

Cohesive markers are essential when it comes to text interpretation as they create links across sentence boundaries, pairing and chaining related items. According to McCarthy (1991), reading is a complex process as readers must interpret connections and make sense of them. Furthermore, based on their understanding of the world, they must build the world of the text in an active manner; they have to make assumptions and continuously evaluate their various interpretations in the connection to the situation and aims of the respective text.

When processing texts, readers also have to recognize textual patterns, coined by McCarthy (1991) as textual segments in order to avoid confusion with grammatical elements and syntactic relations within clauses and sentences. These textual segments are deeply rooted in the readers' cultural knowledge and are exhibited in regularly occurring functional relationships between bits of text (i.e. phrases, clauses, sentences, groups of sentences or even whole paragraphs); it is noteworthy that segments can be isolated by employing labels encompassing a finite set of functional relationships that can occur between any two bits of text, such as phenomenon-example, cause-consequence, instrument-achievement, etc. (see Hoey 1983; McKlee, 2001).

In McCarthy's perspective, when interpreting the relations between textual segments, readers perform cognitive acts, as they might ask text-related questions as it unfolds, they might engage in a dialogue with the author, and the processing of textual segments could be perceived as similar to an exchange in spoken discourse. McCarthy (1991) places textual relationships under the heading of logical sequence relations (i.e. the relationships created by textual segments, such as the subordination of one item to another by various grammatical means or the presence of conjunctions or general vocabulary items can create cause-consequence relationships).

McCarthy (1991) has even gone further in his clause-relational approach and analyzed larger patterns or sequences of relations that are culturally rooted and usually occur in texts (for instance, problem-solution patterns) being often emphasized by grammatical and lexical mechanisms such as subordination and parallelism. Both readers and writers should be aware of these mechanisms and employ them when processing textual relationships that are not readily evident.

3. Research methodology

Having in view that knowledge of writing organization affects, among others, the readers' understanding and perception speed, we considered that the analysis of written texts may represent a useful tool that would enable ESP teachers to enhance the learners' writing and reading skills, by raising their awareness of the main text characteristics (such as coherence and cohesion) and of the core standards that should be met by any well-written text. This is extremely useful since specialized text are even more difficult to understand, having in view "the interfingering between different types of specialized vocabularies and the common vocabulary in use" (Nadrag, 2016, p. 37) that characterize them.

Thus, this section of the paper is aimed at analyzing several written text features in the field of economic sciences, with a focus on lexical density, keywords, most frequently used words and readability, by means of specialized text analysis tools, i.e. "SEO Scout" (<https://seoscout.com/tools/keyword-analyzer>), "Text Analyzer" (<https://www.online->

utility.org/text/analyzer.jsp) and “Analyze My Writing” (<https://www.analyzemywriting.com/>). For the purpose of our research, an authentic text was selected, i.e. a EU Directive that should be consulted by the employees of large companies in order to increase awareness on various issues related to corporate social responsibility – “DIRECTIVE 2014/95/EU OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL of 22 October 2014 amending Directive 2013/34/EU as regards disclosure of non-financial and diversity information by certain large undertakings and groups”.

4. Findings

First of all, a content analysis was performed by “SEO Scout”. The results are shown in the table below:

Table no. 1. Content Analysis by SEO Scout

Word Count	4718
Character Length	30718
Letters	24972
Sentences	397
Syllables	8548
Average Words/Sentence	11.9
Average Syllables/Word	1.9
Lexical Density	54%
Lexical Diversity	21%

Source: Table processed by “SEO Scout” (<https://seoscout.com/tools/keyword-analyzer>)

According to Castello (2008, pp. 49–51), the texts written in the English language are normally characterized by lexical densities above 40%; additionally, the lexical density of non-fiction texts range between 40% and 65% (see Stubbs, 1986, pp. 27–42). It may be noticed that, according to the results issued by “SEO Scout”, the lexical density of the selected corpus is quite high, i.e. 54%, which is typical of informative or technical texts. This fact highlights that the corpus is aimed at transmitting a quite large amount of information, which makes it more complex and difficult to retain.

The specialized text analysis tool assessed the readability of the corpus, and assisted us in determining its difficulty level in terms of understanding the message that it conveys. Word choice, sentence length and structure, average syllables per word are among the elements that can influence a text’s readability score. When it comes to regulations, readability plays a significant role as it can influence the readers’ understanding of and engagement with content.

Table no. 2. Readability by “SEO Scout”

Reading Ease	36.8%
Grade Level	11.1
Gunning Fog	14.4
Coleman Liau Index	15.4
Smog Index	10.7
Automated Reading Index	9.4

Source: Table processed by “SEO Scout” (<https://seoscout.com/tools/keyword-analyzer>)

Table no. 2 shows the readability scores of the analyzed corpus, processed by “SEO Scout”. These scores indicate that the text is difficult to read and that it can be best understood by college sophomores or by college graduates. Therefore, this document can be used as a teaching material in the ESP courses or seminars with the students majoring in Economic Sciences, enrolled in the second or third year of study.

Table no. 3 below, processed by the text analysis tool, includes a top 10 of the keywords in the corpus.

Table no. 3. Keywords by “SEO Scout”

Keyword	Uses	Uses
Undertakings	50	1.1%
Information	47	1.0%
non-financial	45	1.0%
Article	37	0.8%
Directive	35	0.7%
Statement	34	0.7%
Paragraph	34	0.7%
Undertaking	33	0.7%
Report	32	0.7%
European	31	0.7%

Source: Table processed by “SEO Scout” (<https://seoscout.com/tools/keyword-analyzer>)

It is noteworthy that almost all keywords are nouns (except the adjectives “non-financial” and “European”) usually encountered in the economic or legal field.

Another interesting processing was performed by “Text Analyzer”, which dealt with the frequency of words.

Table no. 4. Frequencies of words by “Text Analyzer”. Top 30

order	unfiltered word count	occurrences	Percentage
1.	the	367	7.8503
2.	of	206	4.4064
3.	and	174	3.7219
4.	to	167	3.5722
5.	in	127	2.7166
6.	a	69	1.4759
7.	financial	59	1.2620
8.	non	49	1.0481
9.	information	47	1.0053
10.	undertakings	46	0.9840
11.	by	45	0.9626
12.	for	44	0.9412
13.	l	42	0.8984
14.	on	41	0.8770
15.	article	40	0.8556
16.	shall	39	0.8342
17.	this	38	0.8128
18.	that	36	0.7701
19.	be	35	0.7487
20.	directive	35	0.7487
21.	as	34	0.7273
22.	is	34	0.7273
23.	statement	34	0.7273
24.	paragraph	34	0.7273
25.	or	33	0.7059
26.	with	33	0.7059
27.	should	32	0.6845

28.	report	32	0.6845
29.	european	31	0.6631
30.	member	28	0.5989

Source: Table processed by “Text Analyzer” (<https://www.online-utility.org/text/analyzer.jsp>)

Surprisingly, although most keywords are nouns, the most frequently used words in the text are grammatical items, such as definite and indefinite articles (“the”, “a”), prepositions (“of”, “to”, “in”, “by”, “for”, “on”, “with”), conjunctions (“and”, “or”), pronouns (“this”, “that”), auxiliaries (for instance, “shall”). These words are usually used in order to highlight various relationships between lexical words and phrases. It should be noted that the most frequently used lexical words are adjectives (“financial”, “European”) and nouns (“information”, “undertakings”, “article”, “directive”, “statement”, “paragraph”, “report”, “member”). The main role of these lexical items is to convey meaning within the text.

Although the most frequently used words in the text are grammatical items, the text analysis tool “Analyze My Writing” shows that lexical items hold a greater share in the corpus:

Table no. 5. Shares held by parts of speech in the corpus (by “Analyze My Writing”)

Part of Speech	Percentage
Nouns	31.17%
Adjectives	9.87%
Verbs	10.06%
Adverbs	2.13%
Prepositions	14.33%
Pronouns	0.47%
Auxiliary Verbs	2.89%

Source: Table processed by “Analyze My Writing” (<https://www.analyzemywriting.com/>)

The statistics from the table above highlight that more than half of the parts of speech in the text are lexical items, showing increased lexical density, typical of written texts.

In its turn, lexical density, which is “the number of lexical items as proportion of the number of running word” (Halliday, 1985: 64), deals with the structure and complexity of communication, estimating the linguistic complexity of a text and influencing its readability, memorability and retention. Furthermore, scholars explain that written English texts usually have lexical densities above 40% (Castello, 2008, pp. 49–51), with the non-fiction ones ranging between 40% and 65% (Stubbs, 1986, pp. 27–42). The lexical density of the corpus is quite high, i.e. 56%, typical of expository writing (informative or technical texts). Therefore, the analyzed text conveys a quite large amount of information, which hinders its retention and enhances its complexity.

5. Conclusions

The analysis of written texts by text analysis tools can be extremely useful to ESP teachers when they select their teaching materials. These instruments can indicate critical pieces of information, in terms of text difficulty levels, such as lexical density, readability scores, keywords, most frequently used words.

Furthermore, the understanding of the meanings conveyed by written texts can be enhanced by knowledge of writing organization, which makes text analysis a useful instrument in teaching and developing the students’ writing and reading skills. Thus, awareness of core features such as coherence and cohesion, and knowledge of the characteristics of well-written texts could underlie the development of good readers and writers of English texts.

6. References

- “Analyze My Writing”, [online] Available <<https://www.analyzemywriting.com/>> [Accessed 10th November 2020]
- Afzaal, M., Kaibao Hu, Ilyas Chishti, M. and Imran, M. 2019. “A Study of Pakistani English Newspaper Texts: An Application of Halliday and Hasan’s Model of Cohesion: A Discourse Analysis”. *“International Journal of English Linguistics”*, 9(5), pp. 78-88
- Beaugrand, R., Dressler, W. U. 1981. *“Introduction to Text Linguistics”*. London: Longman.
- Castello, E. 2008. *“Text Complexity and Reading Comprehension Tests”*. Peter Lang
- “Directive 2014/95/EU of the European Parliament and of the Council of 22 October 2014 amending Directive 2013/34/EU as regards disclosure of non-financial and diversity information by certain large undertakings and groups”, [online] Available at <<https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A32014L0095>> [Accessed 10th November 2020]
- Eggins, S. 1994. *“An Introduction to Systemic Functional Linguistics”*. London: St. Martin’s Press.
- Halliday, A. K. M., Hassan, R. 1976. *“Cohesion in English”*. Longman.
- Halliday, M.A.K. 1985. *“Spoken and Written Language”*. Victoria: Deakin University Press.
- Halliday, A.K. M. 1985. *“An Introduction to Functional Grammar”*. London: Arnold.
- Halliday, M.A.K. & Hasan, R. 1985. *“Language, Context and Text: Aspects of Language in a Social-Semiotic Perspective”*. Victoria: Deakin University
- Halliday, M.A.K. 1987. “Spoken and Written Modes of Meaning”. In: Horowitz, Rosalind & Samuels, S. Jay (eds.) *“Comprehending Oral and Written Language”*
- Hoey, M. P. 1983. *On the Surface of Discourse*. London: George Allen and Unwin.
- Janaszkiwicz, P., Rózewski, P. 2019. “The method of multidimensional approach to text summarization”. *“Procedia Computer Science”*, 159, pp. 2189-2196.
- McCarthy. M. 1991. *Discourse analysis for language teachers*. Cambridge. Cambridge University Press.
- McKlee, A. 2001. *Textual Analysis: A Beginner’s Guide*, University of Technology Syney, 127, pp. 139-149
- Nadrag, L. 2019. “Teaching Business English to Non Economics Majors”. *Ovidius University Annals, Economic Sciences Series*, XIX(2), pp. 322-326
- Nadrag, L. 2016. “Teaching Legal English: A Contrastive Analysis (Romanian-English)”. *“The Proceedings of the International Conference Globalization, Intercultural Dialogue and National Identity”*, Boldea, I. (ed.), Vol. III, Section: Language and Discourse. Arhipelag XXI Press, Tirgu Mures, pp. 36-48
- Searle, J. *Speech Acts*. London: Cambridge University Press, 1969.
- “SEO Scout”, [online] Available at <<https://seoscout.com/tools/keyword-analyzer>> [Accessed 10th November 2020]
- Stubbs, M. 1986. “Lexical density: A technique and some findings”. In Malcolm Coulthard (ed.). *Talking about Text*. University of Birmingham: English Language Research. pp. 27–42.
- “Text Analyzer”, [online] Available at <<https://www.online-utility.org/text/analyzer.jsp>> [Accessed 10th November 2020]