

Text Analysis Tools in ESP Teaching. Case Study

Alina Buzarna-Tihenea (Gălbează)

“Ovidius” University of Constanta, Faculty of Letters, Romania
alina_buzarna84@yahoo.com

Abstract

This study, structured into two main sections (a theoretical part and a practical one) deals with the important role played by text analysis tools in ESP courses and seminars, when choosing appropriate teaching materials. The first section of the paper tackles theoretical concepts related to lexical cohesion, reiteration and collocations, and the roles played by vocabulary within discourse. The practical section analyses the results issued by two text analysis tools, i.e. Voyant Tools and SEO Scout, which processed an authentic corpus from the economic field, showing, among others, its lexical density, lexical diversity and collocates. These results provide vital information about the complexity of the analyzed corpus and about its appropriate use in ESP classes.

Key words: text analysis tools, ESP, teaching, economic language, lexical cohesion

J.E.L. classification: Z13

1. Introduction

In ESP teaching and learning, the specialized vocabulary plays a vital part as far as the efficient acquisition of the reading, writing, listening and speaking skills is concerned (see Nadrag, 2016, pp. 36-37). Having in view that ESP texts are usually aimed at specific audiences and that they are also focused on the achievement of particular purposes, by using text analysis tools, one can better understand natural language processing (with a focus on specialized fields). Moreover, if teachers employ these tools, appropriate authentic discourse samples can be identified and used more efficiently in ESP teaching. Thus, text analysis (usually performed by various specialized software) can support teachers to improve the (ESP) teaching and learning processes, especially in terms of syllabus creation, choice of appropriate classroom materials and teaching methods, techniques and activities.

Scholars such as Halliday and Hassan (1976), Halliday (1985; 1989), Cook (1989), McCarthy (1991), McCarthy and Carter (1994), De Celce-Murcia and Olshtain (2000) tackled in their research the contribution brought by text/ discourse analysis to language teaching, which marked a shift from Saussure’s traditional approach focusing on “langue” to the one that deals with actual language products and meaning creation, professing that language learning cannot take place out of context. In this regard, text analysis tools deconstruct meaning creation in order to show how language functions in a certain communicative environment. Learners acquire language skills in order to be able to communicate in various contexts (Nadrag, 2019, pp. 322-324). Thus, ESP teaching and learning processes should also focus on how a text makes sense, on how it holds together; consequently, the main purpose of grammar and vocabulary teaching is to help learners understand the contribution of these skills to the cohesion and coherence of a text and to the construction of meaning.

In order to better understand the semantic and pragmatic functions of a text, its lexical and grammatical features should be analyzed. Additionally, this knowledge should be applied to the ESP teaching and learning process (especially in terms of curriculum design, lesson materials and assessment), in order to enhance the learners’ language skills (with focus on their field of specialization). Thus, the results provided by text analysis tools can contribute to language awareness and improve the quality and relevance of ESP classes.

2. Theoretical background: roles played by vocabulary within discourse and lexical cohesion

Across time, the linguists' opinions on the roles played by syntax and vocabulary in language learning varied largely, ranging from the focus on syntax professed in formal linguistics (which viewed languages as governed by rules that native speakers apply in creative and original ways) (see Chomsky, 1965) to the focus on vocabulary, supported by those scholars who argued that both everyday language and the professional one are quite predictable, formulaic and dominated by routine (see Hoey, 1983; Nattinger and DeCarrico, 1992). Since word meanings heavily depend on context (Celce-Murcia, 2000), it can be deduced that knowledge of a certain word implies the skills needed in order to properly use it in a syntactic, semantic and pragmatic way. In this regard, McCarthy (1984 p. 14) tackles several approaches to vocabulary teaching based on the functions fulfilled by the order of lexical items, showing the role played by the lexis:

"The belief that vocabulary skill is clearly more than understanding the componential features of words and recognizing their typical collocations, more than the ability to define a word and slot it into a sentence, leads me to propose that the key to a new approach in vocabulary teaching lies in an examination of syntagmatic and paradigmatic relations of collocation and set between lexical items:

above sentence level

across conversational turn-boundaries

and within the broad framework of discourse organization". (McCarthy 1984, pp.14)

Therefore, from McCarthy's perspective (1991; 1994), the relations formed at the level of collocations and sets of lexical items are constantly influenced by a dynamic reclassification process. The role played by lexis in written discourses, especially as far as the creation of cohesive textual relations is concerned, has also been revealed through exercises demanding (re)ordering the sentences and paragraphs.

The research conducted on lexical cohesion has also emphasized the importance of teaching lexis in discourse. In this regard, as far as vocabulary teaching is concerned, Crombie (2010) professes that more emphasis should be placed on conjunctions and learners should also be provided with opportunities to discern the roles that lexical items play when creating of semantic relationships.

It is noteworthy that speakers are not constantly required to creatively negotiate and renegotiate at interpersonal level when using vocabulary; the acquisition of fixed expressions can also contribute to the maintenance of discourse relationships (see Nattinger and DeCarrico, 1992). Thus, in order to develop lexical discourse skills, fixed expressions should also be taken into consideration as they create discourse relationships. Moreover, by providing a relatively stereotyped and stable response to an event considered as repetitive and formulaic, they also serve communicative purposes. Since they are predictable and usually contextually "fixed", these so-called "prefabricated discourse-sensitive units" reduce the interlocutors' effort of creating new lexical meaning, playing thus an essential role in discourse construction (see Carter, 2002, p. 223).

As far as lexical cohesion is concerned, some research performed by linguists (Pomerantz, 1984; Pearson, 1986) showed that conversational functions involving agreement/disagreement patterns can be fulfilled by synonyms, hyponyms and antonyms. For instance, Pearson (1986) emphasized that agreement or disagreement is not typically expressed by phrases such as "I agree" or "I disagree" but rather by using certain lexical relations between turns.

According to McCarthy (1991, p. 65), "related vocabulary items occur across clause and sentence boundaries in written texts", which represents a major feature of a coherent and meaningful discourse. The above-mentioned scholar quotes Halliday and Hassan (1976), who studied vocabulary patterns, described lexical cohesion, analyzed various types of lexical relationships that occur within a discourse, and reached the conclusion that "the relations between vocabulary items in texts (...) are of two main kinds: reiteration and collocation" (McCarthy, 1991, p. 65).

Regarding reiteration, in English "it is necessary to vary repetitions and use different synonyms or appropriately related words in order to achieve good style" having in view that "all languages have a stock of general words" that are "highly frequent" and "tend to occur more often in informal spoken discourse" in order to make cohesion (De Celce-Murcia and Olshtain, 2000, p. 83).

Furthermore, these two scholars also explain that in any type of discourse, interacting lexical chains (with words whose meanings range from general to specific) are used, which create “cross-clausal lexical relationships” (p. 84) and enhance text cohesiveness and coherence. Thus, these scholars conclude that the analysis of lexical chains (in connection to semantic relationships) increases the awareness of the role played by lexicons in discourse cohesiveness and coherence.

In their turn, collocations are defined by De Celce-Murcia and Olshtain (2000, p. 83) as words that combine at the sentence level forming semantic and structural connections that become routines or chunks accessed for comprehension and production. They are also essential in discourse analysis, as they “reflect both local word-combining tendencies” typical of any language, and “more general content schemata or information structures” shared by all users of each language (De Celce-Murcia and Olshtain, 2000, p. 83). Carter (2002, p. 81) describes collocations as “recognizable lexico-semantic word-meaning relation”; in addition, he adds that “it is extremely difficult to define in any systematic way the nature of such collocational relations because (...) some patterns are distinctly semantico-syntactic and others are more generally probabilistic”.

McCarthy (1991) suggests that it is extremely useful to train language learners in associating synonyms and antonyms (by means of simple cue and response drills, for instance). Awareness of the communicative value of lexical relations – which can begin at an early language learning stage, as soon as the necessary vocabulary is acquired – can enhance language learning. Furthermore, within a discourse, in order to achieve coherence, lexical items should be selected based on the shared knowledge about the topic of the respective discourse, its modality and genre, its register and audience.

3. Research methodology

During ESP courses and seminars dedicated to the students majoring in economic sciences, the teacher should create opportunities for students to acquire effective English written and oral communication skills and to understand specialized documents, from this professional field (i.e. economic sciences). Therefore, students must develop their understanding of the language specific to the field of economics, in order to use English effectively in their future profession. The teacher should be able to prioritize what to teach and, for this purpose, we consider useful the analysis of some documents in the economic field; this text analysis (performed in our study by specialized software) will highlight the type of language used, lexical density, lexical diversity, collocations, etc.

For the purpose of this study, we have chosen to analyze an authentic text, i.e. an EU Directive that should be taken into account by companies when drafting some of their most important reports (such as consolidated financial statements, various types of reports, annual financial statements, etc.) – “Directive 2013/34/EU of the European Parliament and of the Council of 26 June 2013 on the annual financial statements, consolidated financial statements and related reports of certain types of undertakings, amending Directive 2006/43/EC of the European Parliament and of the Council and repealing Council Directives 78/660/EEC and 83/349/EEC (Text with EEA relevance)”. The text analysis was performed by two specialized software, i.e. Voyant Tools (“a web-based reading and analysis environment for digital texts”, see <https://voyant-tools.org/>) and SEO Scout (<https://seoscout.com/tools/keyword-analyzer>).

4. Findings

The analysis conducted by the Voyant tools specialized software showed that the chosen text consists of 31,526 total words and 2,289 unique word forms, with 43.6 average words per sentence and a vocabulary density of 0.073.

The most frequent words in the corpus are “article” (764), “financial” (394), “shall” (295), “undertaking/s” (568), “statements” (247), “member” (230) and “point” (201). A hierarchy of the words that occur most frequently in the corpus (top 20) was issued by the specialized software as follows:

Table no. 1. “Most frequent words in the corpus”

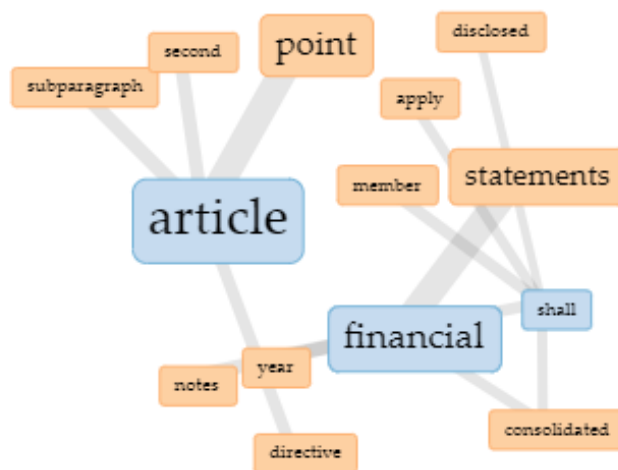
Term	Count
Article	764
Financial	394
Shall	295
Undertakings	290
Undertaking	278
Statements	247
Member	230
Point	201
consolidated	198
directive	171
States	163
European	108
Assets	106
Balance	106
Sheet	103
Value	101
Report	100
Information	99
Accordance	94
profit	91

Source: Table processed by Voyant Tools (see <https://voyant-tools.org/>)

The table above shows that 16 words out of 20 are nouns; three are adjectives (i.e. “financial”, “consolidated” and “European”) and one auxiliary verb (“shall”). This reveals that nouns contribute heavily to text cohesion. Moreover, most of these nouns denote economic concepts: “undertaking”, “statements”, “assets”, “balance”, “sheet”, “value”, “report”, “profit”.

The specialized software also generated a word cluster revealing the most frequent word combinations. It is worth mentioning that these are typical of the economic and legal fields:

Figure no. 1. Word cluster



Source: Figure processed by Voyant Tools (see <https://voyant-tools.org/>)

In addition, the Voyant Tools software issued a table with Corpus Collocates, showing the terms that occur “more frequently in proximity to keywords across the entire corpus” (see <https://voyant-tools.org/>).

Table no. 2. “Corpus Collocates”

Term	Collocate	Count (context)
Financial	Statements	245
Member	States	163
Consolidated	Financial	122
financial	Year	58
Member	State	58
consolidated	Management	28
Financial	Instruments	21
Undertakings	Included	19
Statements	Shall	19
Financial	position	17
shall	Apply	17
Consolidated	balance	13
Undertaking	Concerned	12
Undertakings	Shall	10
Undertaking	Shall	10
Undertaking	governed	10
directive	directive	10

Source: Table processed by Voyant Tools (see <https://voyant-tools.org/>)

It is noteworthy that all the collocates in the above table belong to the economic and legal field, which reveals that the two areas are tightly interconnected in the analyzed corpus. Moreover, this also indicates that the text is highly specialized and it can be fully understood by the professionals activating in these two fields. Therefore, this text is not recommended to the students enrolled in the first year of study. However, some excerpts from this document could be used in the ESP courses or seminars designed for the students enrolled in the second or third year of study.

The complexity of the analyzed corpus is also highlighted by the following figure (generated by the Voyant tools software), which indicates that the longest sentence has 48 words (there are two such sentences). The figure shows a top of the longest sentences, which range from 22 to 48 words.

Figure no. 2. “Top 20 of the longest sentences”

Phrases	Count	Len
<input type="checkbox"/> it is of particular importance that the commission carry out appropriate consultations during its preparatory work including at expert level the commission when preparing and drawing u...	2	48
<input type="checkbox"/> the nature of which is clearly defined and which at the balance sheet date are either likely to be incurred or certain to be incurred but uncertain as to their amount or as to the date on w...	2	41
<input type="checkbox"/> of parent and subsidiary undertakings to be included in a consolidation and which on a consolidated basis do not exceed the limits of at least two of the three following criteria on the b...	2	39
<input type="checkbox"/> shall be disclosed separately in the consolidated balance sheet or in the notes to the consolidated financial statements that difference shall be calculated as at the date on which that m...	2	37
<input type="checkbox"/> would be seriously prejudicial to any of the undertakings to which it relates member states may make such omissions subject to prior administrative or judicial authorisation any such o...	2	37
<input type="checkbox"/> any payment whether made as a single payment or as a series of related payments need not be taken into account in the report if it is below eur 100 000 within a financial year	2	35
<input type="checkbox"/> at least two of the three following criteria a balance sheet total eur 20 000 000 b net turnover eur 40 000 000 c average number of employees during the financial year 250	2	33
<input type="checkbox"/> capital and reserves i subscribed capital unless national law provides that called up capital is to be shown under this item in which case the amounts of subscribed capital and paid up ...	2	33
<input type="checkbox"/> a the exempted undertaking and without prejudice to paragraph 9 all of its subsidiary undertakings are consolidated in the financial statements of a larger body of undertakings	2	27
<input type="checkbox"/> applying the exchange rate published in the official journal of the european union as at the date of the entry into force of any directive	2	25
<input type="checkbox"/> is disclosed in the notes to the financial statements or in accordance with point b of this paragraph at the foot of the balance sheet	2	25
<input type="checkbox"/> annex iii a article 9 b annex iii b article 9 c annex iii c article 9 d annex iii d article 9 e	2	24
<input type="checkbox"/> from the obligation to draw up consolidated financial statements and a consolidated management report except where any affiliated undertaking is a public interest entity	2	24
<input type="checkbox"/> the operational activities that are governed by a single contract license lease concession or similar legal agreements and form the basis for payment liabilities	2	24
<input type="checkbox"/> a subscribed capital unpaid of which there has been called unless national law provides that called up capital is to be shown under	2	23
<input type="checkbox"/> be shown separately for each item 1 trade debtors 2 amounts owed by affiliated undertakings 3 amounts owed by undertakings with which the	2	23
<input type="checkbox"/> c i 3 c ii 1 2 3 and 4 c iii 1 2 3 and 4 d ii 2 3 and 6	2	23
<input type="checkbox"/> the subsidiary undertaking as laid down by the laws of the member state in accordance with chapter 2 of directive 2009 101 ec	2	23
<input type="checkbox"/> by the undertakings included in the consolidation or by persons acting in their own names but on behalf of those undertakings and	2	22
<input type="checkbox"/> income from other investments and loans forming part of the fixed assets with a separate indication of that derived from affiliated undertakings	2	22

Source: Figure processed by Voyant Tools (<https://voyant-tools.org/>)

In order to measure the lexical density and the lexical diversity of the corpus, the specialized software SEO Scout was also used, and it highlighted the following data in terms of content analysis:

Table no. 3. Content Analysis

Character Length	196206
Letters	153324
Sentences	4766
Syllables	52133
Average Words/Sentence	6.6
Average Syllables/Word	1.8
Lexical Density	56%
Lexical Diversity	9%

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

It should be noted that lexical diversity deals with the variety of lexical words that convey meaning (i.e. nouns, adjectives, verbs and adverbs) within a text and it is one of the elements that can indicate text complexity and readability. Furthermore, lexical words should not be confused with grammatical items (such as conjunctions, pronouns and articles), which (among many others) aim at highlighting various relationships (also see <https://textinspector.com/help/lexical-diversity/>). Duran et al. (2004, p. 220-242) associate lexical diversity with other concepts, such as "flexibility", "vocabulary richness", "verbal creativity", "lexical range and balance", which "indicate that it has to do with how vocabulary is deployed as well as how large the vocabulary might be". The lexical diversity of the text is quite low, i.e. 9% (which is often encountered in specialized 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 the results issued by the two specialized software, i.e. Voyant Tools and SEO Scout, was extremely useful, as it revealed vital pieces of information regarding the appropriate use of the analyzed corpus in ESP seminars and courses. Students should be exposed to as many authentic texts as possible, from their field of specialization, in order to enhance their language skills, with a focus on the respective professional field. However, ESP teachers should take into account elements such as lexical density, lexical diversity and lexical cohesion when choosing their course or seminar materials, in order to make appropriate decisions, as these materials should be selected in accordance with the students' level of knowledge in the specialized field and with their language level.

6. References

- Carter, R. 2002. *Vocabulary: Applied Linguistic Perspectives*. London: Routledge.
- Castello, E. 2008. "Text Complexity and Reading Comprehension Tests". Peter Lang.
- De Celce-Murcia, M, Olshtain, E. 2000. "Discourse and Context in Language Teaching: A Guide for Language Teachers". Cambridge: Cambridge University Press.
- Chomsky, N. 1965. *Aspects of the theory of syntax*. Cambridge: The MIT Press.
- Cook, G. 1989. *Discourse*, Oxford University Press.

- Crombie, W. 2010. *Genre-based writing instruction: Towards an integrated model*. University of Waikato.
- "Directive 2013/34/EU of the European Parliament and of the Council of 26 June 2013 on the annual financial statements, consolidated financial statements and related reports of certain types of undertakings, amending Directive 2006/43/EC of the European Parliament and of the Council and repealing Council Directives 78/660/EEC and 83/349/EEC (Text with EEA relevance)", *EUR-LEX*, [online] Available at <<https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex%3A32013L0034>> [Accessed 10th November 2020]
- Duran, P., Malvern, D., Richards, D., Chipere, N. 2004. "Developmental Trends in Lexical Diversity". *Applied Linguistics OUP* 2(2), pp. 220-242.
- Halliday, A. K. M. 1989. "*Spoken and Written Language*". Oxford: Oxford University Press.
- 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, A. K. M., and Hassan, R. 1976. *Cohesion in English*. Longman.
- Hoey, M. P. 1983. *On the Surface of Discourse*. London: George Allen and Unwin.
- McCarthy. M. 1991. *Discourse analysis for language teachers*. Cambridge. Cambridge University Press.
- McCarthy. M. 1984. "A New Look at Vocabulary in EFL". *Applied Linguistics*, 5(1), pp. 12–22.
- McCarthy M., and Carter, R. 1994. "*Language as discourse: Perspectives for language teaching*". Longman Publishing, New York.
- 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
- Nattinger J. R., and DeCarrico, J. S. 1992. "*Lexical phrases and language teaching*". Oxford University Press.
- Pearson, E. 1986. "Agreement/disagreement: an example of results of discourse analysis applied to the oral English classroom". "*International Review of Applied Linguistics*", 74
- Pomerantz, A. 1984. "Pursuing a response". In J. M. Atkinson & J. Heritage (Eds.), *Structures of Social Action*. Cambridge: Cambridge University Press.
- SEO Scout, [online] Available at <<https://seoscout.com/tools/keyword-analyzer>> [Accessed 5th 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 Inspector, [online] Available at <<https://textinspector.com/help/lexical-diversity/>> [Accessed 5th November 2020]
- Voyant Tools, [online] Available at <<https://voyant-tools.org/>> [Accessed 5th November 2020]