Pearson Correlation Coefficient in Studying the Meaning of a Literary Text

Oksana Melnychuk [1], Ivan Bekhta [1,2] and Mariia Tkachivska [3]

[1] *Lviv Polytechnic National University, Stepan Bandera Street, 12, Lviv, 79000, Ukraine*
[2] *Ivan Franko National University of Lviv, Universytetska Street, 1, Lviv, 79000, Ukraine*
[3] *Vasyl Stefanyk Precarpathian National University, Shevchenko Street, 57, Ivano-Frankivsk, 76018, Ukraine*

Abstract

This research constitutes the engagement of Pearson correlation coefficient in studying the meaning of a literary text through the statistical textual analysis – correlation of words as parts of speech under the limits of the structure of a literary text (narrative): *Subject (proper nouns)* → *Action (verbs)* → *Object (common nouns)* → *Description / Evaluation (adverbs /adjectives)*. Pearson correlation coefficient is used to establish ties between the most frequent words in corpora (expose general structure) and correlated words (declare meaningful components of the structure) in terms of parts of speech categories. Quantitative data proves the significance of formal structure, which is the initial stage in the multidimensional process of interpreting the meaning of a literary text (narrative). The most frequent words found in two researched corpora – A. Byatt's novels: "Children's book" and "Possession: a romance" – constitute general textual structure, bringing to light connection with correlated words as parts of speech, and their merit. As far as parts of speech are meaningful within the sentence structure they are able to form definite "structure skeleton" in a literary text (narrative) beyond individual author's lexical choice. Quantitative data and their computer processing ensure the disclosure of the meaning of a literary text as a logical process that operates on statistics.

Keywords

Pearson correlation coefficient, meaning, literary text, parts of speech, correlation, frequency

1. Introduction

Analysis of the meaning of a literary text (narrative) requires close attention to its language and also the way it is built under the scope of computer-based discourse analysis, especially statistical textual analysis[1, 2]. The idea of textual analysis involving its structure is not new. It starts in structural linguistics flourishing with a number of formal models proposed by G. Genette, A.- J. Greimas, V. Propp. Linguists, defining literary text (narrative) as a "sequence of events or actions" payed attention to a core structural element – *an action* expressed by verbs. They argued that action is always connected to the one who does it – a doer or *subject* expressed by proper nouns or pronouns. Than the action is directed on *an object* (common noun) and may be *described* or/and *evaluated* (adverbs and adjectives). Such theoretical background gives rise to the following structure of a literary text (narrative) – ***SAO structure*** – *Subject (proper noun, pronoun)* → *Action (verb)* → *Object (common noun)* → *Description/Evaluation (adjective, adverb).* Functioning in a literary text, individual words and expressions, often repeated and correlated, reveal significant information and express formal structure the author uses to make the meaning of a literary text (narrative) well-disposed [3].

Nowadays quantitative data analysis available in computer processing make it possible to evaluate objectively meaningful components of textual structure, which reveals the meaning of a literary text in connection to the parts of speech [4]. The aim of the article is two-fold: *first,* to establish Pearson

---

COLINS-2023: 7th International Conference on Computational Linguistics and Intelligent Systems, April 20–21, 2023, Kharkiv, Ukraine
EMAIL: melnychuk_oksanadm@ukr.net (O. Melnychuk); ivan.a.bekhta@lpnu.ua (I. Bekhta); mariatkachivska@gmail.com (M. Tkachivska)
ORCID: 0000-003-4619-363x (O. Melnychuk); 0000-0002-9848-1505 (I. Bekhta); 0000-0001-9989-9156 (M. Tkachivska)

CEUR Workshop Proceedings (CEUR-WS.org)

correlation coefficient between frequent words in corpora (elements of SAO structure) and correlated words (meaningful components of literary text) using Voyant Tool web browser; *second*, to define weighty correlated parts of speech categories based on their quantity.

The author does not declaratively express the meaning in a literary text; it is hidden in textual fabric – in all the words that appear in mutual connection [5]. Using digital textual analysis tools often does not give us concrete or direct information about texts as complete meaningful units but about the words that need to be calculated to expose general structure. Thus, SAO structure hides important and interesting complexities, however, which provide insights on several topics of central interest to both literary text (narrative) analysis and applied linguistics.

2. Related works

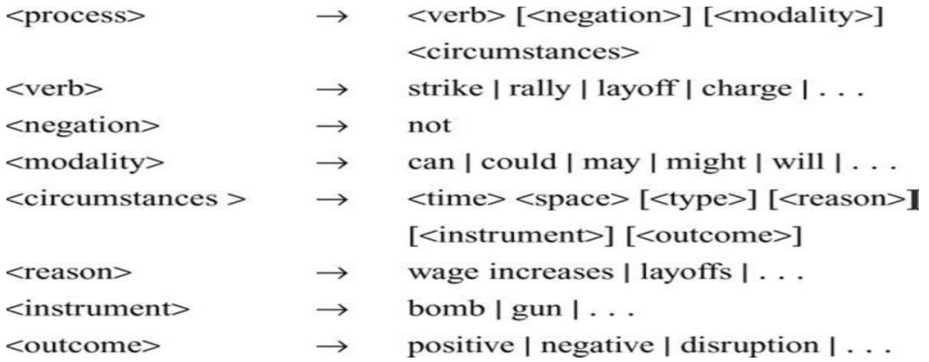
Related works forming theoretical background explain the role of meaning of a literary text (narrative) under SAO structure that comprises parts of speech. The section grounds the need to use Pearson correlation coefficient as a statistical measurement, which determines correlation of words as parts of speech (semantic entities) in literary text (narrative).

2.1. Meaning of a literary text: parts of speech correlation and the SAO structure

The meaning of a literary text (narrative) is a multidimensional and complex phenomenon. It includes many qualitative and quantitative aspects contributing to textual meaning interpretation [6]. Literary (narrative) texts are those where the distinctive traits of the narrative genre are quantitatively predominant. The properties of a literary text (narrative) include macro (semantic) structures, which map onto surface (syntactic) structures through parts of speech: nouns, verbs, adverbs and adjectives [7]. General meaningful sentence structure depicts a subject (actor) performing an activity that affects another entity (object) and uses this construction to depict actions (events) [8]. Y. Wang emphasizes the significance of triple structure: there is a subject (the protagonist or main character), an action (what the subject does), and an object (what the action is directed towards) [9]. Our way of approaching literary text (narrative), starts from the study of frequent parts of speech, completed with statistic correlations – Pearson correlation coefficient – that are united under the following triple SAO structure in literary text (narrative): *Subject ↔Action↔Object.* Emphasizing parts of speech , we will get *Nouns – proper names and pronouns (Subject) ↔ Verbs (Action) ↔ Nouns – common nouns (Object*).

This scheme is the basic analytical unit and assumes that tying formal components cohesively together follows the language specific practices involving part of speech correlation through Pearson correlation coefficient. This basic SAO structure may be extended to comprise adjectives and adverbs (evaluations and descriptions): *Nouns – proper names and pronouns* (*Subject)* ↔ Verbs (*Action*) ↔ *nouns* (*Object*) ↔ *Adjectives /adverbs (Evaluation /Description).*

In literary text (narrative), the parts of speech connection is due to SAO structure providing a sort of literary text (narrative) meaning, which formally may be rewritten as in Figure 1.



<process> → <verb> [<negation>] [<modality>] <circumstances>
<verb> → strike | rally | layoff | charge | . . .
<negation> → not
<modality> → can | could | may | might | will | . . .
<circumstances > → <time> <space> [<type>] [<reason>] [<instrument>] [<outcome>]
<reason> → wage increases | layoffs | . . .
<instrument> → bomb | gun | . . .
<outcome> → positive | negative | disruption | . . .

Figure 1: The meaning of a literary text under SAO structure

The action in SAO structure includes such semantic parameters as verb, process, negation, modality, circumstances (time, space), reason, instrument, and outcome. Thus, the subject provides as action directed on an object. The subject has a certain reason to do something through verb (process) in time and space with the help of an instrument under some circumstances and to get or not to get an outcome. The action also may have a modality and negation. The SAO may be expanded to include both description/evaluation, involving adjectives/adverbs. Both evaluation and description would be alternative elements of a story and could be attached to any object, in particular, events, actors, or physical objects. These parameters become concrete words as parts of speech to describe fiction world depending author's choice and ideas, historical period and desired effect.

2.2. Parts of speech significance in literary text

The central role of verbs is acknowledged by the fact that literary texts (narratives) are a particular kind of action discourse, that is, discourse, which is interpreted as a sequence of actions denoted by verbs and their properties [10]. M. Toolan also argued the significance of verb in literary text: "what is said will not be the core of a story; that, rather, what is done will be. The "what is done" then becomes (or may become) the core narrative text of actions while the "what is said" becomes evaluative commentary on those actions" [11]. For Sh. Rimmon-Kenan, the something that happens, [is] something that can be summarized by a verb or a name of an action. W. Labov stresses that narrative is one of the methods of running again through previous experience by matching a verbal arrangement of clauses to the sequence of events. For M.A.K. Halliday, processes denoted by verb is grouped into three main classes: (1) doing (or material), further divided into happening (being created); creating, changing doing (to), acting; (2) sensing (or mental), further divided into seeing, feeling, thinking (3) being (or relational), further divided into symbolizing, having identity, having attribute. S. Chatman also figures out events as actions and happenings, where actions are nonverbal physical acts, speeches, feelings, perceptions, and sensations of characters [12].

The verb having a "radiative power" is the locus of much of the semantic and grammatical information in the clause [13]. The verb is like a node or a link, and other words (parts of speech) are supposed to be connected to it. Being a necessary element to build a meaningful statement, according to L. Tesnière, a verb is the node of a sentence or of a group of words. From a semantic point of view, a verb expresses an action made or undergone, in other words, a change of state from A to B. "It acts as bridge between the subject (the agent → character) and the object (complement)" [14, p. 168]. Thus, the action of literary text denoted by a verb is constructive center of SAO structure – verb's valence characteristics determine which parts of speech will accompany it, what quantitative correlation the parts of speech will have to it and how they will be characterized semantically.

The significance of noun (proper nouns, common nouns) or pronouns is expressed by Subject (proper nouns) who performs an action and Object (common nouns). P. Geach and A. Gupta claim that the meanings of nouns involve "criteria of identity" [15, p. 474]. A. Wierzbicka proposes that the *primes thing(s)* and *people,* which may by subjects, provide a grammatical prototype for nouns [16]. Linguists distinguish three kinds of singular referring expressions: personal pronouns, definite descriptions and proper names, which are exclusive as fixed points in a dynamic fictional world. It is like a label of an information file we keep about a character. They are the condition for making knowledge and communication possible beyond the private ground [17].

Proper names in a literary text (narrative) play the role of markers of time and space. They reflect the fiction world of a definite social group in a certain era. Proper names in a literary text (narrative) concretize and unite all actions and characters into one single thematic system. Without them, the reader loses a sense of certainty in time and space of textual fabric. The proper name fastens a separate piece of information with the content of the entire text [18]. Proper names contain several types of information, their value is formed as correlation with the object, and in other words, the value of a proper name is identical to established information about the object. Proper names also have the property of a particular reference, as well as a massive number of connotations. The attributes of proper names are implemented in their functions: communicative, appellative, expressive and deictic – "all concepts are nouns" – understanding the semantic content of a noun is understanding "the amount of [defining] notes or elements that there are in the semantic content or idea” [19]. The definition entails that a proper noun indicates an entity without regarding the entities it belongs to.

So, there are at least two universally definable and prevalent parts of speech, which can be called noun (or “nominal”, when there is no contrast with adjectives or adverbs), and verb. The universality of adjectives is not established, although there are broad constructions restricted to descriptions/evaluations of states [16].

The relevance of adjectives is defined by *Evaluation* or *Description* in SAO structure. Adjectives “alter, clarify, or adjust the meaning contributions of nouns”: they can plainly align with nouns, forming complex constituents and linking with other elements to form a noun phrase [20]. At a general level, adjectives gain this capability in virtue of two main characteristics, one of which is semantic and the other is structural. On the semantic side, they suggest properties. On the SAO structure side, they are able to function as *Evaluation* and *Description*, and may tie up with nouns. The result of this combination is a new property, thereby providing a “finer shade of meaning of a literary text” (narrative) than is not possible using the noun alone [10, 11].

The SAO structure involves relations of concepts and ideas expressed with words as parts of speech distributed in a literary text. So, the meaning of a literary text reveals how often a definite word appears in a text to denote certain concept or idea being a kind of formal correlation (for example, a Pearson correlation coefficient) that exhibits explicit ties of different lexical elements contributing to the meaning of a literary text (narrative).

Absolute word frequencies, relative word frequencies, and correlation are formal but significant values used in digital humanities. Following J. C. Tello and J. Pääkkönen, we argue that textual meaning can be identified by information on word (parts of speech) frequency and statistical correlation based on SAO structure [2, 21]. Therefore, scrutiny of textual features is generally considered a prerequisite for literary interpretation. While the computer may lack the ability to detect “qualitative” semantic differences, its promise of a seemingly boundless quantitative analytical scope turns it into a potentially powerful analytic tool.

3. Method

Textual analysis is the most critical method in literary studies [22, 23]. Because it deals with a literary text (narrative), it places greater emphasis on its structure (Subject ↔ Action ↔ Object) expressed as words (parts of speech) [24, 25]. Researchers aim to understand and explain how these SAO structure elements, as parts of speech and their correlation with other parts of speech, contribute to the textual meaning [10; 11; 12]. Under the present research, correlations of the most frequent words (Pearson correlation coefficients) and the parts of speech of these correlations subsidize the meaning of a literary text. The purpose of the statistical textual analysis is to single out frequent words and define related parts of speech involving computer processing of Pearson’s correlation coefficient to contribute revealing the meaning of a literary text (narrative) based on SAO structure. Two methods were employed to collect data to the present study. The first is quantitative text analysis to define word frequency using web-browser Voyant Tool. The second is the study of Pearson correlation coefficient generated by Voyant Tool in terms of parts of speech related to SAO structure.

3.1. Procedure

The corpora of the present study cover the novels “The children’s book” [26] and “Possession: a romance” [27] written by A. Byatt, a British novelist, poet and Booker Prize winner. Procedure for conducting a textual analysis includes:

- determining the type of textual analysis: once the sample has been selected, the type of analysis is determined as a calculation of Pearson’s correlation coefficient of words in the textual corpus to detect parts of speech (verb, noun, adjective, adverb) significance in terms of concrete values;
- reducing the text to words. Two novels, “The children’s book” and “Possession: a romance”, were converted into txt files as two digital corpora and uploaded into Voyant – a tool for digital text processing;

- extracting the most frequent 10 words (Terms 1) in corpora “The children’s book” and “Possession: a romance”. We used Voyant Tool “trends” which generate frequent words as visual charts showing 10 textual segments and indexes of relative frequency for word distribution analysis;
- defining the parts of speech categories of extracted words as “proper noun”, “common noun”, “verb”, and “adjective/adverb” under SAO structure in each corpus. Category “proper noun” includes not only names of the characters but also “a doer of an action”: e.g. *men, helper, grosser, boy, miner* etc. The textual contexts were checked to define parts of speech categories correctly in case the meaning of words was ambiguous;
- exploring the relationship between frequent words (the parts of speech categories under SAO structure) and other words (Terms 2) in a literary text using Pearson’s correlation coefficient (Terms 1 ↔ Terms 2) and applying Voyant Tool “correlation”. We limited each Term 1 to have only 15 correlated words – Terms 2. The correlation of frequent words establishes the values of correlations and their significance among correlated parts of speech. Values approaching 1 are noteworthy and mean that word frequencies vary in synchrony (they rise and drop together); values approaching -1 mean that term frequencies vary inversely (one rises as the other drops). Values approaching 0 indicate little or no meaningful correlation;
- examining the measure of the significance of the correlation value. A significance of 0.5 or less indicates a strong correlation, allowing us to reject a null hypothesis that values are randomly distributed. The validity of this measure depends on the assumption about the normal distribution of the data;
- defining the parts of speech categories (“proper noun”, “common noun”, “verb”, “adjective/adverb”) of correlated words, their quantity and prevalence while correlating with the most frequent words in each corpus.

3.2. Pearson correlation coefficient

Pearson correlation coefficient is a measure of linear correlation between two sets of data. It is the ratio between the covariance of two variables and the product of their standard deviations [3; 6]. It is substantially a normalized measurement of the covariance. The coefficient always has a value between −1 and 1. Textual analysis measures how closely word frequencies correlate. The correlation of the most frequent words and other words in a corpus manifests the meaning of a literary text in terms of parts of speech dependencies.

4. Results and discussion

This section waves around a computer-assisted case study of the words as parts of speech categories representing Pearson correlation coefficients in researched corpora. The results are illustrated as visualization of word frequency in 10 textual segments (Figures 2, 3, 4, 7, 8, 9), tables containing the values of correlation and values of significance (Table 1, Table 2), and the charts exhibiting the quantity of correlated words due to parts of speech categories (Figures 5, 6, 10, 11).

4.1. Corpus “Possession: a romance”: Pearson correlation coefficient and quantitative data analysis under SAO structure

The analysis of Pearson correlation coefficient starts with the analysis of word frequency in each corpus. The most frequent words including proper nouns, a common noun, verbs and adjectives in corpus “Possession: a romance” are *said* (941); *like* (522); *know* (504); *maud* (398); *ash* (381), *think* (339); *thought* (297); *little* (392), *roland* (377), *time* (368). Further, we group frequent words according to parts of speech categories: verbs, proper nouns, common nouns and adjectives/adverbs. The diagrams show the most frequent proper nouns (Fig. 2), verbs (Fig. 3), common nouns and adjectives (Fig. 4) in 10 textual segments. The diagrams including relative frequencies demonstrate that verbs are distributed evenly in textual fragments (except verb *know*). It proves verbs’ importance in providing a SAO structure – the verbs form a kind of an action scheme, a balanced saturation of the literary text (narrative). On the contrary, the quantity of proper names is sharply different in each textual segment. The object *time* is presented in each textual fragment having approximately the same quantity, and the evaluation/description *little* is the highest in the second textual fragment.

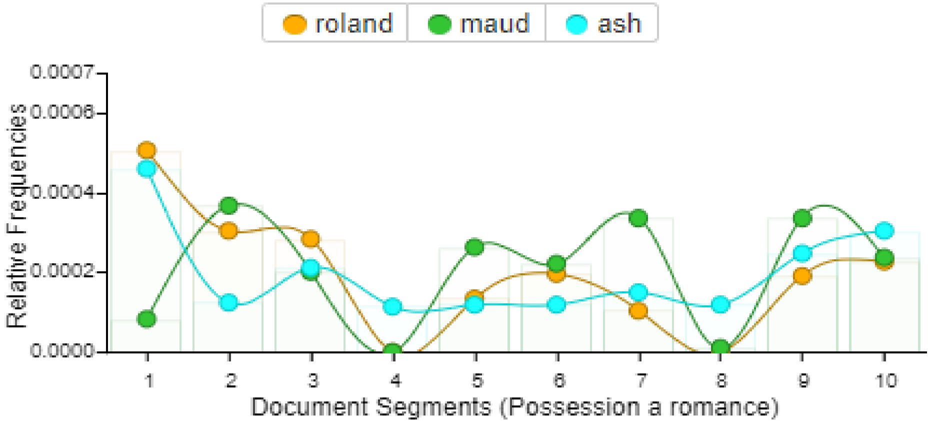


Figure 2: The most frequent proper nouns in corpus “Possession: a romance”

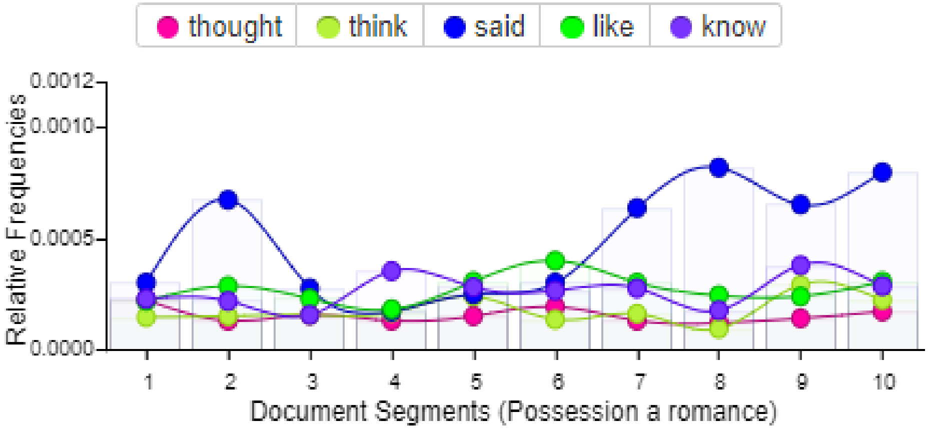


Figure 3: The most frequent verbs in corpus “Possession: a romance”

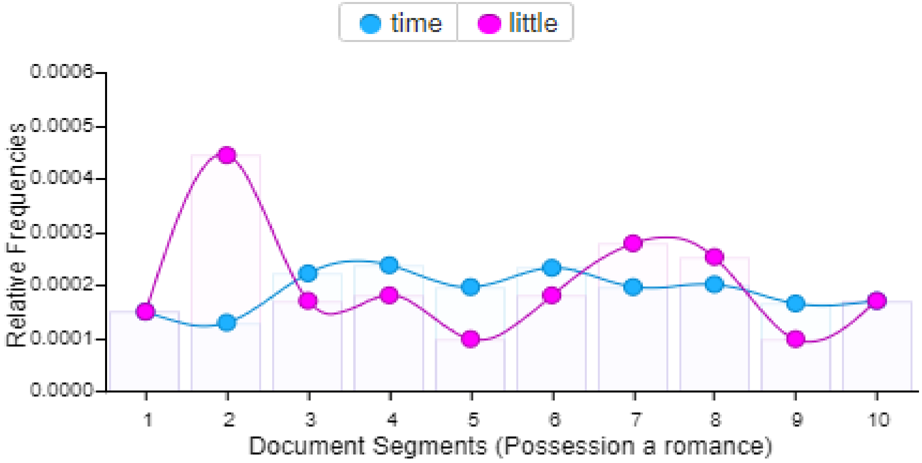


Figure 4: The most frequent common nouns and adjectives in corpus “Possession: a romance”

These frequent words compose extended SAO structure in corpus “Possession: a romance” as following:

Subject (*maud, ash, roland*) ↔ Action (*said, like, know, think thought*) ↔ Object (time) ↔ Descripion/Evaluation (*little*)

Each of these components correlate with a number of words as parts of speech. Using Pearson correlation coefficient in corpus “Possession: a romance” we defined the values of correlation and significance for selected frequent word in a corpus (15 correlations for each word). Pearson correlation coefficient is generated by Voyant Tool. The results are shown as a table containing the most frequent words under SAO structure (Table 1).

Table 1

Pearson correlation coefficient in corpus “Possession: a romance”

Frequent word/ number	Pearson correlation coefficient		
Element of SAO Structure (Terms 1)	Related words (Terms 2)	Correlation	Significance
Subject (proper noun)			
Maud			
1	isn’t	0,9396923	0,0000537
2	built	0,86928916	0,0010874
3	I’m	0,84601897	0,0020335
4	curly	0,8124352	0,0042887
5	car	0,79982734	0,005473642
6	clothed	0,79216015	0,0062987
7	arm	0,79015803	0,006528032
8	bit	0,77595407	0,008327718
9	expect	0,77303183	0,008737342
10	george’s	0,77273464	0,008779782
11	covered	0,7661926	0,009751313
12	dressing	0,76357704	0,010160117
13	end	0,7615354	0,010487494
14	buried	0,7604152	0,010670231
15	listen	0,75541043	0,0115140
Ash			
1	applications	0,9178341	0,0001804
2	arranged	0,91193366	0,0002363
3	actual	0,85975754	0,0014238
4	advisory	0,85975754	0,0014238
5	amazing	0,85975754	0,0014238
6	affair	0,8564658	0,001555841
7	acquired	0,8234388	0,003415447
8	argued	0,8126585	0,004269597
9	1853	0,8126585	0,004269597
10	1856	0,8126585	0,004269597
11	aggression	0,8126585	0,004269597
12	apricot	0,8126585	0,004269597
13	ariachene’s	0,8126585	0,004269597
14	1986	0,7831885	0,0073723
15	aged	0,7831885	0,0073723
Roland			
1	henry	0,9213048	0,0001524
2	blond	0,90000472	0,000386449

3	paper	0,89243954	0,0005133
4	affair	0,8902855	0,0005542
5	confident	0,88657725	0,0006300
6	edition	0,8846337	0,000727
7	carpets	0,8834149	0,0007005
8	randolph	0,8705399	0,0010481
9	london	0,8692129	0,0010899
10	floor	0,8663786	0,0011832
11	duly	0,8636723	0,001277724
12	elderly	0,8636723	0,001277724
13	pine	0,86150235	0,0013573
14	excited	0,84899026	0,001888194
15	carlyle	0,84260106	0,0022105
Action (verb)			
Said			
1	knows	0,8786299	0,0008179
2	ends	0,8688386	0,00110193
3	prick	0,8229225	0,0034532
4	country	0,8131771	0,00422533
5	protected	0,8090084	0,0045905
6	kissed	0,77819663	0,00802273
7	heard	0,7767839	0,008213923
8	grew	0,7755604	0,00838209
9	grandmother	0,7679386	0,00948496
10	cradle	0,7610763	0,01056213
11	lantern	0,7610763	0,01056213
12	laughter	0,7610763	0,010562125
13	lawns	0,7610763	0,010562125
14	child	0,76013106	0,010716934
15	miner	0,75726753	0,011195682
Like			
1	inaccessible	0,90858716	0,0002732
2	animal	0,8977402	0,0004221
3	hole	0,888688254	0,0006235
4	breathing	0,8785134	0,0008209
5	jet	0,8780896	0,0008320
6	boy	0,8584599	0,0014748
7	approach	0,8497918	0,001850305
8	hissed	0,8303317	0,002938348
9	exhausted	0,82830024	0,0030736
10	intent	0,828300024	0,0030736
11	brooch	0,7997295	0,0054836
12	broke	0,7963401	0,0058387
13	experiments	0,7883501	0,006740079
14	dining	0,7849032	0,007157828
15	beat	0,782487	0,007461395
Know			
1	incoherent	0,8218782	0,003530742
2	artful	0,7804074	0,0077298
3	articulate	0,7804074	0,0077298
4	bless	0,7804074	0,0077298

5	citation	0,7804074	0,0077298
6	coals	0,7804074	0,0077298
7	consideration	0,7804074	0,0077298
8	contradictory	0,7804074	0,0077298
9	decorously	0,7804074	0,0077298
10	deference	0,7804074	0,0077298
11	defined	0,7804074	0,0077298
12	englishmen	0,7804074	0,0077298
13	grosser	0,7804074	0,0077298
14	haunted	0,7804074	0,0077298
15	home’s	0,7804074	0,0077298
Think			
1	smoked	0,89646953	0,0004428
2	spark	0,89646953	0,0004428
3	incoherent	0,87647665	0,0008751
4	problems	0,87647665	0,0008751
5	hotel	0,87036884	0,0010534
6	superstitious	0,86586165	0,0012009
7	appropriate	0,85971165	0,0014256
8	calmer	0,8532674	0,0016923
9	cherubs	0,8532674	0,0016923
10	compulsive	0,8532674	0,0016923
11	helper	0,8532674	0,0016923
12	sauce	0,8532674	0,0016923
13	straining	0,8532674	0,0016923
14	sealed	0,8340457	0,0027021
15	feminism	0,8235279	0,003408952
Thought			
1	continuing	0,93145853	0,0000888
2	arcane	0,91275305	0,0002279
3	handed	0,8975664	0,0004249
4	period	0,8777275	0,0008415
5	thought	0,8747158	0,000241
6	breadwinner	0,86496395	0,0012319
7	carries	0,86496395	0,0012319
8	coincided	0,86496395	0,0012319
9	ferny	0,86496395	0,0012319
10	regularly	0,86496395	0,0012319
11	script	0,86496395	0,0012319
12	thinkers	0,86496395	0,0012319
13	pleasant	0,85727495	0,0015226
14	jeans	0,85556453	0,0015934
15	melusina	0,84553415	0,002058021
Object (common noun)			
Time			
1	reason	0,95924824	0,0000114
2	shape	0,90168244	0,0003625
3	life	0,8627234	0,0013120
4	rings	0,8464182	0,0020135
5	imagine	0,83277196	0,0027815
6	supposing	0,8143953	0,004122658

7	cast	0,79608	0,005866638
8	respect	0,79215896	0,0062989
9	ruddy	0,78366476	0,0073123
10	harder	0,7789868	0,007917174
11	men	0,7789868	0,0079171
12	believes	0,7761436	0,008301629
13	human	0,7759228	0,008332037
14	below	0,77022594	0,009143847
15	delighted	0,76385945	0.010798283
Evaluation/Description (adj/adv)			
Little			
1	funds	0,95213115	0,0000216
2	cap	0,9374048	0,0000622
3	doors	0,9033379	0,0003339403
4	ensure	0,9011337	0,000370418
5	frightful	0,9011337	0,000370418
6	hairy	0,9011337	0,000370418
7	instinct	0,9011337	0,000370418
8	hedgehog	0,8977526	0,0004219
9	access	0,8845944	0,0006736
10	east	0,8725159	0,0009880
11	knocked	0,8725159	0,0009880
12	advice	0,8707238	0,0010424
13	craft	0,8678781	0,0011331
14	hen	0,8676193	0,00118603
15	craftsman	0,8662977	0,0014402

The table demonstrates that the correlation of all presented words is significant – it is no less than 0,8 having relevant value of significance – less than 0,5. The words with high correlation values are *knows*, *ends*, *prick*, *country*, *protected* (for *said*); *inaccessible* (for *like*); *incoherent* (for *know*); continuing (for *thought*); *isn't* (for *maud*); *applications, arranged* (for *ash*), henry blond (for *Ronald*).

Frequent words correlate with all of the researched parts of speech categories but each of the word "draws" different quantity of proper nouns, common nouns, verbs, adjectives and adverbs (Fig. 5). The highest figures of parts of speech categories are concentrated in *Action* (verb) element of SAO structure. Action "attracts" mostly nouns and verbs. Common nouns, which present *Object* are not so numerous.

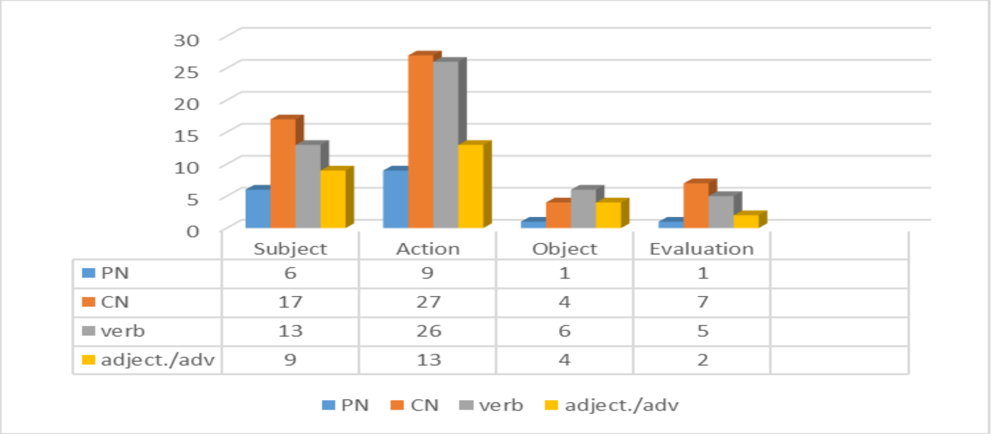


Figure 5: Parts of speech correlation under SAO structure in corpus "Possession: a romance"

To make the results more accurate we depicted correlated parts of speech categories as percentage (%). Figure 6 demonstrates that subject (frequent proper names) mostly correlates with proper nouns; action (frequent verbs) – with common nouns and verbs; object (frequent common nouns) – with verb; and evaluation (frequent adjectives) – with common nouns. Proper nouns and adjectives do not have high percentage while correlating with *Subject*, *Object*, and *Evaluation/Description*.

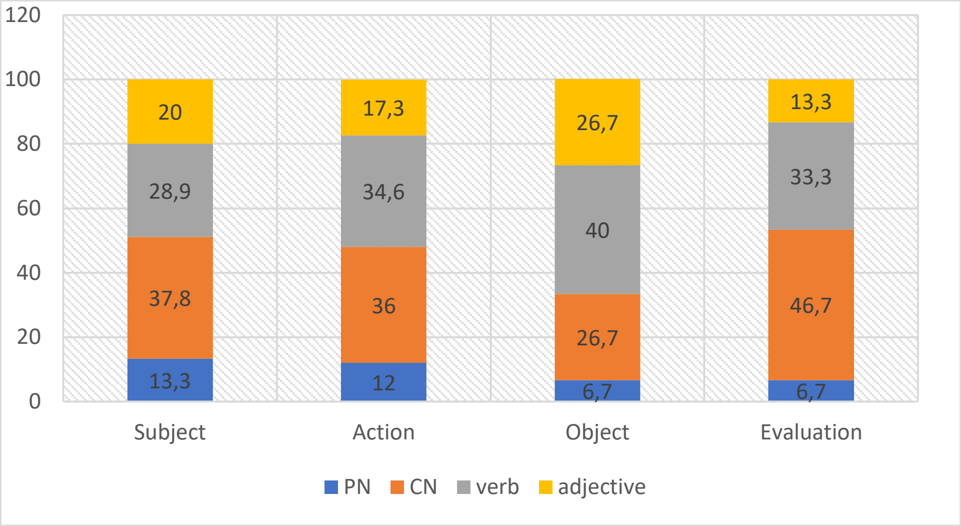


Figure 6: Parts of speech correlation under SAO structure (%) in corpus “Possession: a romance”

4.1. Corpus “The children’s book”: Pearson correlation coefficient and quantitative data analysis under SAO structure.

The most frequent words covering proper nouns, common nouns, verbs and adjectives in corpus “The children’s book” are: *said* (2023); *like* (821); *dorothy* (543); *tom* (528); *thought* (520); *know* (489); *philip* (479); *think* (424); *little* (397); *things* (354). Grouped frequent words as parts of speech categories are presented as proper nouns (Fig. 7), verbs (Fig. 8), common nouns and adjectives (Fig. 9).

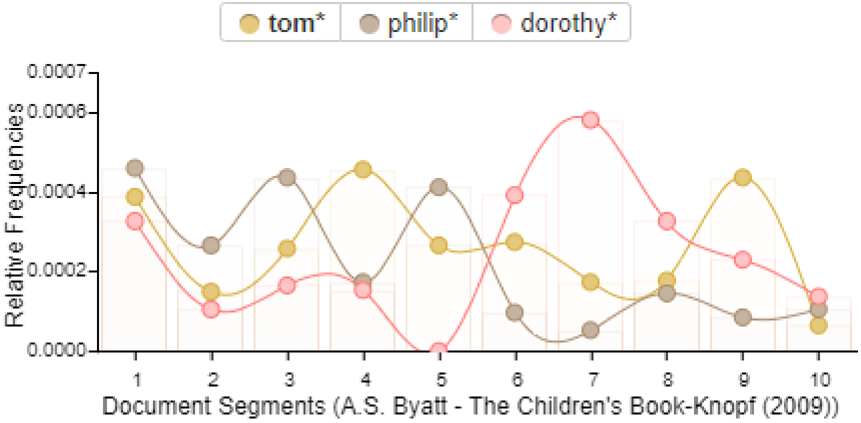


Figure 7: The most frequent proper nouns in corpus “The children’s book”

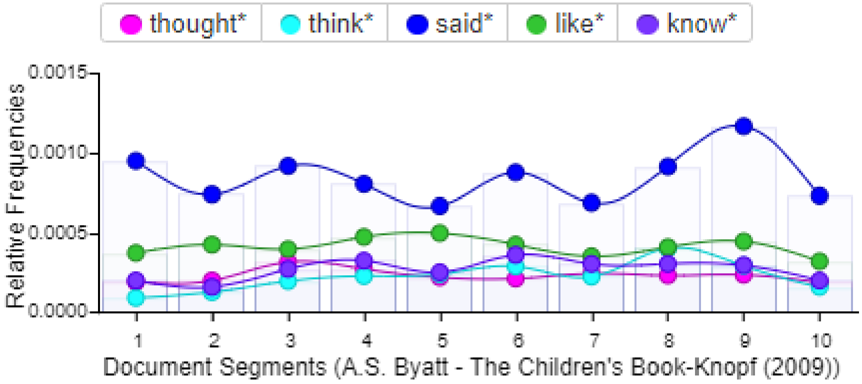


Figure 8: The most frequent verbs in corpus “The children’s book”

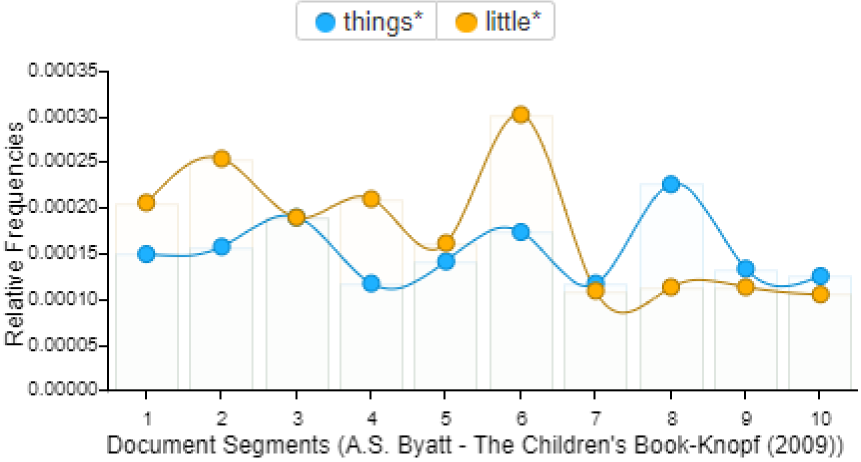


Figure 9: The most frequent adjectives and common nouns in corpus “The children’s book”

The diagrams show how the frequent words are distributed in textual segments: the verbs are almost the same in each textual fragment (except the word *know*, as it is in previous corpus) that help the reader predict the unfolding of the scene. Proper nouns, common nouns and adjectives have sharp fluctuations that signifies their instability or variability.

The frequent words of the corpus compose extended SAO structure as following:

Subject (dorothy, tom, philip) ↔ Action (said, like, know, thought, think) ↔ Object (things) ↔ Description/Evaluation (little)

Further, the frequent words are taken to establish correlations with other words in the corpus “Children’s book”. The results containing values of correlation and its significance are summarized in Table 2.

Table 2

Pearson correlation coefficient in corpus “The children’s book”

Frequent word/ Number (Terms 1)	Pearson correlation coefficient		
Element of SAO Structure (Terms 1)	Related words (Terms 2)	Correlation	Significance
Subject (proper nouns)			
Dorothy			
1	confined	0,8472941	0,0019702
2	cousins	0,8338751	0,0027127
3	beer	0,821122	0,0035875
4	brother	0,7955094	0,0056656

5	bohemian	0,7955094	0,0059281
6	companion	0,791239	0,0059281
7	cherubs	0,783321	0,006403537
8	stern	0,78332144	0,006742209
9	concentrated	0,7643854	0,007355542
10	dances	0,7643854	0,008822043
11	classes	0,7643854	0,0090222611
12	collars	0,7643854	0,10032507
13	come	0,75365496	0,10032507
14	desires	0,7509045	0,10032507
15	capable	0,74825895	0,011820855
Tom			
1	gallery	0,9132047	0,0002233
2	keeper	0,8757968	0,0008938
3	let’s	0,8757968	0,0008938
4	doubtful	0,8654017	0,0012167
5	hello	0,8654017	0,0012167
6	hair	0,8654017	0,0012167
7	necklace	0,8654017	0,0012167
8	didn’t	0,8578555	0,0014990
9	persuaded	0,8575924	0,0015097
10	allowed	0,85391515	0,0023114
11	second	0,84074134	0,0027338
12	october	0,83353394	0,0032108
13	stroke	0,82630396	0,0032108
14	edge	0,82630396	0,0033056
15	printed	0,82496035	0,0039203
Philip			
1	gripping	0,92839617	0,0001054
2	orange	0,92839617	0,0001054
3	downstairs	0,9215292	0,0001507
4	chimneys	0,91334325	0,0002219
5	fitted	0,90867895	0,0002721
6	pattern	0,9000763	0,0003860
7	hedges	0,895135	0,0004653
8	gloves	0,8854391	0,0007575
9	bicycle	0,8810241	0,0009338
10	mines	0,87437385	0,0012078
11	benignly	0,86565936	0,0012338
12	manor	0,8649106	0,0013609
13	wanted	0,8614043	0,0015066
14	lesson	0,85766715	0,0018633
15	morris	0,8495146	0,0021563
Action (verb)			
Said			
1	nurseries	0,9700613	0,0000033
2	hard	0,9401429	0,0000522
3	aspects	0,9074871	0,0002862
4	desk	0,8813762	0,0007489
5	rejected	0,8665187	0,0011785
6	fairytales	0,86146086	0,0013588

7	recovering	0,83969057	0,0023698
8	happening	0,8278009	0,0031075
9	conducted	0,826825	0,0031746
10	confident	0,826825	0,0031746
11	lustre	0,8256674	0,003255531
12	fits	0,8234412	0,003415275
13	I'd	0,82142603	0,0035646
14	proposing	0,81395954	0,0041591
15	marriage	0,808829	0,004606761
Like			
1	flowers	0,91845423	0,0001751
2	glow	0,8917597	0,0005259
3	accompanied	0,8550317	0,0016160
4	devil	0,84645975	0,0020115
5	acceptable	0,7987976	0,0055797
6	cheered	0,7987976	0,0055797
7	fingers	0,787381	0,006855731
8	attending	0,77948433	0,007851215
9	gets	0,7768593	0,008203639
10	body	0,7750972	0,008446392
11	glimpsed	0,77155185	0,0089501
12	crest	0,77003944	0,009171335
13	bowls	0,76985914	0,009197959
14	fragments	0,7608098	0,010605605
15	clutched	0,75349385	0,011849292
Thought			
1	rocks	0,95874316	0,0000120
2	read	0,9509246	0,0000239
3	slow	0,9135194	0,0002202
4	coming	0,85929275	0,0014419
5	colony	0,85650945	0,0015540
6	beginning	0,8555952	0,0015921
7	camping	0,84898806	0,0018882
8	purposes	0,84898806	0,0018882
9	staggered	0,84357536	0,0021590
10	sodden	0,83611345	0,0025768
11	chamberlain	0,8230791	0,003441779
12	sun	0,81792986	0,00383487
13	bleached	0,8065464	0,0048165
14	detached	0,80072427	0,005382436
15	stretching	0,80055135	0,005399933
Think			
1	argued	0,92063814	0,0001575
2	went	0,85731614	0,001520935
3	strange	0,8550248	0,0016163
4	talk	0,854361	0,0016447
5	heart	0,8517629	0,0017594
6	barriers	0,84632635	0,0020181
7	soul	0,831341	0,002872755
8	situation	0,8231588	0,0034359
9	boat	0,80421656	0,005037626

10	sword	0,8041112	0,005047795
11	mermaid	0,80218565	0,005236166
12	know	0,7969573	0,005236166
13	companion	0,7965497	0,0057728
14	supposed	0,7887631	0,0058163
15	good	0,78757775	0,0066912
Know			
1	imogen’s	0,7968629	0,005782919
2	fragments	0,7957306	0,005904276
3	comic	0,7938005	0,006115285
4	curled	0,77758361	0,0083343985
5	drew	0,75548583	0,01150103
6	argued	0,7450523	0,013407478
7	good	0,7286243	0,01651701
8	feelings	0,72459614	0,016840832
9	kind	0,71900505	0,01776766
10	absence	0,71440244	0,019111926
11	absent	0,70867527	0,0202701
12	interest	0,69232273	0,021777874
13	imogen	0,6910697	0,026505237
14	admitted	0,6826844	0,026894113
15	formal	0,6802213	0,029597443
Object (common noun)			
Things			
1	fetch	0,92374337	0,0001348
2	comforting	0,92351073	0,0001364
3	frequently	0,9134588	0,0002208
4	choice	0,88396806	0,0006878
5	firing	0,87208664	0,0010008
6	imagination	0,8606471	0,0013896
7	know	0,8599656	0,0014158
8	mermaid	0,8598922	0,0014158
9	afternoons	0,8598922	0,0014158
10	cooling	0,8598922	0,0014158
11	dish	0,8598922	0,0014158
12	irritably	0,8598922	0,0014158
13	coast	0,8598922	0,0014158
14	should not	0,8598922	0,0014158
15	frosty	0,8567037	0,0015460
Evaluation/Description (adj/adv)			
Little			
1	dozed	0,90823495	0,0002773
2	hadn’t	0,8885088	0,0005896
3	eye	0,88297594	0,0007108
4	iacy	0,877131	0,0008674
5	clear	0,8602178	0,0013975
6	away	0,8594129	0,0014060
7	dropped	0,8577897	0,001437293
8	cauldron	0,8459127	0,00151744
9	evidence	0,8450388	0,0020389

10	dresses	0,8406161	0,0020832
11	hounds	0,8406161	0,00231835
12	impression	0,8378658	0,00231835
13	ill	0,8365345	0,0024738
14	circular	0,8301295	0,0025517
15	amiably	0,8300202	0,002951618

The table exposes that the correlation of all presented words is significant – it is no less than 0,8 having relevant value of significance – less than 0,5. The words with high correlation values are *nurseries*, *hard*, *aspects* (for *said*); *flowers* (for *like*); *rocks*, *read*, *slow* (for *thought*).

Frequent words correlate with all of the researched parts of speech categories but each of the word "attracts" different quantity of proper nouns, common nouns, verbs, adjectives and adverbs (Fig. 9). The highest figures of parts of speech categories are concentrated in *Action* (verb) element of SAO structure. Action "draws" mostly nouns and verbs. Common nouns, which present *Evaluation/Description* are not so numerous.

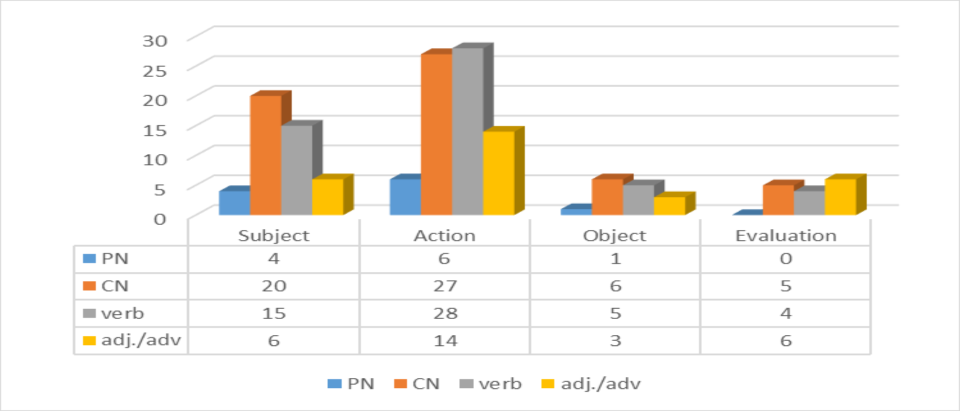


Figure 10: Parts of speech correlation under SAO structure in corpus "The children's book"

To make the results more exact we illuminated correlated parts of speech categories as percentage (%). Figure 10 demonstrates that subject (frequent proper names) mostly correlates with common nouns; action (frequent verbs) – with common nouns and verbs; object (frequent common nouns) – with verbs; and evaluation (frequent adjectives) – with adjectives.

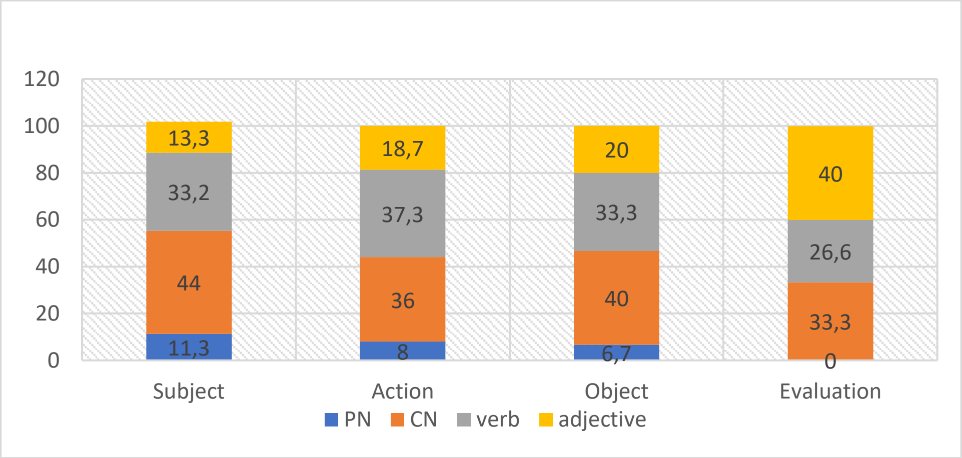


Figure 11: Parts of speech correlation under SAO structure (%) in corpus "Possession: a romance"

Proper nouns and adjectives do not have high percentage while correlating with *Subject*, *Object*, and *Evaluation/Description.*

We see approximately the same quantity of correlated parts of speech categories in both corpora under SAO structure. It means that the Pearson correlation coefficient does not characterize the author's style but contributes meaning exposing in textual structure.

5. Conclusion

Taken together, these results suggest that the Pearson correlation coefficient is the significant quantitative index in the study of the meaning of a literary text through the statistical correlation of words, which forms the basis for the semantic analysis of a literary text under SAO structure. The resulting picture is one that raises a number of noteworthy questions about the centrality of verb and nouns meaning in relation to Action and Subject in literary text (narrative) under the scope of statistical textual analysis.

The most frequent words in researched corpora (the novels "Possession: a romance", "The children's book") are parts of speech categories which reveal the meaning and correspond to interrelation within the structure of literary text (narrative) – *Proper nouns (Subject) ↔ Verbs (Actions) ↔ Object (Common nouns) ↔ Evaluation/Description (Adjectives/Adverbs)*. The most frequent words in the corpora are the verbs *said*, *like*, and *thought*. These results clearly show that the most frequent words in corpora suggest a high Pearson correlation coefficient (0,8-0,9) that is noteworthy (less than 0,5). This research proves the idea about centrality of the verbs embodied in *Action* and connected to *Object* (common nouns) in a literary text (narrative) in spite of the author's style.

By carefully examining the data, it was found that the most frequent words in each corpus correlate with words as definite parts of speech: mostly with nouns and verbs, to a lesser extent with adjectives. In perspective, the investigation of such correlations may be broadened as semantic analysis of the parts of speech categories. Calculation of the Pearson correlation coefficient of the words in a literary text might be addressed in future studies involving both quantitative aspects (e.g. Spearman correlation) and qualitative parameters of literary textual interpretation or cognitive modelling.

6. References

[1] R. Chartier, Genealogies of the study of material texts: the French trajectory, Textual Cultures, 14/1 (2021) 20–25.

[2] J. C. Tello, Grammatical, lexical, semantic, and textual annotation, in The Novel in the Spanish Silver Age, in: J. C. Tello (Ed.), A Digital Analysis of Genre Using Machine Learning, Bielefeld University Press, Bielefeld, 2021, pp. 179–190. doi:10.1515/9783839459256

[3] U. Varadarajan, B. Dutta, Models for narrative information: A Study, 2020. URL: https://arxiv.org/

[4] R. Nicewander, Thirteen ways to look at the correlation coefficient, The American Statistician, 42/1 (1988) 59–66.

[5] R. Odin, V. Hediger, Textual analysis and semio-pragmatics, in: R. Odin, V. Hediger (Eds.), Spaces of Communication: Elements of Semio-Pragmatics, Amsterdam University Press, Amsterdam, 2022, pp. 141–156. doi: 10.1515/9789048538669-002.

[6] M. W. Monroe, Using quantitative methods for measuring inter-textual relations in Cunei form, in: V. B. Juloux, A. R. Gansell, and A. di Ludovico (Eds.), CyberResearch on the Ancient Near East and Neighboring Regions: Case Studies on Archaeological Data, Objects, Texts, and Digital Archiving, Brill, Leiden, Boston, 2018, pp. 257–280. doi: 10.1163/9789004375086_010

[7] A. Goldstone, Teaching quantitative methods: what makes it hard (in literary studies), in: M. K. Gold, L. F. Klein (Eds.), Debates in the Digital Humanities, University of Minnesota Press, Minneapolis, 2019, pp. 209–223. doi:10.5749/j.ctvg251hk.22.

[8] A. Pawley, The depiction of sensing events in English and Kalam, in: H. Bromhead, Z. Ye (Eds.), Meaning, Life and Culture: In Conversation with Anna Wierzbicka, 1st ed., ANU Press, 2020, pp. 381–402. doi:10.2307/j.ctv1d5nm0d

[9] Y. Wang, Narrative Structure Analysis: A Story from "Hannah Gadsby: Nanette", Journal of Language Teaching and Research, 11/5 (2020) 682–687. doi: 10.17507/jltr.1105.03.

[10] Franzosi R. (Ed.), On quantitative narrative analysis. SAGE Publications, Inc., London, 2012. doi:10.4135/9781506335117

[11] Toolan M., Narrative and narrative structure, in: K. Allan (Ed.), The Routledge Handbook of Linguistics, Routledge, 2015, pp. 236–249.

[12] T. Ogata, T. Akimoto. Post-narratology through computational and cognitive approaches, IGI Global, Japan, 2019.

[13] C. Goddard, Prototypes, polysemy and constructional semantics: The lexicogrammar of the English verb climb, in: H. Bromhead, Z. Ye (Eds.), Meaning, Life and Culture: In Conversation with Anna Wierzbicka, 1st ed., ANU Press, 2020, pp. 13–32. doi:10.22459/MLC.2020.01.

[14] V. B. Juloux, A qualitative approach using digital analyses for the study of action in narrative texts: KTU 1.1–6 from the Scribe Ilimilku of Ugarit as a case study, in: V. B. Juloux, A. R. Gansell, A. di Ludovico (Eds.), CyberResearch on the Ancient Near East and Neighboring Regions: Case Studies on Archaeological Data, Objects, Texts, and Digital Archiving, BRILL, 2018, pp. 151–193. URL: https://www.jstor.org/stable/10.1163/j.ctv4v349g.14.

[15] Epstein B., Sortals and criteria of identity, Analysis, 72/3 (2012) 474–478.

[16] A. D. Andrews, On defining parts of speech with Generative Grammar and NSM, in: H. Bromhead, Z. Ye (Eds.), Meaning, Life and Culture: In Conversation with Anna Wierzbicka, 1st ed., ANU Press, 2020, pp. 333–354. doi: 10.2307/j.ctv1d5nm0d.24

[17] D. N. Djenar, M. C. Ewing, H. Manns, Referring to self and other, in: D. N. Djenar, M. C. Ewing, H. Manns (Eds.), Style and Intersubjectivity in Youth Interaction, 1st ed., De Gruyter, Berlin, 2018, pp. 23–63.

[18] W. G. Lycan, Metaphysics and the paronymy of names, American Philosophical Quarterly, 55/4 2018 405–419. doi: 10.2307/45128634.

[19] M. García-Carpintero, Semantics of fictional terms, Teorema: Revista Internacional de Filosofía, 38/2 (2019) 73–100.

[20] J.Giacon, Adjectives – Gayrrda, in: J. Giacon (Ed.), Wiidhaa: An Introduction to Gamilaraay, ANU Press Languages, 2020, pp. 119–126.

[21] J. Pääkkönen, Data do not speak for themselves: interpretation and model selection in unsupervised automated text analysis, in: A. Licastro, B. Miller (Eds.), Composition and Big Data, University of Pittsburgh Press, Pittsburgh, 2021, pp. 245–261.

[22] O. Melnychuk, N. Bondarchuk, I. Bekhta, O. Levchenko. Quantitative features of the words representing nonverbal behaviour in Ian McEwan’s fiction, Proceedings of the 6th International Conference on Computational Linguistics and Intelligent Systems (COLINS 2022). Volume I: Main Conference, Gliwice, May 12-13, 2022, Poland, pp. 461–470.

[23] O. Melnychuk, N. Bondarchuk, I. Bekhta, O. Levchenko, N. Yesypenko, N. Hrytsiv. The Quantitative Parameters in Computer-Assisted Approach: Author’s Lexical Choices in the Novels by Martin Amis, in: Proseedings of IEEE 17th International Conference on Computer Science and Information Technologies (CSIT), Lviv, 10-12 November, 2022, pp. 89–92.

[24] C. Luck, Rewriting Language: How Literary Texts Can Promote Inclusive Language Use, University College, London, 2020.

[25] L. C. Lawyer, The verb and verbal morphology, in: A Grammar of Patwin, University of Nebraska Press, 2021, pp. 228–331.

[26] A.S. Byatt, Possession: A Romance. Vintage, London, 2018.

[27] A.S. Byatt, The Chidren’s Book. Vintage, London, 2018.