

Language Corpora as Data Base and Verification Tool for Cognitive Linguistic Research

Svitlana Lyubymova

The Department of Foreign Languages,
National Polytechnic University, Odessa, Ukraine
E-mail: elurus2006@gmail.com

Abstract: The work presents a corpus analysis carried out within the framework of cognitive linguistics. The application of corpus analysis in the study of language representation of a sociocultural phenomenon was chosen in relation to the objectives to verify a hypothesis about the phenomenon's stereotypic and symbolic nature and to trace its evolution in culture. The paper gives a general classification of corpora, briefly describes the corpus approach and corpus-application techniques in cognitive linguistics research. Our study of sociocultural stereotypes as reflected in the corpora data is based on a quantitative and qualitative analysis of the lexical representation of stereotypes. The results are formalized in distribution patterns and frequency tables showing categorical features of these stereotypes and the alterations in its cognitive structure since its emergence. The acquired results and implemented hybrid approach to the study of cognitive characteristics of linguistically represented stereotypes might provide important support for researchers interested in studying the cognitive nature of linguistic phenomena.

Keywords: corpus analysis, cognitive linguistics, sociocultural stereotypes, collocation, context, distribution, frequency, assessment.

1. Introduction

With the advances of computer technologies, data mining for various linguistic disciplines has greatly enlarged. At the turn of the century, we witnessed a breakthrough in the use of computer corpora in linguistic research. The large data stores known as "corpora" exist in a computer-readable form of textual collections assembled for the purpose of linguistic investigation. Nowadays corpora are used for a wide range of studies: cognitive linguistics, semantic and pragmatic research, synchronic and diachronic studies, cross-linguistic research, discourse analysis, etc. Corpora provide researchers with convenient tools to conduct empirical studies of repeatable outcomes and ensure valid judgments about linguistic phenomena.

The popularity of corpora for empirical research has put forward the question about the position of corpus linguistics among language sciences. A decade or more ago, there was a lot of controversy about corpus linguistics status. Scientists argued, whether corpus linguistics is "just a newly emerging methodology for studying language" or "a new research enterprise, and in fact a new philosophical approach" [1]. Those who admit disciplinary status of corpus linguistics claim that "a corpus is not merely a tool of linguistic analysis but an important concept in linguistic theory" [2]. Others consider it a methodology, "no more, but also no less" [3].

Over the past 50 years, various definitions of corpus linguistics have been offered: corpus linguistics is a tool, a method, a methodology, a methodological approach, a discipline, a theory, a theoretical approach, a paradigm (theoretical or methodological), or a combination of these [4]. Improving methodological and analytical practice corpus linguistics defines its position as "both a new method (in terms of computer-aided descriptive linguistics) and a new research discipline (in terms of a new approach to language description)" [5].

Over a relatively short period, corpus linguistic methods have been embraced by a wide range of disciplines. In 2017, at the conference in the University of Birmingham it was recognized that corpus linguistics has a transformative effect on various areas of linguistics, among them historical linguistics, child language acquisition, critical discourse analysis, and others. Today, corpus technology and corpus-linguistic theorizing have advanced to such an extent that we are witnessing

an appearance of new sub-disciplines, for example, corpus pragmatics and corpus-based discourse studies.

Operating in the digital textual universe, in which the humanities and social sciences exist today, the classic definition of a corpus, as a digital database compiled by linguists for the purposes of linguistic analysis, has become increasingly difficult to uphold and corpus-linguistics will definitely merge with the digital humanities movement. The future is seen in creating multiple corpora with an emphasis on spoken, multilingual and nonstandard data [6].

Nowadays, it is admitted that corpus linguistics not only provides various branches of linguistics with sophisticated tools for analysis, but has also left a profound mark on linguistics as a discipline. Traditional descriptive work can now be carried out to higher empirical standards. More importantly, new areas of linguistic inquiry have been opened up to rigorous empirical investigation, and corpus-based research has given a general boost to usage-based theoretical frameworks of all kinds [7].

The advent of corpus linguistics in 1970s, as well as cognitive linguistics, was motivated by the development of computer technologies that considered language study as a means of creating models and a mode of gathering and analyzing data. Complementary in many ways, corpus linguistics and cognitive linguistics consider the language from the point of its social and psychological nature. Investigating organization, processing, and conveying information by language means, cognitive linguistics applies linguistic corpora as database for empirical research and adopts the system of methods and principles worked out by corpus linguistics.

A cognitive linguistics view on the language representation of sociocultural concepts and stereotypes requires access to big datasets to draw valid conclusions of the problem under investigation. As an indirect method of study, corpus analysis avoids bias of introspectiveness, providing means for polyfactorial research of social constructs that includes qualitative as well as quantitative methods. This research employs stylistic, syntactic, pragmatic, semantic examinations that verify research hypotheses and direct researchers to new theoretical assumptions.

The aim of our research is to determine the changes in the perception of sociocultural stereotypes and verify the cognitive structure of the stereotype on the results of corpus analysis.

2. Language Corpora: A Case Study

Since the first computerized corpus, the Standard Corpus of Present-Day Edited American English (the Brown Corpus), dozens of various corpora appeared.

The Brown Corpus, compiled by W. Nelson Francis and Henry Kučera in 1960s, contained a million words in 2000-word samples that represented a wide range of genres of published written texts in American English produced during the year. The pattern of the Brown Corpus was applied for compiling the Lancaster-Oslo/Bergen Corpus of British English in 1970s. In the 1980s the Collins-Birmingham University International Database (COBUILD) introduced a new progressive corpus design to read large quantities of printed texts of 20 million words. Created at the same time the Longman/Lancaster English Language Corpus outreached COBUILD in number. It comprised 30 million words. The British National Corpus is the product of collaboration of Longman, Oxford University Press and Computer Service, Cambers Harrap, the University of Lancaster, and the British Library. Appearing at the beginning of 1990s, the BNC started with 100 million words sampled from 1960 [8]. Now it counts a 100 million words collection of samples of written and spoken language from a wide range of sources.

Comparable in size and distribution of genres to the British National Corpus, the American National Corpus (ANC) is a large-scale database of American English. The ANC represents written materials and transcripts of spoken data produced from 1990 onward. As such, the ANC is the definitive record of contemporary American English for the creation of dictionaries, thesauri, and materials for linguistic research, as well as lexicography, computational linguistics research, corpus linguistic research, and a resource for the development of English language teaching materials. Appearing in October 2003, the ANC's first release of 11 million words of data was met with enthusiasm by the computational linguistics and linguistics communities [9]. The ANC's the

second and last release of 2005 comprises 22 million words. It is available for research and education for a licensing fee.

A fully open and unrestricted for any use, the Open American National Corpus (OANC) of 15 million words is an electronic collection of American English that includes texts of all genres and transcripts of spoken data produced from 1990 onward [10].

Composed in 1990 the Corpus of Contemporary American English (COCA) surpassed the database of ANC in size and growth: in 2017, COCA counts 560 million words with the year growth of 20 million words. Representing different written or spoken genres of American English COCA supports free access via web interface [11]. The corpus was created by Mark Davies, Professor of Corpus Linguistics at Brigham Young University, Provo, Utah. In addition to COCA, Professor Davies created other corpora: Time Magazine Corpus of 100 million words (1923-2006), Corpus of Historical American English of 400 million words (1810-2009), Global Web-Based English of 1.9 billion words (2012-2013), which is followed by New Web-Based (2017-2018), and 9 more different corpora.

One of the challenges to corpus creators and users is the fact that so many corpora quickly become “stale”. Corpora may represent the language from 10-20 years ago, but there is nothing from the last year or two. In May 2016, Professor Davies released the NOW corpus (“Newspapers on the Web), which uses automated scripts running every day to add texts to the corpus. Each day these scripts get 10,000-15,000 URLs from Google News. A corpus like NOW allows researchers to investigate billions of words of data and see language change as it occurs [12].

The list of newly emerging corpora is constantly enlarging with monolingual and parallel corpora in different languages, e.g. Arabic-French Parallel Text, Egyptian Arabic Informal Web Data, Amharic Text Resources, Somali Text Resources, etc. European languages are represented in new specialized corpora, such as German Children’s Handwriting corpus, TAC KBP English Source Material, etc. There is a strong tendency in many countries to create national language corpora representing comprehensively different communication forms of a national language.

2.1. Criteria of a Language Corpus

Corpus design criteria fully depend on its character. One of their essential characteristics is representativeness, which is seen as the ability of a corpus to reflect all features of the data domain relevant for linguistic analysis proportionally to the frequency of the phenomenon occurring in the data domain. The phenomenon under study may be any unit within the sphere of grammar or lexis, discourse analysis, language variation, etc. The understanding of the data domain is analogous to that of context in cognitive linguistics: it manifests semantic, grammatical and communicative cohesion of the word or collocation with their textual environment.

Another important criterion is balance. A well-balanced corpus must include spoken and written language. A corpus-creator chooses texts according to specific criteria, such as period of text production, full or sample presentation, etc.

To meet the requirements of investigators a well-constructed general corpus must be large enough. For work on the lexicon even a million words is insufficient. A general corpus must represent the language as faithfully as possible within the definite period [13].

The structure of the data representation is of great importance. Marked-up corpora basically involve indication of main divisions of the text, and may include part-of-speech tagging for every word, syntactic, or semantic annotation. Raw corpora, or untagged corpora is sequence of running words (tokens) of the text represented in a corpus. There are other special characteristics: type-token ratio shows variety of the vocabulary; lexical density indicates percentage of running words; key word in the context (KWIC) is concordance for input word in a definite context [14].

Corpus design and corpus tools for data retrieval and statistics for testing the data is crucial in a cognitive-linguistics research [15]. Considering criteria, we classify linguistic corpora in the following table.

Table 1. Classification of Corpora.

Criterion	Type
Size	small (to 1 million words); medium (from 1 to 10 million words); large (more than 10 million words)
Language of Texts	English, German, French, Swedish, Ukrainian, etc.
Quantity of Languages	monolingual, parallel, multilingual
Mode	spoken, written, balanced
Data Character	general, specialized (dialect, idiolect, sociolect, etc.)
Temporal characteristic	diachronic, synchronic
Annotation	marked-up (annotated), raw
Access	free, commercial, closed
Application	research, illustrative, learner, translation, etc.
Dynamism	dynamic, static
Authorship	one author, two and more

Our study of the cognitive aspect of a sociocultural stereotype is conducted on the data of three corpora created by M. Davies: Corpus of Contemporary American English, Corpus of Historical American English, and the Time Magazine Corpus. In the above-mentioned classification, the corpora are large and balanced, monolingual, annotated, well-adapted to conduct linguistic research.

2.2. Corpus approach

The great advantage of corpus-linguistic analysis lies in the objectivity that comes out of a large amount of naturally occurring language data produced by a variety of native speakers. A leading pioneer in the field of language corpora, Jan Svartvik noted that the corpus approach meets the requirements of practitioners in vast amounts of easily accessible and real data. He wrote, "Linguistic competence and performance are too complex to be adequately described by introspection and elicitation alone" [16]. Asserting the objectivity of corpus data, Starvik mentioned other significant advantages of corpus linguistics:

- easy verification of results, as well as sharing them with other researchers;
- possibility to study variation between dialects, registers and styles;
- opportunity to observe frequency of naturally occurred linguistic items;
- applicability of information in language teaching and language technology (machine translation, speech synthesis etc.);
- accountability of various linguistic features, not just selected features;
- free access to the data all over the world [16].

Viewing "naturally-occurring" language as a credible source for the investigation of linguistic structures, corpus linguistics provides an opportunity to obtain and analyze data quantitatively and qualitatively. By means of corpora, it is possible to track the development of specific features of the research subject and investigate discourse markers for the purposes of sociolinguistic or discourse-analytical research. But the major contribution of the corpus approach is to document the existence of linguistic constructs that are not recognized by linguistic theories [17].

The importance of empirical study in cognitive-linguistic research has been marked by D. Geeraerts (2006), R. Gibbs (2007), D. Glyn (2008), and others. These researchers attempted to bridge the gap between cognitive linguistics and corpus linguistics. Being a flexible framework, cognitive linguistics constitutes a cluster of many partially overlapping approaches, among them corpus approach [18], which grants efficient methods to study the meaning transference in communication. Though meanings do not present themselves directly in the corpus data, they can

be sampled and processed empirically. Actual usage appears in corpora in the form of spontaneous, nonelicited language data or as an online and elicited form in experimental settings [18].

The empirical character of corpus-linguistic approach, does not exclude intuition and introspection in the research. “Intuition will still be considered an essential input, it will play a big part, for instance, in selecting the phenomenon that the linguist will choose to investigate, and ultimately it will have an important role when it comes to evaluating the evidence in corpus”[19].

In fact, there are two acknowledged approaches in corpus analysis: corpus-based and corpus-driven. Corpus-based research undertakes the validity of pre-defined hypothesis that come from linguistic theory and the researcher’s intuition. Corpus-driven research is an inductive method that emerges from analysis and interpretation of corpus data. When the research plan allows both corpus based and corpus driven analyses, they are combined, in a hybrid approach cumulating the merits of deduction and induction [20].

Primarily, the methodological distinctions between these approaches were described by Elena Tognini-Bonelli. She writes: “Traditionally, linguistic theories are the result of reflection by a scholar after absorbing a great deal of experience of language and languages, and testing the implications and consequences with the reference to the intuition of competent or native speakers. The precise nature of language experience is impossible to ascertain because it figures only as part of the linguist’s credentials, and not as part of the methodology research. If a corpus was to be used to evaluate one of these classes of theories, the theory would have to be put into an explicit form so that those aspects of corpus patterning that it covered could be distinguished from those where the theory did not cover, or was the variance with, the evidence” [21]. The scholar sees any type of work, related to corpus analysis, corpus-based. She attributes the term corpus-based to the methodology that is applied to expound, test or exemplify the theories and descriptions that were formulated before. This methodology indicates where minor corrections and adjustments can be made. It is also a valuable source of quantitative evidence.

Typically, the corpus-based approach prioritizes the information received mainly by syntactic patterns [21].

The corpus-driven approach aims to derive linguistic categories from recurrent patterns and the frequency distributions that emerge from the language in context. Observed language phenomena are considered in the verbal context as well as situation context and wider context of culture [21]. Schematically these approaches can be shown in the following way:

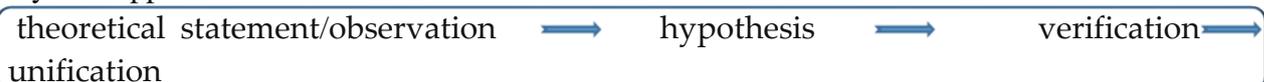
Corpus-Based Approach



Corpus-Driven Approach



Hybrid Approach



Our study presents a hybrid approach. While the cognitive status of the examined phenomenon as a fixed in culture typified image of a social group was ascertained introspectively, the alterations in its cognitive structure and changes in social perception of the stereotype were revealed in the course of corpus analysis. Based on distributive and statistical procedures the research incorporates quantitative and qualitative components of lexical units’ distribution in the text. The results are

formalized in distributional patterns, which reflect a repeated linguistic phenomenon in correlation with language as semiotic system.

3. Corpus Methods in Cognitive-Linguistic Research

The new data-mining affordances have evoked advances in the methodology of linguistic analysis. Corpus examination was adopted by cognitive linguistics as a methodology that provides rigorous means of describing culturally bound concepts and stereotypes within the framework of cognitive linguistics.

A wide range of corpus-application techniques have been developed by the German corpus linguists S. Gries and A. Stefanovich. The scholars' assumption that intuition does not give the full picture of the observed word sense, and that corpus analysis provides a detailed account of minimally different senses of a particular word, corresponds the standpoint of pragmatics from which meaning appears in communication. A researcher can draw "a behavioral profile" of the word on the corpus data. The profile is compiled by detecting morphological features, the syntactic properties, semantic characteristics of the referents of the elements co-occurring with the instances of the word, collocates co-occurring in the same clause, a paraphrase of the word's meaning in the citation. Corpus data help to distinguish senses in terms of formal patterns [22].

A framework of corpus analysis proposed by S.Gries and D. Divijak includes four consecutive steps:

- 1) the search and excerption of instances of the examined phenomenon;
- 2) the analysis of the found instances;
- 3) summarizing the results in tables;
- 4) processing the results statistically and descriptively [23].

Worked out by D. Glynn the multivariate statistical technique for corpus analysis arguably represents indices of a conceptual structure. Using a simple multivariate statistical technique allows to identify patterns of associations between the different lexemes, different speech contexts, and encyclopaedic semantic features of the words representing the concept. The received patterns assort with a range of the extralinguistic, formal, and semantic features [24].

Not restricted to quantitative dimensions, the corpus approach supports descriptive methods, that include hermeneutic procedure of the text interpretation. The aim of qualitative corpus analysis is retrieval of authentic examples of the language phenomenon under investigation, interpreting these empirical data in depth, and applying the ensuing insights to a broad range of intellectual exploration. The qualitative methods include genre studies, discourse and conversational analysis, ethnography of communication, contrastive analysis, semantic and pragmatic analysis [25].

Assuring accuracy of inference, quantitative analysis is applied to detect the variety of aspects, such as theme frequencies the examined phenomenon found, the prevalence of particular features and proportion of them in the cognitive structure of the phenomenon, etc. The raw numbers from the simple calculations indicate the prevalence of a phenomenon, while generalization is made on calculation of percentages and averages [26]. Combining quantitative and qualitative methods of corpus analysis assures a profound description of the phenomenon under investigation.

4. Results and Discussion

The application of corpus analysis in our study was determined by the objectives: 1) to verify the cognitive status (a stereotype) of the sociocultural phenomenon of flappers; 2) to trace its evolution.

The stereotype of *Flapper* refers to a standardized representation of a social group and a subculture of young American women in the 1920s that set up new standards of beauty and behavior. In a cognitive linguistic paradigm, a stereotype is treated as a mental construct that correlates with the linguistic picture of the world.

The unit of analysis is the word *flapper* that forms a grammatical and semantic environment, or semantic domain, that indicates functional knowledge of the word. To detect latent factors

influencing the meaning and implications of the polysemantic word, we examine broad and narrow contexts. In a broad, as well as a narrow one, the lexical meaning of the word acquires a specific sense that may include pragmatic component, known as connotation. The broad context represents the theme associated with specified reality. The narrow context, within the limits of syntactic structure or proposition, sets the typical collocability of the word that is determined by semantic and syntactic features of the word. Within the abundance of defined contexts, these two determine the behavioral profile of linguistic phenomena. While a thematic (or broad) context limits the language phenomenon by relevancy in meaning and implication, the phrase (or narrow) context is arbitrary for the word. Obviously, distributional knowledge is ultimately knowledge based on frequencies of occurrence, frequencies of co-occurrence, and dispersion characteristics [27].

The most identifiable sample for study is collocation. Collocations can be characterized in terms of their frequency, position and their idiomaticity. Specific collocational patterns are contextualized and examined in the detailed way using patterns W+W of the concordance [28].

Co-occurrence of the words in collocations is conditioned by semantic agreement, according to which the more substantial element in the collocation forms more combinations. Apart from specific differentiating semes, words in collocations, have at least one seme in common [29]. S. Johansson explains the co-occurrence of the elements in collocations by mental lexicon of a person that stores semi-preconstructed phrases that correspond the particular case [30]. Collocability of words is determined also by expressive, stylistic, pragmatic and situational factors.

In analyzed corpora, the word *flapper*, firstly mentioned in 1840 in the meaning of “a fan”. Over the next decades the word is noticed in the meaning of “the servant, drawing master’s attention to the conversation” (1890s) and “a fin” (1900s). The meaning of “a young girl” is noted in 1917.

Table 2. The frequency of the word *flapper* in analyzed corpora.

period	1840s	1850s	1860s	1870s	1880s	1890s	1900s	1910s	1920s	1930s	1940s	1950s	1960s	1970s	1980s	1990s	2000s
occurrence	1					1	1	1	56	17	9	15	5	12	7	6	22
frequency per 1m	0,06					0,05	0,04	0,05	2,8	0,69	0,37	0,61	0,21	0,50	0,28	0,21	0,74

The frequency of occurrence in the texts of 1920s is the highest. This is explained by popularity of the social and cultural phenomenon in this decade. Even at theatre performances, feminine personages of A. Dumas and B. Shaw were dressed as flappers.

In the succeeding years, the word is quite rare in occurrence. However, in 2000s the usage of word *flapper* has grown. The fact shows a renewed interest in the cultural phenomenon of flappers.

The period of popularity is represented in the Time Magazine Corpus by variety of themes:

- cinema (22) –about 20%;
- biographies of flapper-creators (19) – 17 %, among which are the names of J. Held (8) F.S. Fitzgerald (4);
- theatre (15) – 13%;
- high society life (10) – 9%;

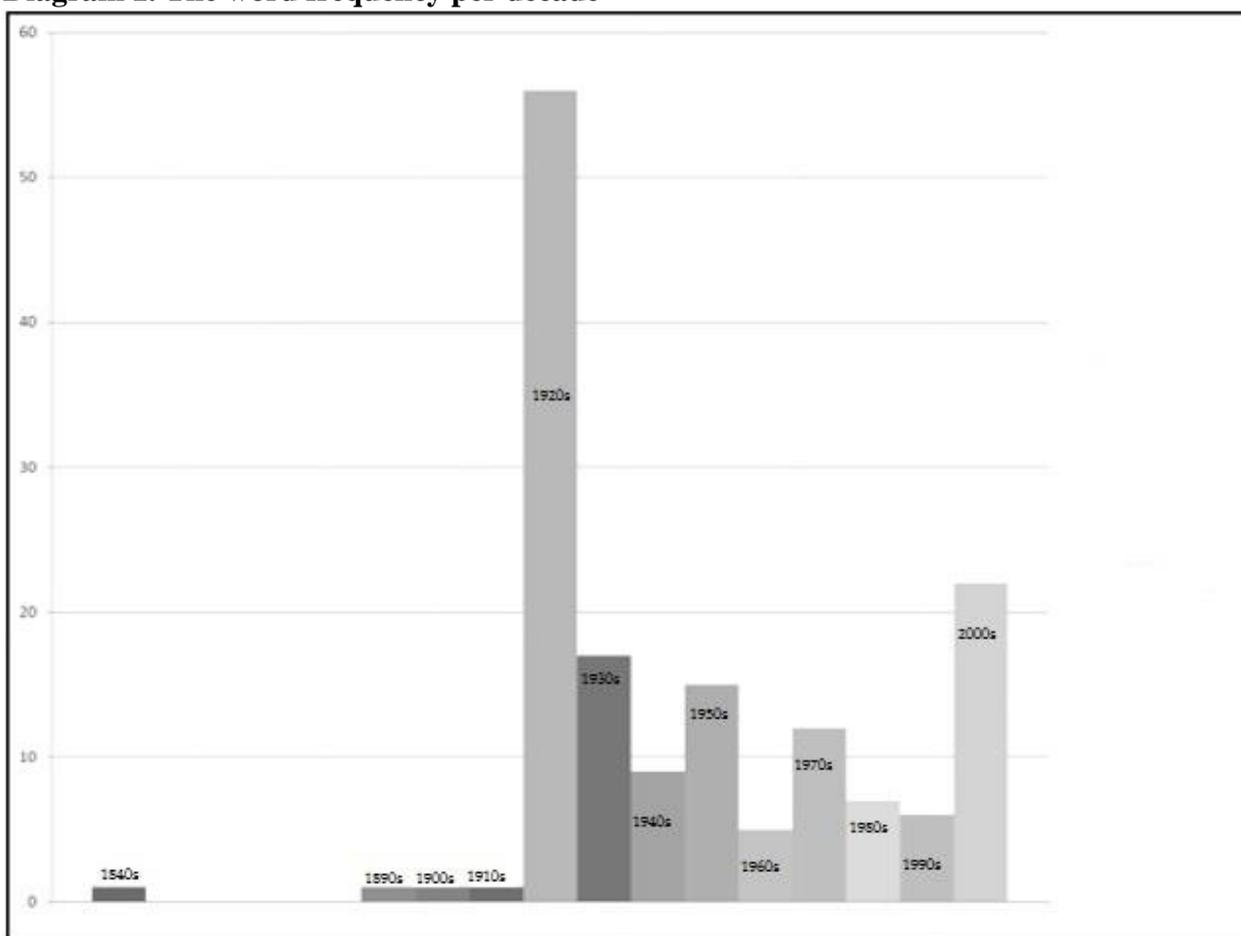
- politics and economy (10) – 9%;
- fashion (9) – 8%;
- chronical of events (6) – 5%.

The next decades witnessed disappearance of the phenomenon of flappers, thus the frequency of occurrence of the word *flapper* decreased, as well as the variety of themes. The concordance of the Time Magazine Corpus comprises 114 samples from which 112 instances represent the word in the meaning of “a young woman of 1920s”.

Table 3. Thematic frequency in Time Magazine Corpus

1920s	1930s	1940s	1950s	1960s	1970s	1980s	1990-2003
frequency of the word <i>flapper</i>							
44	12	17	10	11	10	5	3
variety of themes							
14	6	5	6	6	4	5	2
commonly occurring themes							
theatre	chronicle of events	high society events	cinema	fashion	cinema	history of 1920s	fashion

Diagram 1. The word frequency per decade



The representative frequency of the thematic contexts concerning cinema and fashion increases in the middle of the 20th century. The quantity proportion of the contexts, in which the word *flapper*

denotes the referent, to contexts presenting literary or cinema image, sets up to 1:3 (13:44). To receive this proportion, all analyzed contexts were divided into two groups: representing the referent and the image of a flapper.

The concordance of COCA (from 1990 till 2011) is 127 samples of the word *flapper*, from which 88 % represent the meaning of “a young and daring American woman of 1920s”. The remaining 12% of instances exemplify the meaning of “a valve”, “a servant”, or “ a fledgling”.

Table 4. The frequency of occurrence of the word *flapper* in COCA

section	TV transcripts	fiction	popular magazines	newspapers	scientific periodical
standard frequency per 1 m	0,17	0,46	0,44	0,21	0,17
frequency of occurrence in the section	15	39	40	18	15

The highest frequency of the word *flapper* is observed in popular magazines that depict iconic image of a typical flapper.

Table 5. Thematic frequency in COCA.

period	1990-1994	1999-1999	2000-2004	2005-2009	2010-2011
standard frequency per 1 m	0,25	0,25	0,23	0,42	0,32
frequency of occurrence in the period	26	26	24	43	8
themes	flapper’s description (11) historical period(6) cinema (3) history of cinema(2) fashion(3)	historical period (6) fashion (6) flapper’s description (4) history of cinema (2) style (2) entertainment (1) cinema(1)	historical period (11) fashion(7) exhibitions (2) clubs (1) cinema (1) flapper’s description (1)	flapper’s description (10) мода (10) historical period (2)	fashion (5) flapper’s description (3)
commonly occurring themes	flapper’s description, historical period	historical period, fashion	historical period, fashion	fashion, flapper’s description	fashion, flapper’s description

As seen from the above table, since 2005 the flapper image has activated the public interest. Iconized as a symbol of the “Roaring 20s”, a flapper is treated mostly as a fashionable icon, thus being represented most frequently in a thematic context of fashion. The typical elegant flapper style has influenced the collections of famous fashion designers, as Maggie Norris, Roberto Cavalli, and others.

The prevalence of external characteristics (18) over behavioral description (10) indicates the changes in the cognitive structure of the stereotype that was discovered introspectively.

Ascertained on the data of the Time magazine corpus, the proportion of elements in the cognitive structure of the stereotype is given in their rate per cent. To calculate percentage, the external characteristic is indexed as C_1 , the behavioral description is marked as C_2 ; the combination of traits is C_{1+2} ; whereas C is the criterion of the category.

Table 6. The proportion of elements in cognitive structure of the stereotype.

C_1	C_2	C_{1+2}
total in corpus		
106	50	39
54%	26%	20%
1920s		
7	29	8
16%	66%	18%

Generalizing the characteristics of the sociocultural stereotype *Flapper* manifests in distributions of the word *flapper*. In the Time magazine corpus and in COCA the distribution of the word manifests in collocations:

- with adjectives *popular, wild, thin*, that reflect stereotypic perception of behavioral and external characteristics;
- with nouns *dress, look, ball, skirt, dance, decade, era, hat*;
- with verbs *dress, name, accept, dance, wear*, that outline the mode of behavior.

In attributive collocations the word *flapper* fulfills a defining role (*Flapper* + N) or is defined by adjectives (A + *Flapper*).

Table 7. Descriptors of the word *flapper* found in corpora

period	descriptors		features described
1920s	frisky, robust, final, perfect, little, enduring, the veriest, US, Soviet, particularly degraded, bare-kneed naughty American, campus,		adolescence naughtiness persistence moral taint popularity
1930s	bad-mannered, perennial		persistence moral taint

1940s	little, fiftyish, Charleston-mad, hot and cold, 60-year-old		giddiness impulsiveness indifference persistence
1950s	75-year-old, a mad-cap, famed, flat-hipped		persistence publicity impulsiveness slenderness
1960s	F. Scott Fitzgerald, fun-loving, ultimate, fresh-faced, aging, 1920's		persistence gaiety adolescence historic period
1970s	money-hungry, man-chasing, jazz-age, free-spirited, pretty, vapid, iridescent, genuine, good-hearted		greediness freedom historic period attractiveness lack of taste style authenticity empathy
1990s	nubile, definitive, young, 40s, scantily clad, cherub-faced, Roaring 20's, perfect, bobbed-haired, wild, short-skirted, very passable, late night, quintessential, typical American, awakened, naked, twenties		sensuality beauty style extravagancy historic period adolescence
2000s	big-eyed, free-spirited, aging, Marlene Dietrich		beauty freedom persistence style

The descriptors presented in the table reflect the typical views of flappers. If in 1920s-1930s flappers were considered *frisky*, *naughty*, *degraded*, and *bad-mannered*, since 1950s they have been described as *mad-cap*, *wild*, and *free-spirited*. Contemporary Americans do not consider flappers immoral and spoiled as it was in 1920s.

Proper names (F.S. Fitzgerald, Marlene Dietrich) in the function of an attribute, indicate the creators and popularizers of a flapper culture in literature and cinema. Ambivalence in the perception of flappers, observed in texts of 1970s, demonstrates attributing to the girls diametrically opposed qualities: they are seen as *money-hungry*, as well as *good-hearted* girls.

The combinations with adjectives *veriest*, *definitive*, *genuine*, *quintessential*, *typical*, *passable* indicate typicality of flappers' characteristics, thus proving the status of the social and cultural phenomenon as a stereotype.

Functioning as modifiers, attributive clauses specify the opinion of society toward the behavior and intellectual abilities of the girls: *a flapper who would not be caught dead with a conventional notion about sex*; *a flapper who had four husbands and bought books with jackets to harmonize with her draperies*; *a flapper who literally sinks her teeth into nice young men*.

As a descriptor, the word *flapper* attributes to a defined noun the specific qualities intrinsic to the social category. In the pattern N₁ + N₂, the close connection between the elements results in semantic unity that possesses the qualities of both elements. Thus we observe generalization and abstraction of traits typical of stereotyping.

Table 8. The word *flapper* in the function of attribute

period	determined words	attributed characteristics
1920s	grandmother, comedy, Cleopatra, visitor, bandit, parts, world, voters, costumes, princes, saints, electorate	popularity, typicality, criminal activities, specific fashion style, female voting right
1930s	youth, vote, bride, fad	typicality adolescence, waywardness
1940s	parts, successor, era (4), days	popularity, characteristic of the historic period
1950s	era, tone, apparel, Jack	characteristic of the historic period, typicality, specific fashion style, source of revenue
1960s	trend (2), rage, dresses	specific fashion style, popularity
1970s	era	characteristic of the historic period
1980s	finery, era	typicality, characteristic of the historic period
1990s	frocks, dress (2) , look, period, drawings, cut, era	specific fashion style, characteristic of the historic period
2000s	dress (6), cut, gun molls, wigs, mode, outfits, dreams, style, film look, era, party skirt, gown	fashion style, characteristic of the historic period, criminal activities, specific behavioral mode

As seen from the table, the most frequently manifested semantic feature is appearance of flappers (*a flapper trend, mode, style*). In 1920s, the society associated flappers with a criminal world of gangsters (*a flapper bandit*), or with activities of suffragists (*voters, electorate*). Since 1940s, a flapper image becomes the symbol of the 1920s.

In collocations of the pattern *flapper* + of + N (*look of a flapper, the story of a flapper, the stereotype of flapper, influences of flapper, silhouette of flapper, the time of the flapper, the age of the flapper, the era of flapper*) the word *flapper* attributes its specific features to a determined noun. In these combinations, the word *flapper* characterizes appearance, behavior, or historic period.

The collocations of the patterns A + *flapper* + N, A₁ + A₂ + *flapper* + N, or A₁ + A₂ + A₃ + *flapper* + N, determine usually external feature that is intrinsic to flappers.

Table 9. Semantic features of the word *flapper* in attributive function

period	collocations	semantic features
1920s	the rapt flapper attention the bony, jazzy flapper figure	exaltation, slenderness,
1930s	famed flapper moron	popularity, folly
1940s	underslung flapper figure	leanness
1950s	the flapper era, flapper apparel	characteristic of the historic period, typicality of style
1960s	the flapper trend, the flapper dresses, winter's flapper rage	typicality of style, exaltation
1970s	low-cut flapper lines	staggering dress style
1980s	flapper finery, flapper era	characteristic of the historic period, typicality of style
1990s	boyish flapper creations, 20s flapper dress, 40s flapper era, Roaring Twenties flapper look, cloche flapper hats, a fringed flapper dress, the old flapper style, a black beaded chiffon flapper dress, the usual flapper'n'gangster show	slenderness, characteristic of the historic period, typicality of style
2000s	skimpy flapper outfits, a twenties flapper item, Maggie Norris flapper dress, a 20's style flapper dress, slim glittery flapper dress, gold vintage silk-velvet flapper dress, a crystal-hemmed flapper dress	typicality of style

The typical features presented in the table are slender figure (*bony, boyish, slim*), staggering style (*skimpy, low-cut*) of typical flapper attire (*fringed, beaded, crystal-hemmed, chiffon, silk-velvet*). In the Time magazine article of 1926 the wife of the president Madison (Dolly Madison) is categorized as a flapper reasoning from qualities intrinsic to flappers and assessed as frivolous: *Last week Mrs. Harry Atwell Colman, Washington woman, focused her lorgnets on Dolly Madison, wife of the fourth President. "If Dolly Madison were alive today, "said Mrs. Colman, "she would have been classed as a flapper.... She was frivolous, used rouge, dipped snuff, and... played cards for high stakes.... Quite another sort of woman was Martha Washington, a quiet lady, gentle and demure..."*

An evaluative attitude to flappers manifests itself in comparison to their qualities and actions. Cognitive similarity bases on bringing the conformity of heterogeneous essences into prominence.

The comparison with a flapper arises from appearance (*look like a flapper, like an iridescent flapper, much like vapid blonde flapper*) and behavior (*as the flapper dodges lectures, as the veriest flapper of today, as the money-hungry, men-chasing flapper, as naughty American flapper, as the good-hearted flapper, drink like a flapper, as a quintessential flapper*). A strongly negative attitude to a present-day politician is given in comparison to denunciation of flappers: *a tough-talking conservative who dismisses Townsend's candidacy the way a dowager would dismiss a flapper*.

The reflection of real situations in predications present a cognitive schema with actants corresponding to the Critic (society) and the Offender (flappers). Sometimes, especially in the texts of later years, a flapper is presented as the model that established new rules and standards in behavior and look. A role of trendsetter is represented in predications with the verbal phrases *give a vogue, set a mode, become vanguard of fashionableness*.

In propositions, the word *flapper* in the function of the subject is accompanied by such verbal phrases as *catch a magnet, play verbal hide-and-peek, annoy mother, marry chauffeur, beset, vote, disappear to start life together, become a symbol, begin to flourish, dodge "lectures", be really important, begin to decline, become vanguard of fashionableness, loose popularity, dance*. These are the examples of public judgment of flappers. They are depicted as active, seductive, extravagant, and even fatal for men. Popularity and symbolism of the flapper phenomenon reflects in phrases *become a symbol, begin to flourish, become vanguard of fashionableness, be really important*.

In the function of the object in propositions, the word *flapper* shows its qualities in correlation to the actions of the subject. The word *flapper* is used with the verbs *create, seduce, dismiss, give a vogue, not to entitle, involve with, measure against, be old for, picture, marry*, that outline mostly visual image of a flapper, or express condemnatory and suspicious attitude to them.

As a predicate, the word *flapper* realizes its conceptual meaning. The predicate indicates the characteristic of the category the subject belongs to. In this function, the word *flapper* endows the subject by outer and inner characteristics of itself. Categorizing the subject, the predicate can supply the assessment evident in the context. In analyzed examples, it is rejection, as well as approval.

As observed in Time magazine corpus, the collocations *flapper era* and *flapper dress(es)* acquired a stable character. In the first collocation, the word *flapper* indicates the most prominent feature of the modern period of American society that has accepted new standards of behavior laid down by flappers. The collocation *flapper dress(es)* reflects the tendency of the later years to visualize the image of a flapper. The self-sufficiency and freedom of flappers as categorical features are reflected in the word combinations *flapper Philippine prince* and *Soviet flapper*.

The distribution of the word *flapper* (F) in analyzed corpora is presented in the following table, where absolute majority (80%) is attributive word combinations.

Table 10. The distribution of the word *flapper* in analyzed corpora

attributive word combinations (F=flapper)				prepositional word combinations	predicative word combinations		
A + F	F + N	A1 + A2 + F + N	F + of + N		<i>flapper</i>	<i>flapper</i>	<i>flapper</i>
				N + Prep + <i>flapper/ flapper + Prep</i>	subject	object	predicate
61	63	21	8	+ N	15	11	6
153				6	32		
80%				3%	17%		

The collocations *Flapper Age*, *flapper era*, *flapper dresses* reflect symbolic character of the stereotype, which is associated by contemporary Americans with adventurous and prosperous period of 1920s.

4. Conclusions

Individual introspection in cognitive linguistic research is not enough to support the research hypothesis. Systematic collection of authentic language examples represented by corpora provides the powerful and rigorous means of testing researchers' hypotheses and validating introspective findings.

Our hypothesis of the stereotypic and symbolic nature of the sociocultural phenomenon "flapper" was verified in corpus analysis. The analyzed 241 corpora samples of the polysemantic word *flapper* represent the meaning of "a young, daring American woman of the 1920s" by 93% of the concordance.

Based on the results of corpus analysis, we can state that in 1920s the society assessed primarily the behavior of flappers. Flappers were conceived as disturbers of the public peace in 1920s.

In modern American society, the emotional evaluation of this sociocultural stereotype differs considerably from the assessment during the period of stereotype's formation. Now the external characteristics prevail in the cognitive structure of the stereotype.

In contemporary American culture, a flapper represents the symbol of the historic period, as well as a fashion trend. Flappers are considered constitutors of new standards and rules of behavior in modern American society. The categorical traits of flappers have become the guideline of public assessment of modern American women.

References:

- [1] Leech, G.N. (1992). Corpora and theories of linguistic performance. *Directions in corpus linguistics. Proceedings of Nobel Symposium 82nd*, Stockholm, Sweden, 4-8 August, 105-122.
- [2] Stubbs, M. (1993). British traditions in text analysis: From Firth to Sinclair. *Text and Technology: In honour of John Sinclair*; M. Baker, F. Francis and E. Tognini-Bonelli Eds., John Benjamins: Amsterdam, the Netherlands, 1-36.
- [3] Gries, St. (2007). Corpus-based methods and cognitive semantics: The many senses of to run. St. Gries, A. Stefanowitsch Eds. *Corpora in Cognitive Linguistics. Corpus-Based Approaches to Syntax and Lexis*. Mouton de Gruyter: Berlin, Germany, 57-99.
- [4] Taylor, Ch. (2008). What is corpus linguistics? What the data says. *ICAME Journal*, 32: 179-200.
- [5] Mukherjee, J. (2005). English ditransitive verbs: Aspects of theory, description and a usage-based model. In *Language and Computers 53*, Ch. Mair, Ch. F. Meyer and N. Oostdijk Eds. Amsterdam/New York: Rodopi.
- [6] McIntyre, D. (2017). Just what is corpus stylistics? In *Corpus Linguistics International Conference*, University of Birmingham, UK, 25-28 July 2017; 21. <http://paulslals.org.uk/ccr/CL2017ExtendedAbstracts.pdf>, Retrieved on 11.12.2018.
- [7] Mair, C. (2017). Downsizing and upgrading: Why we need more spoken, more multilingual and more nonstandard corpora. *Corpus Linguistics International Conference*, University of Birmingham, UK, 25-28 July 2017; 20. <http://paulslals.org.uk/ccr/CL2017ExtendedAbstracts.pdf>, Retrieved on 13.10.2018.
- [8] Crystal, D. (2003). The Corpus Revolution. *The Cambridge Encyclopedia of the English Language*. 2nd ed. Cambridge University Press: New York, USA.
- [9] Ide, N. & Suderman, K. (2004). The American National Corpus First Release. In *Proceedings of the Fourth Language Resources and Evaluation Conference (LREC)*, Lisbon, Portugal Centro Cultural de Belem, Lisbon, Portugal, 26-28 May 2004; 1681-1684.

- [10] *Open American National Corpus*. (2018). <http://www.anc.org/data/oanc/> Retrieved on 25.05.2018.
- [11] *The Corpus of Contemporary American English (COCA) and the American National Corpus (ANC)*. (2018). <https://corpus.byu.edu/coca/compare-anc.asp> (16, May, 2018). Retrieved on 25.05.2018.
- [12] Davies, M. (2017). The new 4.3 billion word NOW corpus, with 4–5 million words of data added every day. *Corpus Linguistics International Conference*, University of Birmingham, UK, 25-28 July 2017; 523-524. <http://paulslals.org.uk/ccr/CL2017ExtendedAbstracts.pdf>. Retrieved on 11.12.2018.
- [13] Crystal, D. (2003). The Corpus Revolution. *The Cambridge Encyclopedia of the English Language*. 2nd ed. Cambridge University Press: New York, USA.
- [14] Kenny, D. (2009). Corpora. In *Translation Studies*. Routledge Encyclopedia of Translation Studies, 2nd ed. M. Baker, G. Saldanha Eds; Routledge: London, UK, 59-62.
- [15] Grondelaers S., Geeraerts D. & Speelman D. (2007). A case for a cognitive corpus linguistics. In *Methods in Cognitive Linguistics*. M. Gonzalez-Marquez, I. Mittelberg, S. Coulson, M.J. Spivey Eds. John Benjamins Publishing: Amsterdam/ Philadelphia, 149-170.
- [16] Svartvik, J. (1992). Corpus linguistics comes of age. In *Directions in Corpus Linguistics. Proceedings of Nobel Symposium 82nd*. Stockholm, Sweden, 4–8 August, 1991. J. Svartvik Ed. Mouton de Gruyter: Berlin/New York, 1992; pp.7–14. doi.org/10.1515/97831108672757.
- [17] Biber, D. (2015). Corpus-based and corpus-driven analyses of language variation and use. In *The Oxford Handbook of Linguistic Analysis*, 2 ed. B. Heine, H. Narrog Eds. Oxford University Press: Oxford, UK, 193–224. doi.org/10.1093/oxfordhb/9780199677078.013.0008.
- [18] Geeraerts, D. (2006). Methodology in cognitive linguistics. In *Cognitive Linguistics: Current Applications and Future Perspectives*. G. Kristiansen, M. Achard, R. Dirven, F.J. Ruiz de Mendoza Ibanez Eds. Mouton de Gruyter: Berlin/New York, 21-49.
- [19] Tognini-Bonelli, E. (2001). *Corpus Linguistics at Work*. John Benjamins Publishing: Amsterdam/ Philadelphia.
- [20] Biber, D. (2015). Corpus-based and corpus-driven analyses of language variation and use. In *The Oxford Handbook of Linguistic Analysis*, 2 ed. B. Heine, H. Narrog Eds. Oxford University Press: Oxford, UK, 193–224. doi.org/10.1093/oxfordhb/9780199677078.013.0008.
- [21] Gries S. (2008). Th. Behavioral profiles: corpus-based approach to cognitive semantic analysis. In *Cognitive linguistics*. Evans V. and Pourcel S. Eds. John Benjamins: Amsterdam/Philadelphia, 57-75.
- [22] Biber, D. (2015). Corpus-based and corpus-driven analyses of language variation and use. In *The Oxford Handbook of Linguistic Analysis*, 2 ed. B. Heine, H. Narrog Eds. Oxford University Press: Oxford, UK, 193–224. doi.org/10.1093/oxfordhb/9780199677078.013.0008.
- [23] Gries, St. (2007). Corpus-based methods and cognitive semantics: The many senses of to run. St. Gries, A. Stefanowitsch Eds. *Corpora in Cognitive Linguistics. Corpus-Based Approaches to Syntax and Lexis*.: Berlin, Germany: Mouton de Gruyter, 57-99.
- [24] Glynn, D. (2008). Usage-based cognitive semantics. A corpus-driven empirical technique for conceptual analysis. *Proceedings of International Congress in Cognitive Linguistics*, Tambov, Russia, 8-10 October 2008, 55-58.
- [25] Hasko, V. (2013). Qualitative Corpus Analysis. *The Encyclopedia of Applied Linguistics*, 1st ed. C.A. Chapelle Ed. Wiley-Blackwell Publishing: Chichester, West Sussex, UK, 4758-4763.
- [26] Reef, D., Lacy, S., Fico, F. G. (2009). *Analyzing Media Messages: Using Quantitative Content Analysis in Research*, 2nd ed. Lawrence Erlbaum Associates, Publishers: Mahwah/ New Jersey/ London.
- [27] Gries, S. T. (2012). Corpus linguistics, theoretical linguistics, and cognitive/psycholinguistics: Towards more and more fruitful exchanges. *Corpus Linguistics and Variation in English. Theory and Description*. J. Mukherjee, M. Huber Eds. Brill, 41-63.
- [28] Tognini-Bonelli, E. (2001). *Corpus Linguistics at Work*. John Benjamins Publishing: Amsterdam/ Philadelphia.

- [29] Gak, V. G. (1998). K probleme semanticheskoy sintagmatiki [On the Problem of Syntagmatics]. *Jazykovye preobrazovanija*: Moscow, 272-297.
- [30] Johansson, S. (1995). On the Role of Corpora in Linguistic Research. *The European English Messenger*; Vol. 4(2): 19-25.

Bibliography:

- Biber, D. (2015). Corpus-based and corpus-driven analyses of language variation and use. In *The Oxford Handbook of Linguistic Analysis*, 2 ed. B. Heine, H. Narrog Eds. Oxford University Press: Oxford, UK, 193–224. doi.org/10.1093/oxfordhb/9780199677078.013.0008.
- Crystal, D. (2003). The Corpus Revolution. *The Cambridge Encyclopedia of the English Language*. 2nd ed. Cambridge University Press: New York, USA.
- Created by Mark Davies Corpora. <https://corpus.byu.edu/corpora.asp>, Retrieved on 29.04.2018.
- Davies, M. (2017). The new 4.3 billion word NOW corpus, with 4–5 million words of data added every day. *Corpus Linguistics International Conference*, University of Birmingham, UK, 25-28 July 2017; 523-524. <http://paulslals.org.uk/ccr/CL2017ExtendedAbstracts.pdf>. Retrieved on 11.12.2018.
- Gak, V. G. (1998). K probleme semanticheskoy sintagmatiki [On the Problem of Syntagmatics]. *Jazykovye preobrazovanija*: Moscow, 272-297.
- Geeraerts, D. (2006). Methodology in cognitive linguistics. In *Cognitive Linguistics: Current Applications and Future Perspectives*. G. Kristiansen, M. Achard, R. Dirven, F.J. Ruiz de Mendoza Ibanez Eds. Mouton de Gruyter: Berlin/New York, 21-49.
- Glynn, D. (2008). Usage-based cognitive semantics. A corpus-driven empirical technique for conceptual analysis. *Proceedings of International Congress in Cognitive Linguistics*, Tambov, Russia, 8-10 October 2008, 55-58.
- Gries, St. (2007). Corpus-based methods and cognitive semantics: The many senses of to run. St. Gries, A. Stefanowitsch Eds. *Corpora in Cognitive Linguistics. Corpus-Based Approaches to Syntax and Lexis*. Mouton de Gruyter: Berlin, Germany, 57-99.
- Gries, S. (2008). Th. Behavioral profiles: corpus-based approach to cognitive semantic analysis. In *Cognitive linguistics*. Evans V. and Pourcel S. Eds. John Benjamins: Amsterdam/Philadelphia, 57-75.
- Gries, S.T. (2012). Corpus linguistics, theoretical linguistics, and cognitive/psycholinguistics: Towards more and more fruitful exchanges. *Corpus Linguistics and Variation in English. Theory and Description*. J. Mukherjee, M. Huber Eds. Brill, 41-63.
- Grondelaers S., Geeraerts D. & Speelman D. (2007). A case for a cognitive corpus linguistics. In *Methods in Cognitive Linguistics*. M. Gonzalez-Marquez, I. Mittelberg, S. Coulson, M.J. Spivey Eds. John Benjamins Publishing: Amsterdam/ Philadelphia, 149-170.
- Hasko, V. (2013). Qualitative Corpus Analysis. *The Encyclopedia of Applied Linguistics*, 1st ed. C.A. Chapelle Ed. Wiley-Blackwell Publishing: Chichester, West Sussex, UK, 4758-4763.
- Ide, N. & Suderman, K. (2004). The American National Corpus First Release. In *Proceedings of the Fourth Language Resources and Evaluation Conference (LREC)*, Lisbon, Portugal Centro Cultural de Belem, Lisbon, Portugal, 26–28 May 2004; 1681–1684.
- Leech, G.N. (1992). Corpora and theories of linguistic performance. *Directions in corpus linguistics. Proceedings of Nobel Symposium 82nd*, Stockholm, Sweden, 4-8 August, 105-122.
- Johansson, S. (1995). On the Role of Corpora in Linguistic Research. *The European English Messenger*; Vol. 4(2): 19-25.
- Kenny, D. (2009). Corpora. In *Translation Studies*. Routledge Encyclopedia of Translation Studies, 2nd ed. M. Baker, G. Saldanha Eds; Routledge: London, UK, 59-62.
- Mair, C. (2017). Downsizing and upgrading: Why we need more spoken, more multilingual and more nonstandard corpora. *Corpus Linguistics International Conference*, University of

- Birmingham, UK, 25–28 July 2017; p. 20.
<http://paulslals.org.uk/ccr/CL2017ExtendedAbstracts.pdf>, Retrieved on 13.10.2018.
- McIntyre, D. (2017). Just what is corpus stylistics? In *Corpus Linguistics International Conference*, University of Birmingham, UK, 25–28 July 2017; p. 21.
<http://paulslals.org.uk/ccr/CL2017ExtendedAbstracts.pdf>, Retrieved on 11.12.2018.
- Mukherjee, J. (2005). English ditransitive verbs: Aspects of theory, description and a usage-based model. In *Language and Computers 53*, Ch. Mair, Ch. F. Meyer and N. Oostdijk Eds. Rodopi: Amsterdam/New York.
- Open American National Corpus*. (2018). <http://www.anc.org/data/oanc/> Retrieved on 25.05.2018.
- Reef, D., Lacy, S., Fico, F. G. (2009). *Analyzing Media Messages: Using Quantitative Content Analysis in Research*, 2nd ed. Lawrence Erlbaum Associates, Publishers: Mahwah/ New Jersey/ London.
- Stubbs, M. (1993). British traditions in text analysis: From Firth to Sinclair. *Text and Technology: In honour of John Sinclair*; M. Baker, F. Francis and E. Tognini-Bonelli Eds., John Benjamins: Amsterdam, the Netherlands, 1-36.
25.05.2018.
- Svartvik, J. (1992). Corpus linguistics comes of age. In *Directions in Corpus Linguistics. Proceedings of Nobel Symposium 82nd*. Stockholm, Sweden, 4–8 August, 1991. J. Svartvik Ed. Mouton de Gruyter: Berlin/New York, 1992; pp.7–14. doi.org/10.1515/97831108672757.
- Taylor, Ch. (2008). What is corpus linguistics? What the data says. *ICAME Journal*, 32: 179-200.
- The Corpus of Contemporary American English (COCA) and the American National Corpus (ANC)*. (2018). <https://corpus.byu.edu/coca/compare-anc.asp> (16, May, 2018). Retrieved on
- Tognini-Bonelli, E. (2001). *Corpus Linguistics at Work*. John Benjamins Publishing: Amsterdam/ Philadelphia.

**Сп. „Реторика и комуникации“, брой 38, януари 2019 г.
Rhetoric and Communications Journal, Issue 38, January 2019**