EMPIRIC STUDY OF VERBALIZED SOCIOCULTURAL STEREOTYPES

Lyubymova S.
PhD, Associate Professor,
Doctoral Student
ORCID ID 0000-0001-7102-370x
Kyiv National Linguistic University
elurus2006@gmail.com

Abstract. One of the objective empirical tool to draw conclusions in cognitive-linguistics research is corpus analysis. This article reports on data analysis applied in the study of language representation of sociocultural stereotypes. Application of this method is conditioned by the task to objectively describe categorial features of sociocultural phenomena and to verify the hypothesis of their stereotypic and symbolic nature. Sociocultural stereotypes are characterized by various explicit and implicit information comprising cultural schemata, emotional evaluation, allusions, and images, which can be revealed in the language. Since the appearance of language corpora, data-mining in investigation of represented in the language sociocultural phenomena has been simplified and methodology has become more sophisticated. Hybrid approach in this study is conditioned by the fact that cognitive status of the examined phenomena as fixed in culture typified images of social groups have been ascertained introspectively, while cognitive structure and changes in social perception of the stereotypes are revealed in the course of corpus analysis. Methodologically the work presents polyfactorial research that combines analysis of semantic and quantitative dimensions of verbalized sociocultural stereotypes. Based on distributive and statistical method, the research incorporates quantitative and qualitative techniques of verbal material. The results are formalized in frequency tables showing categorial features of the sociocultural stereotypes. As empirical method of study, corpus analysis avoids bias of introspectiveness. The acquired results and methods in study of linguistically represented stereotypes could assist the researchers interested in verbalized sociocultural phenomena.

Keywords: sociocultural phenomena, empiric study, corpus analysis, categorical features, cognitive stricture.

Introduction

With the advances of computer technologies, the data mining for various linguistic disciplines has greatly enlarged. At the turn of the previous century, we have witnessed a breakthrough in the use of computer corpora in linguistic research. 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 study of repeatable outcome and ensure valid judgments about linguistic phenomena.

Methodologically corpus analysis is defined as “a new method (in terms of computer-aided descriptive linguistics) and a new research discipline (in terms of a new approach to language description)” [1]. Over a relatively short period, corpus-linguistics 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, cross-cultural research [2]. Today, corpus technology and corpus-linguistic theorizing have advanced to such an extent that we are
witnessing an appearance of new sub-disciplines, as corpus pragmatics and corpus-based discourse studies.

The new abilities in data-mining has evoked advances in methodology of linguistic analysis. Corpus examination was adopted by cognitive linguistics as a methodology that provides rigorous means for describing culturally bound concepts and stereotypes within the framework of cognitive linguistics. As 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. Corpus analysis employs stylistic, syntactic, pragmatic, semantic examinations that verify research hypothesis and direct the researcher to new theoretical assumptions.

Since the first computerized corpus, the Standard Corpus of Present-Day Edited American English (the Brown Corpus), dozens of various corpora in different languages appeared. Composed in 1990, the Corpus of Contemporary American English (COCA) is one of the biggest in size and growth database: in 2017, COCA counts 560 million words with the year growth of 20 million words. Representing different written and spoken genres of American English, COCA supports free access via web interface [3]. 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 [4]. The Corpus has become popular among researchers in functional linguistics field [5].

To avoid “staleness” inherent in other corpora, Professor Davies released in May 2016 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 [6].

The aim of the research is to determine categorial features of sociocultural phenomena. The Corpus of Contemporary American English, created by M. Davies, is analyzed in this research. This corpus is large and balanced, monolingual, annotated, well-adapted to conduct linguistic research of cognitive aspect of a sociocultural phenomena.

Methodology

The importance of empirical study in cognitive-linguistic research, has been marked by D. Geeraerts (2006), R. Gibbs (2007), D. Glyn (2008), and others. The researchers attempted to bridge the gap between cognitive linguistics and corpus linguistics. Cognitive linguistics constitutes a cluster of many partially overlapping approaches, among them corpus approach, 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 [7]. The semantic meanings of lexical units that represent in different types of discourse sociocultural stereotypes are decoded and interpreted in the context.

There are two acknowledged approaches in corpus analysis: corpus-based and corpus-driven. Corpus-based research undertakes the validity of pre-defined hypothesis that came from linguistic theory and the researcher’s intuition. Corpus-driven research is inductive method that emerge from analysis and interpretation of the corpus data. When the research plan allows both corpus based and corpus driven analyses, they are combined, in a hybrid approach cumulating merits of deduction and induction [8]. This work presents the example of a hybrid approach. While the cognitive status of the examined phenomena as fixed in culture typified images of social groups were ascertained introspectively, the cognitive structure and changes in social perception of the stereotype were revealed in the course of corpus analysis. Based on distributive and statistical procedure the research incorporates quantitative and qualitative components of lexical units’ distribution in the text. The results are formalized in distributional patterns reflecting repeated linguistic phenomenon in correlation with language as semiotic system.
A wide range of corpus-application techniques have been developed by German corpus linguists S.Gries and A. Stefanovich. The scholars’ assumption that intuition does not give the full picture of 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.

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 [9].

Worked out by D. Glynn the multivariate quantitative technique for corpus analysis arguably represents indices of conceptual structure. Using this technique allows to identify patterns of associations between the different lexemes, different speech contexts, and encyclopaedic semantic features of the words representing sociocultural phenomena. Received patterns assort with a range of the extralinguistic, formal, and semantic features [10].

Not restricted to quantitative dimensions, 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 [11].

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. Received from simple calculations raw numbers indicate the prevalence of a phenomenon, while generalization is made on calculation of percentages and averages [12]. Combining quantitative and qualitative methods of corpus analysis assures a profound description of a phenomenon under investigation.

Results and Discussion

The object of my investigation is sociocultural phenomena, verbalized in American public culture discourse. The hypothesis is that provoking mostly negative emotions and criticized in public sociocultural phenomena become stereotypes. The sociocultural stereotypes are characterized by various explicit and implicit information comprising cultural schemata, emotional evaluation, allusions, and images, which are represented in the language.

The researchers assume that stereotypes can be revealed on all levels of language. As subjectively determined idea, a stereotype reflects on a syntactic level of language, in forms of judgements about certain features of stereotyped objects [13], or it can be represented by a lexical item that codifies and interprets a category of social world [14]. The evidence of a stereotype can be also found in connotations of the word, that make a stereotype prominent as “social meaning” in a definite context [15].

In cognitive linguistic paradigm, the stereotype is treated as a mental construct that correlates with the linguistic picture of the world. Nouns, that name stereotypes, denote the clusters of descriptive and evaluative features, including those of the character, physical appearance and typical behavior of the stereotyped groups. These nouns act as labels of social categories that codify extensive net of attributes, implications and beliefs. A word (or
a group of words), connected with a stereotype, stimulates activation of the stereotype’s content in a certain context, thus forming the center of semantic and cognitive associations.

The unit of analysis is a word that denotes a stereotype. It forms 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 word, broad and narrow contexts are examined. In a broad, as well as a narrow context, the lexical meaning of the word acquires specific sense that may include pragmatic component. The narrow context, within the limits of syntactic structure or proposition, sets typical collocatability of the word that is determined by semantic and syntactic features of the word.

Within the abundance of defined contexts, these two determine behavioral profile of linguistic unit. 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 [10].

The phenomena under study are W.A.S.P., Valley girl, and hipster, which represent certain social groups with their own ethics not always accepted in American society. The choice of subject of the research was conditioned by the aims of major work devoted to the study of cognitive features of stereotyping represented in language. The investigation of the stereotyped sociocultural phenomena is based on detection of individual associations, attitudes, and estimations, which in total represent the cultural and pragmatic meaning that incorporate the content of sociocultural stereotypes. Personified images of social groups are represented by generic character and visual image that are tend to be described on the data of Corpus of Contemporary American English [16].

The sociocultural stereotype represented by the acronym (abbreviation) WASP is composed by initial letters of words in a phrase “White Anglo-Saxon Protestant” pronounced together as a word. The acronym WASP denotes the upper class or elite group with disproportionately affluent economic and political advantages in American society. From the American Revolution to the 1930s, the WASPs, especially those with a clear ideology of close, upper-class ties, dominated America in all social aspects, in major areas like politics, economy, and culture. Nearly all immigrants before the 1950s were assimilated under an Anglo-conformity model. Since the beginning of the twentieth century, however, progressive WASPs supported greater meritocracy and a more diverse establishment, and gradually took the lead in promoting change. Thus, the WASP establishment has experienced a retreat, as more non-WASPs, especially Catholics and Jews, entered the elite group. Despite the decline, Protestant dominance, perpetuated in American institutions, still remains [17].

The referential meaning of the word WASP indicates the category of people in American society whose privileged status makes them criticized. WASPs are presidents (George Washington, Franklin Delano Roosevelt, George H.W. Bush, Lyndon B. Johnson), political leaders (Nelson Rockefeller, John Foster Dulles, John McCain), senators (Prescott Sheldon Bush, Henry Cabot Lodge), and businessmen (Averell Harriman, Phil Gramm). They are educated in elite private schools and universities, live in fashionable districts of New York, Philadelphia, and Chicago.

Used in its figurative sense, acronym WASP means a social winged insect, which has a narrow waist, a sting, and is typically yellow with black stripes. On the ground of identical spelling, a social category is compared to a very social kind of stinging predatory insects. A wasp image is a visual symbol of the stereotype, which implies ironic and disrespectful opinion of people that achieve leading positions in the society owing to their origin, rather than diligence and self-cultivation. To receive everything from the birth, without efforts, contradicts the realization of “American Dream”, thus evokes hostile attitude.

The total concordance (occurrence of the word) of the word WASP in the Corpus of Contemporary American English is 2259. The total number includes the instances of the
word in the meaning “a flying insect with a sting”. The following table presents the 190 cases studied by us.

Table 1. WASP in Corpus of Contemporary American English.

<table>
<thead>
<tr>
<th>years</th>
<th>quantity</th>
<th>attributes</th>
<th>features</th>
</tr>
</thead>
<tbody>
<tr>
<td>2015-2017</td>
<td>13</td>
<td>lazy</td>
<td>character – 87 %</td>
</tr>
<tr>
<td>2010-2014</td>
<td>25</td>
<td>conservative</td>
<td>appearance – 13 %</td>
</tr>
<tr>
<td>2005-2009</td>
<td>23</td>
<td>new</td>
<td></td>
</tr>
<tr>
<td>2000-2004</td>
<td>22</td>
<td>modern</td>
<td></td>
</tr>
<tr>
<td>1995-1999</td>
<td>23</td>
<td>classic</td>
<td></td>
</tr>
<tr>
<td>1990-1994</td>
<td>84</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

As we see, though nowadays the word occurs less, if to compare with the 1990s, it is still in use denoting elite social group in American society. The attributes found in the corpus showed that generic character of WASP is much more important for the society than its visual image. The word is characterized by mostly negative pragmatic meaning, which is definitely shown in the next example from the Corpus of Contemporary American English:

The acronym “Wasp”, from “White Anglo-Saxon Protestant”, is one many Wasps dislike, as it’s redundant - Anglo-Saxons are perforce white - and inexact. Elvis Presley was a white Anglo-Saxon Protestant, as is Bill Clinton, but they are not what anyone means by “Wasp”. Waspiness is an overlay on human character, like the porcelain veneer that protects the biting surface of a damaged tooth.

Another example of an American sociocultural stereotype is Valley Girl. Appeared in 1980s, the stereotype represents an image of a social category of young girls that belong to upper middle class, who live in a fashionable part of Los Angeles, San Fernando Valley. Seemingly beautiful girls, spoiled by parents’ wealth, they spend almost all their time shopping and appearance caring [18]. The image of an affluent girl was made famous through a movie starring Deborah Foreman, entitled “Valley Girl”, and the single of the same name, performed by Moon Unit Zappa. Later the term has become applicable to any American superficial girl, who cares only about social status and personal appearance. The girls use speech patterns associated with thoughtlessness. Their slang is characterized by the excess use of words such as “like, duh, rad, awesome, totally, and oh my god”. Most features associated with valley girls can be found among young, white people certain speech settings, as for example “in peer group youth interactions [19]. The following table presents in the stereotype of Valley Girl in the analyzed corpus.

Table 2. Valley Girl in Corpus of Contemporary American English.

<table>
<thead>
<tr>
<th>years</th>
<th>quantity</th>
<th>attributes</th>
<th>features</th>
</tr>
</thead>
<tbody>
<tr>
<td>2015-2017</td>
<td>8</td>
<td>real</td>
<td>character – 48 %</td>
</tr>
<tr>
<td>2010-2014</td>
<td>6</td>
<td>vintage</td>
<td>appearance – 52 %</td>
</tr>
<tr>
<td>2005-2009</td>
<td>6</td>
<td>perfect</td>
<td></td>
</tr>
<tr>
<td>2000-2004</td>
<td>9</td>
<td>stereotypical</td>
<td></td>
</tr>
<tr>
<td>1995-1999</td>
<td>15</td>
<td>suburban</td>
<td></td>
</tr>
<tr>
<td>1990-1994</td>
<td>10</td>
<td>middle class</td>
<td></td>
</tr>
</tbody>
</table>

As shown in the table, the stereotype still exists in American culture, and in the whole, it is used with the same frequency as in time of its emergence. The visual traits prevail in the image of a Valley Girl. The pragmatic meaning of a sociocultural stereotype
The faces of the audience grew expectant, their interest heightened by her silence. Bill, her buddy, sat smiling up at her from a folding chair in the front row. She said, high and emphatic like a valley girl, “Go-ad! Do you think Mom was right?” There was no sound but the wind, the splash of waves, the toot of a boat too far away to see. She maintained an amused face but knew the story had bombe. The meaning of the stereotype Valley Girl implies extravagant behavior and excessive expressiveness regarded as simpleton’s quality, which are criticized in American society.

The word hipster represents a sociocultural stereotype, which denotes an upper middle-class representative of a youth subculture, who is keen on fashion, modern art, alternative rock, and supports progressive ideas. Someone who is up to date, informed, and poised. The word hipster is generally associated with silly, pretentious slaves of fashion accepted in the society with mockery. Typically progressive in politics, hipsters in the newest time-frame are often renouncing the latest high-style fashions and gadgets. The moniker dates back to the 1940s when jazz was the “hip” music to listen. The table 3 presents the concordance of the word Hipster in the analyzed Corpus of Contemporary American English.

<table>
<thead>
<tr>
<th>years</th>
<th>quantity</th>
<th>attributes</th>
<th>features</th>
</tr>
</thead>
<tbody>
<tr>
<td>2015-2017</td>
<td>139</td>
<td>lifestyle (of a hipster)</td>
<td>character – 7%</td>
</tr>
<tr>
<td></td>
<td></td>
<td>travelling store</td>
<td>appearance – 93%</td>
</tr>
<tr>
<td>2010-2014</td>
<td>142</td>
<td>neighbourhood (hipster) culture</td>
<td></td>
</tr>
<tr>
<td>2005-2009</td>
<td>99</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2000-2004</td>
<td>55</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1995-1999</td>
<td>39</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1990-1994</td>
<td>21</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

As shown in the table, the stereotype’s popularity is growing. Its qualities affect various spheres of American life claiming to be global. In the following example from the Corpus of Contemporary American English, a physical characteristic of the stereotype is emphasized: The suffering is over. It’s time, it really is time, to dance. We booked the DJ, a hipster with slicked back hair and black cigarette pants tapered to elfish, pointy-toed shoes. The visual traits prevail over character.

All examined phenomena represent sociocultural stereotypes. The language evidence of this fact is shown in frequent collocates with the words “perfect, stereotypical, real, culture”. The words, denoting the stereotypes in discourse, code extensive information of different aspects of their existence in culture. Appeared at the end of the 20th century, these stereotypes still exist in modern American culture. The next stage in the study of the sociocultural stereotypes will involve the research of changes and transformations that these sociocultural phenomena undergo in the course of time.

Conclusions

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

Analyzed corpora samples validated the hypothesis of stereotypic character of the phenomena under study. Frequently manifested in the corpus semantic features correlates with cognitive structure of the stereotypes. Simple quantitative analysis results in proportion of visual and character components in cognitive structure of the stereotypes. The reflection of real situations in predications present cognitive schema with actants
corresponding to the Critic (society) and the Offender (a social group). Thus, corpus analysis is a valid method to examine cognitive aspects of represented in discourse sociocultural phenomena.

**ЕМПІРИЧНІ СТУДІЇ ВЕРБАЛІЗОВАНИХ СОЦІОКУЛЬТУРНИХ СТЕРЕОТИПІВ**

Любимова С. А.
Кандидат філологічних наук, доцент, докторант
Київський лінгвістичний університет
ORCID ID 0000-0001-7102-370x
вул. Велика Васильківська, 73, Київ, 03680
elurus2006@gmail.com

Анотація. Одним із об’єктивних емпіричних інструментів для отримання висновків у когнітивно-лінгвістичному дослідженні є корпусний аналіз. У статті представлені дані корпусного аналізу лінгвального матеріалу, що репрезентує соціокультурні стереотипи. Застосування цього методу зумовлене завданням об’єктивно описати категоріальне характеристики соціокультурних явищ та перевірити гіпотезу о їх стереотипності та символічності. Соціокультурні стереотипи характеризуються різноманітною експліцитною та імпліцитною інформацією, що включає культурні схеми, емоційну оцінку, алюзії та образи, які можна розкрити, аналізуючи мовні засоби їх реалізації. З часу появи мовних корпусів, отримання даних у дослідженнях вербалізованих соціокультурних явищ спрощується, а методологія стає більш досконалою. Гібридний підхід у дослідженні обумовлений тим, що когнітивний статус досліджуваних явищ як закріплених у культурі типізованих образів соціальних груп було встановлено інтерпретативно, тоді як когнітивна структура та зміни соціального сприйняття стереотипів виявляються в ході аналізу корпусів. Робота представляє поліфакторне дослідження, що поєднує аналізсмислових та кількісних вимірів вербалізованих соціокультурних стереотипів. Дослідження ґрунтується на дистрибутивно-математичній методиці, що поєднує якісний і кількісний прийоми аналізу лінгвального матеріалу. Результати аналізу надані у таблицях, що репрезентують категоріальні ознаки соціокультурних стереотипів. Як емпіричний метод дослідження, корпусний аналіз дозволяє уникнути упередженості інтерпретативного дослідження. Отримані результати та методи вивчення лінгвістично представленних стереотипів можуть допомогти дослідникам, що вивчають вербалізовані соціокультурні явища.

Ключові слова: соціокультурні явища, емпіричне дослідження, аналіз корпусу, категоріальні ознаки, когнітивна структура.

REFERENCES


Received: 18 August, 2019