www.volsu.ru

DOI: https://doi.org/10.15688/jvolsu2.2016.3.15

УДК 811.111’42
ББК 81.432.1-51

Дата поступления статьи: 12.05.2016
Дата принятия статьи: 24.08.2016

LEXICAL AND STATISTICAL PROCEDURES FOR IDENTIFICATION OF THEMATIC DOMINANTS OF AN AUTHORIAL TEXT IN MEDIA DISCOURSE

Vladislav Valeryevich Voskoboynikov
Postgraduate Student, Department of English Philology,
Volgograd State University
voskoboynikov.v.v@gmail.com, english_philology@volsu.ru
Prosp. Universitetsky, 100, 400062 Volgograd, Russian Federation

Abstract. This article presents the experiential data on the dominant thematic characteristics of texts concerning fashion by English speaking columnists Hamish Bowles and Suzy Menkes, obtained with the aid of semantic and statistical analysis (the AntConc concordancer developed by Dr. Laurence Anthony from Waseda University Japan). The fragments of the texts have been examined from the perspective of functional and semantic representations of nominativeness, process, and attributiveness as the basic mental and linguistic categories. Among the dominant thematic units of the nominativeness are: “fashion” (22 %), “person” (20 %), “art and science” (10 %), “time” (4,5 %), “buildings” (4 %), “space and movement” (3 %), “inter-object relations” (3 %), and “plants and animals” (2 %). The category of process comprises the thematic categories of “movement and transfer” (23 %), “creation and modification” (19 %), “mental processes” (18 %) and “cooperation” (6 %), “speech” (5 %), “possession” (5 %) and “similarity” (4 %). The category of attributiveness is represented by adjectives that belong to thematic categories of “evaluation” (40 %), “color and shades” (13 %), “toponymical features” (10 %), “temporal features” (8 %), “size” (6 %), “materials and fabrics” (4 %), “shape” (3 %), “similarity / difference” (2 %) and “restriction” (2 %); by adverbs, which are frequented by circumstantial adverbs, realizing the meaning as suggested by “where”, “when”, and “how” (39 %), degree adverbs, fulfilling the semantic function of comparison (28 %), connective adverbs, responsible for the logical connection between lexical units (17 %), focusing adverbs that implement the function of restriction (9 %), and stance adverbs that reflect the author’s position in a text (7 %). The data may be used to objectify the lexical and thematic features of the thematic space or serve as reference material for conceptual studies of an author’s style.

Key words: text theme, thematic space, nominativeness, process, attributiveness, lexico-statistic method.

Interest in the conceptual text structure saw its development throughout the 2nd half of the 20th century, and evolved the works on the text theory. Despite a substantial corpus of works covering this subject in Russian linguistic theory, it is worth noting that the studies of the thematic structure and semantics of a text have been largely approached in connection with the literary text. However, the development of journalistic discourse in the 21st century, driven by the advent of new digital means of information distribution coupled with increased institutionalization of the

media sphere and strengthening the role of major specialized media as a source of field-specific events, have contributed to increased attention in research concerning large thematic units beyond literary discourse. Recent research in this field include works on thematic structure of political journalistic discourse by O. Noskova [5], sketches of The New Yorker magazine by N. Petrova [6] and others. Development of computer-assisted methods of text processing has led to a new wave of scientific research, in which the traditional issue of thematic text structure relies on new methodological frameworks, namely corpus technologies and lexical statistics (see: A. Buranova [2], I. Belousov [1]).

Modern linguistics defines "theme of the text" as the notional nucleus of a text; the condensed and generalized contents of a text" [4, p. 17]. However, as noted by K. Belousov, "…thematic space of a text is comprised of statistically-determined entities of more or less unity". Such statistical entities are comprised of dominant unities, among which are elementary units. According to K. Belousov, "in this competition, it is the synthesis of the largest unities that may be called the theme of a text" [1, p. 16].

The current research is aimed at the examination of the *thematic dominants* – statistically frequent semantic groups of lexical units that ensure thematic integrity of a text through representative functional groups of nominativeness, process, and attributiveness in journalistic articles on fashion. The research relies on quantitative and semantic analysis methods for lexical and semantic categorization of the texts concerning fashion industry by two English-speaking journalists Suzy Menkes [Menkes] and Hamish Bowles [Bowles], written for Vogue magazine.

To study the manner in which individual style is manifested in the author's text it is essential to note the single or multiple authorial corpora in order to collect and analyze the statistical data as well as the evidence related to the characteristic use of certain semantic and grammatical categories. Preliminary processing of the text itself must correspond to the goals and objectives of the research, i.e. distribution of grammatical and / or functional categories in single author corpora calls for the so-called part-of-speech tagging (POS-tagging) involving the assigning of a part-of-speech or function mark to every element in the corpus either manually or by computer-aided procedures. An appropriately prepared corpus allows for the generation of word lists arranged by frequency for all of the words or just those with a certain attribute, thus studying the frequency of specific word meanings and the occurrence of word collocations in the author's texts. Statistically relevant lexical units upon further examination should be viewed as a form of concordance, generally defined as a context in which a word or a collocation has been used. Such a procedure is ensured by special software, a corpus manager or a concordancer which can immediately construct all possible contexts for the entry in question. The benefit of studying lexical units in their original context is the ability to determine their semantic, pragmatic and connotative features.

In current research practices part-of-speech tagging is conducted with the online-instrument Part of Speech Tagging (Standard), available at the Xerox linguistic tools page [1]. The results of processing a corpus are represented by a word list, in which each entry is assigned a "tag" – an identification mark of a hypothesized speech part – and a root form, e.g. +vaux (auxiliary verb), +advcmp (comparative adverb), etc. Word lists produced in such a manner are highly accurate; they do, however, demand manual verification for each entry and disambiguation where necessary.

Current study also relies on the AntConc concordance [2] developed by Dr. Laurence Anthony from Waseda University Japan. This software uses corpora, marked-up in accordance with the objectives of the study, in "*.txt" format. The functionality of the concordancer facilitates studying lexemes in their original context, search by word part, collocations or certain categories provided the corpus is sufficiently marked-up and the generation of frequency lists by specified criteria.

The preparation of the mini-corpus for our research consisted of POS-tagging of the text data before assigning each element a mark corresponding to the functional categories of nominativeness, process or attributiveness. The function of nominativeness is associated with the naming of objects; the function of process realizes the meaning of "action" either as a manifestation of energy or a physical, emotional or psychological

"state"; the function of attributiveness implies the meaning of quality or the character of a process (attribute of action), object (attribute of object) or attribute (attribute of attribute).

For studying the dominant thematic groups within the mentioned functions, all lexical units of the corpus are manually assigned a thematic tag and subsequently categorized into major thematic groups. This mark-up in the mini-corpus permits the generation of frequency lists of functional semantic and thematic categories. According to the statistical data obtained, the relative share of nominativeness units amounted to 23 %, 14 % in the case of process units, and attributiveness units secured 12 % of the total word span.

The comparative analysis of the frequency lists demonstrates that the relative shares of each category under examination are closely interrelated in the texts of each of the referenced journalists (nominativeness units – 22 % and 24 %, process units – 14 % and 13 %, and attributiveness units – 14 % and 11 % in the texts by H. Bowles and S. Menkes, respectively). Such slight variation in the figures (1–2%) in respect to the choice of functional categories in both authors' texts indicates the uniformity of lexical and grammatical units of the thematic space under examination. This provides the relevant and objective results within the frames of current research (see Table).

The thematic feature analysis in the category of nominativeness makes it possible to identify the following thematic dominants in the texts concerning fashion:

1. *Fashion (22 %)* – this category comprises lexical units with the meaning of "denomination of clothes and accessories" – 10 % (*coat, beret, boot*); "clothes manufacturing" – 6 %, including such denominations as "fabrics", "sewing techniques", "parts of clothes" (*corduroy, stitch, sleeve*); "general fashion notions" – 4 % (*fashion, line*); "judgmental estimate" – *2 %* (*glamour, elegance, luxury*).

2. *Person (20 %)* – this category describes various physical, mental and social features of a person:

a) *physical* (6 %) features include "body" – 4 % (*body, silhouette, waist*); "gender" – 1 % (*woman, lady, man*); "age" – <1 % (*baby, childhood*);

b) mental (5 %) features include the categories of "cognition" – 1,5 % (*concept, idea*); "traits" – 1,5 % (*verve, rebel*); "emotions" – 1 % (*amusement, anger*); "emotional and physical perception" – 1 % (*taste, comfort*);

c) social (9 %) features describe the following notions: "relationship and social roles" – 2 % (*boyfriend, wife*); "personal names" – <1 % (*Coco [Chanel], [Bettina] Graziani*); "ethnicity" – <1 % (*Italians, Americans*); "professional activities" – 6 %, which comprises the sub-categories such as "general notions" – 1 % (*work, career*) and "profession" – 5 %, which can be further subcategorized into "professions related to the field of fashion" – 3,5 % (*agent, mannequin, couturier*) and "other professions" – 1,5 % (*inventor, boss, waiter*).

3. *Art and science (10 %)* – includes the following subcategories: "general notions" – 2 % (*art, design*); "visual arts" with the subcategories "painting, graphic and photographic arts" – 2 % (*sketch, portrait*) and "pattern" – 2 % (*tile, checkerboard*); "art institutions" – <1 % (*museum*); "styles and genres" – <1 % (*classicism, surrealist*); "color and its characteristics" – 2 % (*turquoise, hue*); "science and scientific notions" – 1 % (*philosophy, X-ray*).

4. *Time (4,5 %)* – this category names the temporal characteristics: "general notions" – 2 % (*date, time*); "period" ~1 % (*day, week, minute*); "days of the week" – <1 % (*Monday, Sunday*); "seasons" ~ 1 % (*fall, spring*); "epoch" – <1 % (*eighties, sixties*); as well as "artifacts of the past" ~ 1 % (*history, heritage, legacy*).

5. *Buildings (4 %)* – this category includes such thematic categories as "residential buildings" – 1 % (*apartment, chateau, manor*); "architectural elements of buildings" – 2 % (*floor, stairway, garden*); "furniture" – 1 % (*rug, chair*).

Distribution of functional categories in the texts of the journalists under examination

Functional category	H. Bowles	S. Menkes
Nominativeness	22 %	24 %
Process	14 %	13 %
Attributiveness – adjectives and adverbs	14 % (10 % and 4 %)	11 % (8 % and 3 %)

6. *Space and movement (3 %)* – this thematic category is comprised of the following sub-categories: "movement" – 1 % (*step, walk*), "space" – 2 % (*position, area, site*) as well as "geographical and geological artifacts" – <1 % (*coast, seaside*).

7. *Inter-object relations (3 %)* – this category describes the units that portray "part-whole" relations – 1,5 % (*section, piece, version*) and "referential bond" – 1,5 % (*instance, example, mark*).

8. *Plants and animals (2 %)* – this thematic group includes the categories "birds and animals" – 1 % (*alpaca, ostrich*) and "plants" – 1 % (*cedar, tulip*).

The statistical analysis procedure allows the identification of the most frequent, i.e. dominant, thematic groups within the functional category of nominativeness in the examined mini-corpus. In addition to the groups listed, the texts encompass still other categories, such as "quantity", "shape", "sound" etc., each represented by a total share of less than 1 %. These statistically insignificant groups are therefore not included into the report.

Thematic systematization of the units that correspond to the function of process suggested categorizing all verb and verbal forms by their predicative function (i.e., notional and functional verbs) before assigning each element of the notional verbs a certain semantic class. This makes possible the analysis of the thematic organization of this functional group, which in turn revealed the following dominant thematic groups:

1. Movement and transfer of objects – this category makes up 23 % of the total number of verbs in the examined texts and embodies the concept of "movement" – 10 % (*move, rotate*) and the "transfer of objects" – 13 % (*bring, carry, spend*).

2. Creation and modification – this thematic groups comprises 19 % of the verbs and describes the processes of "creation" – 9,5 % (*make, produce*) and "modification" – 9,5 % (*amplify, soften*).

3. Mental processes – the dominant thematic sub-categories of this group, which account for 18 % of the total number of the verbs, are "cognition" – 8 % (*comprehend, forget, know*), "desideration and affectivity" – 3 % (*want, like, prefer*), "approval" – 2 % (*approve, reject*) and "naming" – 2 % (*call*).

4. Cooperation – this group makes up 6 % of the verbs and includes such sub-categories as "cooperation" – 2 % (*collaborate, negotiate),* "demonstration" – 2 % (*display, show*) and "connection" – 2 % (*tie, blend*).

5. Speech – this category comprises acts of speech and the vocal transmission of information, making up 5 % of the total verb count (*say, respond*).

6. Possession – this thematic group is represented by 5 % of the total number of the verbs (*own, possess*).

7. Similarity – this category conceptualizes the relation of similarity and representation and makes up 4 % of the total number of the verbs in the mini-corpus (*seem, look, represent*).

The category of attributiveness is represented by adjectives (9 % of all words in the mini-corpus) and adverbs (3 % of all words). Semantic analysis of the adjectives enabled the identification of the following dominant thematic groups:

1. Evaluation – this thematic group, manifesting the function of the author's evaluation of or attitude towards the described object, accounts for 40 % of the analyzed adjectives (*discomfiting, hefty, overwhelming).*

2. Colors and shades – 13 % (*green, blue, black*).

3. Toponymical features – 10 % (*Italian, American, British*).

4. Temporal features – 8 % (*mid-century, postwar, ancient).*

5. Size – 6 % (*maxi, broad, petite).*

6. Materials and fabrics – 4 % (*leather*, *velvet*, *satin*).

7. Shape – 3 % (*slinky, loose, flat).*

8. Similarity / difference – 2 % (*same, similar, different).*

9. Restriction – 2 % (*specific, secret, private).*

Adjectives, which actualize the meaning of "attribute of action" or "attribute of attribute", are represented by the following dominant semantic categories according to the classification of the adverbs by A. Downing [3, p. 505]:

1. *Circumstantial* adverbs, realizing the meaning as suggested by "where", "when" and "how", make up 39 % of the total number of adverbs in the mini-corpus. This category is frequented by the thematic categories of "time" with such sub-categories as "frequency"– 10 % (*often, once*), "time relation" – 6 % (*recently,*

still) and "moment" – 5 % (*now, then*); "manner" – 9 % (*carefully, discreetly*); and "space" with the following subcategories: "position in space" – 5 % (*inside, overhead*) and "direction"– 3 % (*forward, here*).

2. *Degree* adverbs account for 28 % of the adverbs and fulfill the semantic function of "comparison" – 12 % (more, less), "intensification" – 13 % (*totally, especially*) and "attenuation" – 2 % (*somewhat, a bit*).

3. *Connective* adverbs, which make up 17 % of the total number of the analyzed adverbs, ensure the logical connection between the lexical units in a text. The dominant thematic sub-categories are "concession" – 5 % (*however, though*), "reinforcement" – 4 % (*also, even*), "opposition" – 2 % (*instead*), "result" – 2 % (*thus*), and "equation" – 2 % (*meanwhile, too*).

4. *Focusing* adverbs secure 9 % of the total number of adverbs in the texts and represent the function of "restriction". The dominant sub-categories in this group are "reinforcement" – 5 % (*even*) and "restriction" – 4 % (*just, only*).

5. *Stance* adverbs make up 7 % of the adverbs. The function of this thematic category is the reflection of the author's attitude in the text itself. Among the thematic sub-categories of this group are "attitude" – 4 % (*hopefully, seemingly*), "viewpoint" – 1 % (*playfully*) and "possibility" – 1 % (*perhaps*).

The lexico-statistical analysis of the thematic organization of these journalistic texts with the aid of corpus technologies and statistical methods creates the possibility for the identification of the dominant thematic characteristics within a range of functional-and-semantic categories. The data obtained on the basis of authorial corpora related to a certain theme can be used for the objectification of linguistic features of textual unities, though it can be equally instrumental in studies concerning individual authorial characteristics through categorical and thematic systematization of the texts and comparison of the obtained characteristics with the results of a referenced common corpus.

NOTES

[1] Xerox – *Linguistic tools*. Available at: https://open.xerox.com/Services/fst-nlp-tools.

[2] AntConc – *AntConc Homepage*. Available at: http://www.laurenceanthony.net/software.html.

REFERENCES

1. Belousov I.K. Strategii strukturirovaniya tematicheskogo prostranstva teksta [Strategies of Structuring the Thematic Space of Text]. *Vestnik Permskogo universiteta. Rossiyskaya i zarubezhnaya filologiya*, 2014, no. 4, pp. 15-25.

2. Buranova A.I. Tematicheskaya organizatsiya dialektnoy rechi: kvantitativnyi analiz [Thematic Organization of Dialect Speech: Quantitative Analysis]. *Izvestiya Saratovskogo universiteta. Seriya: Filologiya. Zhurnalistika*, 2012, no. 3, pp. 35-38.

3. Downing A., Locke Ph. *English grammar : A University Course.* Second Edition. Oxon, Routledge, 2006. 610 p.

4. Moskalskaya O.I. *Grammatika teksta* [Text Grammar]. Moscow, Vysshaya shkola Publ., 1981. 183 p.

5. Noskova O.A. Smysloporozhdayushchee prostranstvo teksta kak kharakteristika lingvokognitivnogo stilya avtora-publitsista [The Meaning-Evolving Text Space as a Characteristic of the Publicist's Linguocognitive Style]. *Vestnik Kemerovskogo gosudarstvennogo universiteta* , 2012, vol. 4, no. 4 (52), pp. 80-84.

6. Petrova N.Yu. *Osobennosti publitsisticheskikh tekstov malogo formata: na materiale kratkikh zametok zhurnala "The New Yorker": dis. ... kand. filol. nauk* [Pecularities of Small-Size Journalist Texts: as Exemplified by The New Yorker Magazine Sketches. Cand. philol. sci. diss.]. Moscow, 2005. 195 p.

SOURCES

Bowles – Hamish Bowles – Vogue. Available at: http://www.vogue.com/contributor/hamish-bowles/.

Menkes – Suzy Menkes – The New York Times. Available at: http://www.vogue.ru/suzy_menkes/.

ЛЕКСИКО-СТАТИСТИЧЕСКИЕ ПРИЕМЫ УСТАНОВЛЕНИЯ ТЕМАТИЧЕСКИХ ДОМИНАНТ АВТОРСКОГО ТЕКСТА В МЕДИЙНОМ ДИСКУРСЕ

Владислав Валерьевич Воскобойников

Аспирант кафедры английской филологии,
Волгоградский государственный университет
voskoboynikov.v.v@gmail.com, english_philology@volsu.ru
просп. Университетский, 100, 400062 г. Волгоград, Российская Федерация

Аннотация. В статье предлагаются результаты анализа тематических доминант текстов на тему «мода» англоязычных журналистов Х. Боулза и С. Менкес (издание Vogue) с применением методов семантического и статистического анализа. Фрагменты авторских текстов изучались с позиции репрезентации функционально-семантических категорий номинативности, процессуальности и признаковости, актуализированных в категориальной семантике номинативных единиц текста. Тематическими доминантами категории «номинативность» являются группы: «мода» (22 %), «человек» (20 %), «искусство и наука» (10 %), «время» (4,5 %), «здания и сооружения» (4 %), «пространство и движение» (3 %), «межобъектные отношения» (3 %), «флора и фауна» (2 %). Категория «процессуальность» представлена тематическими категориями «движение и перемещение объектов» (23 %), «создание и модификация» (19 %), «ментальные процессы» (18 %), «кооперация» (6 %), «говорение» (5 %), «обладание» (5 %) и «подобие» (4 %). Категория «признаковость» репрезентирована прилагательными с тематическими доминантами «оценка» (40 %), «цвет и оттенки цвета» (13 %), «топонимические признаки» (10 %), «темпоральные характеристики» (8 %), «размер» (6 %), «материалы и ткани» (4 %), «форма» (3 %), «подобие / различие» (2 %) и «ограничение» (2 %); наречиями, среди которых наиболее частотны обстоятельственные наречия, реализующие признак «где», «когда» и «как» (39 %), степени, выполняющие семантическую функцию сравнения (28 %), соединительные, обусловливающие логическую связь между лексическими единицами (17 %), фокусирующие, реализующие функцию ограничения (9 %), и модальные, объективирующие авторскую позицию в тексте (7 %). Полученные с помощью методов корпусной лингвистики статистические данные могут быть использованы для объективации лексико-тематических характеристик текста как сложного тематического единства или выступать в качестве опорного материала при изучении характеристик индивидуально-авторского стиля речи.

Ключевые слова: тема текста, тематическое пространство, номинативность, процессуальность, признаковость, лексико-статистический метод.