Keyword-based Study of Thematic Vocabulary in British Weather News

Nataliya Bondarchuk [1], Ivan Bekhta [2], Oksana Melnychuk[3], Olha Matviienkiv[4]

[1] *Lviv Polytechnic National University, Lviv 79013, Ukraine*
[2] *Lviv Polytechnic National University, Ivan Franko National University of Lviv, Lviv 79013, Ukraine*
[3] *Rivne Medical Academy, Rivne, Ukraine, 33017, Ukraine*
[4] *Ivan Franko National University of Lviv, Lviv, 7900, Ukraine*

Abstract
The survey centers on the examination of keywords and related themes representing weather in four daily British newspapers (The Times, The Guardian, The Daily Mail, The Sun) between 2014 and 2017. The articles in this period mentioning weather news represent the corpus of our research. The goals of the research are the following: expose frequently occurring words (keywords) in the corpus, categorize them into groups according to relevant themes in the text, identify the quantitative content of each lexical-thematic group, as well as determine dominant themes of weather news. The computer software that was used to establish keywords is WordSmith Tools 7.0 with the British National Corpus as a reference corpus. A method for automatic cataloging of keywords is described. The corpus contains 746 324 words taken from 180 newspaper articles under research.

Despite the necessity of keyword study, thematic and quantitative analyses provide deeper insight into text-specific weather-related vocabulary and its textualizing role. The analysis of quantitative data helps to select two dominant lexical-thematic groups – "Weather extremes" and "Weather and people", giving evidence of central themes discussed in weather news. Hence, the resulting major themes are the depiction of adverse weather conditions affecting people's daily life; the representation of the effects of weather disasters on people and their environment. The obtained results highlight the link between a theme/themes and lexical level of the text proving the efficiency of keyword analysis in the research.
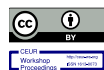
Keywords
keyword analysis, lexical-thematic groups, quantitative analysis, weather, WordSmith Tools, corpus, vocabulary

1. Introduction

Keyword analysis is a widely used method in various sciences and fields, in particular corpus linguistics. Egbert and Biber suggest that it is used "to identify the words that are especially characteristic of the texts in a target discourse domain" [1]. Keyword extraction is an optimal way while clustering, classifying, indexing and visualizing texts of different discourses, genres or text types. However, the application of keyword analysis to the text or corpus requires further interpretation of the results since keywords which are blindly extracted on the basis of their frequencies do not convey

---

relationship with other words/keywords or texts. Considering the key words at the linguistic level, the main idea, as stated by M.Scott, is that keyness is not language dependent, but text dependent [2]. Drawing from this, the advantage of the use of keyword analysis lies in the extraction of text-specific vocabulary. Therefore, in our research generation of keywords is a starting point of the analysis for further categorization of keywords into thematic groups representing weather. We understand thematic groups as groups of lexical units used within the text interchangeably to convey certain semantic meaning [3].

Prior surveys of such nature concentrated more on sentiment and quantitative analysis of weather vocabulary [4, 5, 6], corpus-based analysis of weather metaphors [7] and climate representation [8]. The absence of relevant computer-based studies on the topic constitute the topicality of the survey. This paper presents an interdisciplinary approach that incorporates linguistic and computer-based (statistical) techniques to the analysis of weather-related vocabulary to define dominant themes that are specific to weather news of British press, thus offering prospects for a better understanding of its contextualizing and textualizing role in newspaper discourse.

The primary objective of this paper is to present an argument for the definition of keywords according to different approaches. The secondary aim is to automatically extract keywords form the research corpus and organize them into lexical-thematic groups, as well as to find out their quantitative composition. As a result, dominant themes circulating in the texts of weather news may be defined.

2. Related works

The study of keywords is associated with the works of G. Matore [9], R. Williams [10], M. Scott [11, 12], Tribble [13], P. Baker [14; 15], T. Berber Sardinga [16], M. Bondi [2], A. Wierzbicka [17], L. Jeffriesand, B. Walker [18], J. Sinclair [19, 20], J. Firth [21], N. Fairclough [22], Gries [23], A.Wilson [24], M. Stubbs [25, 26], M. Nelson [27], G. Leech [28], M. Mahlberg [29], J. Culpeper [30], T. McEnery [31], G. Wilcock [32]. The notion of "key word" is multi-faceted and understood in different senses in various disciplines. From a sociological point of view, key words are part of the vocabulary of culture and society [10]. These are the words that have a special status, express an important social meaning and play a special role. From a linguistic point of view, they contribute to the long-lasting search for meaning [19] and are the most important units of the semantic and stylistic structure of the text. In corpus linguistics, keyword is defined as a word which occurs with significantly high frequency in one corpus when compared to some appropriate normative corpus (Scott, 1997; Scott & Tribble, 2006).

Paying attention to the importance of keywords in creating textual content and meaning, it is believed that keywords are lexical units with the greatest semantic content contributing to the structure and semantic framework of the text [19]. This makes keywords an effective method for identifying lexical characteristics of texts [34]. The new research shows a tendency for ambiguity in the terminological definition of keywords, which can be seen in three approaches:

Cultural (Matore, J. Firth, R. Williams, A. Wierzbicka). The first researchers (J. Firth, R. Williams), who discussed key words, were intuitively focused on words which, in their opinion, contain important notions reflecting social or cultural problems. Already in the 1930s, J. Firth proposed to study socially important words that could be called "focal" or "pivotal", and advocated an analysis of the distribution of words, the meanings of which characterize the society in specific contexts, with specific associations and values [21]. R. Williams tried to analyse modern culture by studying key words and established a close link between key words and discursive society [10]. However, while performing this analysis, he focused on historical and social macro contextual factors without paying special attention to text and genre and leaving the methodological tools of text analysis completely out of consideration.

Quantitative (M. Scott). Based on the concept of corpus linguistics, M. Scott differentiated key words by means of statistical characteristics. A word is deemed key if it is used in the text at least as many times as the minimum frequency of occurrence is estimated by the user [11], or key words are words whose frequency of occurrence in the text is exceptionally high, if we compare them with other words [33]. Identification of elements that are repeated with statistically significant frequency is not an analysis or interpretation of the text or corpus, but indicates the elements that need to be investigated

and explained. M. Scott distinguishes three types of key words: proper names, words that people themselves consider to be key words and are indicators of the "aboutness" of a particular text, and especially high-frequency words that are more indicative of style than of subject matter [12]. When talking about the topic and style of the text, as well as the role of key words in their identification, attention is paid to what semantic structures are indicated by key words and in what way the author's view influences them in the process of text creation [2]. M. Scott compares the theme ("aboutness") with the mental meta functions of M. A. K. Halliday [11]. The words gain meaning not from the link between the word and the meaning, but from the intrinsic interaction with other words. Later, P. Baker, using M. Scott's classification, described lexical key words (nouns, adjectives,) as subject words, i.e. words that can be used to identify the topic of the text [15]. However, the key words are not only elements of the conceptual, but also of the grammatical structure of the text. Apart from informational conjugation, they are indicators of communicative intention and micro- or macrostructure of the text. The text is stored in the memory in a set of key words, which are then revealed during its retrieval. Therefore, the notions of key and subject words are not identical.

Lexical-thematic groups include words that constitute components of one main thematic line, the elements of which realise a certain idea, while the key words either serve individual thematic blocks of the text (local thematic words) or implement, together with other text elements, the ideological idea of the whole work (universal key words) [22]. Consequently, key words of the lexical-thematic group are frequently varied units of the lexical organization of the text. They play an important role in the lexical structure of the text, they take part in shaping the content and creating the meaning for an adequate comprehension. The key word as a stimulus word, a source of textual associations, based on linguistic (paradigmatic, syntagmatic) and extralinguistic (thematic) links of lexical units, performs the function of a core, which directs the process of text comprehension. This approach appears to meet the tasks of our research the most.

Phraseological (M. Stubbs). Key words are defined as phraseological units and phrases that are constructed according to similar word models [25]. Assumption of key meaning through frequency can also be seen in word forms, lemmas, and word sequences. This definition can easily be applied to more complex units than words, pointing to current trends in descriptive and theoretical linguistics, in particular phraseology. In essence, key words are not necessarily individual words, they can be clusters or even phrases [20, 29]. A quite different approach was taken by M. Hoey, who, taking the category of text as a basis, showed that lexical links in a speech can be considered as indicators of text structure or potential acceleration or lexical models can reveal textual (as opposed to grammatical) models [36]. Key words are not necessarily the main ones in the textual sense, but they can help to understand the idea of the text by repetition. M.Toolan mentioned repetition as one of the key word figures, which has "a very rich semantic meaning" [37]. The key words are intended to focus the reader's attention on the necessary state of speech in the production of a coherent text. They can act as markers of coherence of this or that text and at the same time give the texts a unique author's style.

3. Methodology

The procedure of our analysis involves the choice of relevant material and methods of the research which combine a computer-based model of keyword analysis with traditional qualitative (in particular, thematic analysis) and quantitative analysis. Thus, in our framework we use three procedural steps: corpus compilation, keyword analysis, thematic and quantitative analyses.

The first step of the investigation was to compile the corpus of the research. The data used in the study is the corpus of weather news selected from British online daily newspapers between 2014 and 2017 (The Times, The Guardian, The Daily Mail, The Sun) which consists of 746 324 words taken from 180 newspaper articles. While compiling a corpus the following criteria were considered: firstly, the timeframe of four years (2014–2017) to represent recent use of the related themes, secondly, open access articles to be easily downloaded. Finally, the texts from theguardian.com/uk [38], thedailymail.co.uk [39], thesun.co.uk [40], thetimes.co.uk [41] were selected by using the search term "weather". The timeframe and the amount of words testify the representativeness and validity of the results. The dataset was made into a text file (.txt) and later imported to the WordSmith software.

Our next step was to extract keywords using this software. For this reason, the British National Corpus (hereinafter – BNC) was chosen as a suitable reference corpus since all data are specific to British English. In addition, BNC is one of the largest corpora which contains 100 million words of text from a wide range of genres (e.g. spoken, fiction, magazines, newspapers, and academic) [42]. The aim of a keyword analysis was to retrieve the words which are statistically relevant for the investigation. Consequently, we constructed two word frequency lists with the help of WordList tool: of a target corpus of weather texts and of a reference corpus (BNC) and generated a keyword list.

According to P. Baker, keyword extraction requires "a way that combines the strength of key keywords with those of keywords but is neither too general or exaggerates the importance of a word based on the eccentricities of individual files [14]. Therefore, we have taken into account both statistically significant (positive, high-frequency) and negative (unusually low frequency) keyword items using log likelihood ratio (Dunning 1993) [43]. The cells in the generated keyword list with negative keywords were shaded in red and had a negative log likelihood value. The reason for taking into consideration words with low frequencies is that our reference corpus consists of a collection of rather small texts. Consequently, the distribution of some words in the text may be uneven and some of the thematic lines might be lost. As stated by Gries, "corpora are inherently variable internally"[23] and low frequency keywords may help us find additional „local" themes of weather news. In this case, the issue may be solved by generating a wordlist for each single text in the research corpus, but it would be very time-consuming.

It is important to note here that the analysis is also restricted to content words only, which, being the units of meaning, we define to be directly related to the identification of the theme. Function words cannot demonstrate the link of lexical units and themes [45]. The extraction of keywords provides insights into their further grouping by themes since the potential of the key lexical units is realized within the whole text: the semantic influence of a single word (sign) is determined only by the whole text [44]. This idea is also supported by Morris and Hirst, who explain that "when a unit of text is about the same thing there is a strong tendency for semantically related words to be used within that unit" [46].

Thus, our last step was to group the keywords into lexical-thematic groups to define dominant themes related to the representation of weather in the news of online press. To this end, we worked out the following procedure: the context of each word from the keyword list is checked using the concordancer and the word is put into the appropriate group manually.

Two problematic issues that arose during this step were that of how to group 1) the words that could fit in multiple thematic groups and 2) the words that do not fit into any one of them. We applied a systematic hierarchical decision-making procedure and critical analysis to solve this issue: if a word could fit into several thematic groups, it was categorized into each of them; and if a word did not have an appropriate thematic group to be categorized, it was left out. In this step quantitative analysis was also used to further explore and identify dominant/prevalent themes of weather news by finding quantitative content of each thematic group. According to G. P. Cantos, in quantitative research, linguistic features are classified and counted [47]. In recent years commensurate attention has been paid to mixed-method studies of a text which use both quantitative and qualitative data [44]. As a result, the combination of keyword, lexical-thematic and quantitative approach in our research opened new opportunities for in-depth analysis of British weather news.

4. Experiment

Having compared two word frequency lists, we have obtained a computed list of keywords (Table 1). The list of keywords was limited to 300 words. The presented list in Table 1 provides only the first 45 keywords and their frequencies, as they will be further investigated during their thematic grouping.

Table 1

The list of keywords and their frequencies

Keyword/Frequency	Keyword/Frequency	Keyword/Frequency
SNOW 119	COOLER 15	CLIMATE 32

COUNTRY 57	SPELLS 13	TRAFFIC 11
SHOWERS 46	MORNING 39	WARNED 27
HEAVY 56	DOWNPOURS 15	GALES 15
WINDS 46	POLICE 25	DELAYS 17
FORECASTERS 38	CAR 24	COAST 17
CONDITIONS 57	CHILD 8	ICY 11
WARNING 37	PM 8	FLOODING 23
UPDATED 17	DORIS 8	SUNSHINE 28
ARCTIC 14	BRITS 18	STORM 34
WEEKEND 50	COMMUTERS 10	RAIN 68
HIGHS 13	CHRISTMAS 9	HEATWAVE 33
TODAY 40	FLOODS 34	EXTREME 39
HEAT 66	TRANSPORT 14	DAMAGE 28
LOCALS 21	DISASTER 19	BRITAIN 19

Scotts' classification of keywords into *proper nouns*, *aboutness keywords* and *high-frequency words* [12] is relevant for our study. The first type of keywords is usually represented in the corpus of our research by place names (Britain), names of nationalities (Brits) and names of storms (Doris). Aboutness keywords, the words that have semantic correlations with the main ideas and central themes of the text, are the most numerous. The third type constitutes words with high frequencies that are considered to be more indicators of style than theme. However, the objective of our study consists in grouping keywords into thematic groups rather than classifying them by types.

As defined earlier, our next step was to group the keywords by thematic categories representing weather, which consists in classifying the lexical units according to the thematic groups and quantifying them. Thus, the computed list of keywords was divided into five groups, each of which was classified thematically. As a result, the following groups were organized: *"Weather extremes", "Climate", "Weather and people", "Weather and nature", "Weather phenomena"*.

5. Results

The first thematic group "Weather extremes" consists of key lexical units that denote weather catastrophes that cause destruction of material objects, casualties and even death of people. In the corpus of the studied texts we find weather cataclysms of *hydro- (floods, tsunami)* and *atmospheric (blizzards/snowstorm, drought, hailstorm, heatwave, snow avalanche, showers, downpours, thundersnow, storm, duststorm) origin*, which accordingly constitute two subgroups of the group. This group also collects words denoting: *protective equipment* (shelter, sandbag); *locations* (nomenclatures: homes, businesses, region, village, area, country, town, adjectives: local, tropical, central, coastal); *means of transport and infrastructure* (road, bridge, building, traffic, speed, travel, highway, train, boat, van); *general vocabulary* (extreme, cancellation, delay, condition, incident, recovery, collapse, highs); *meteorological terms* (icy, Doris, cyclone, warning); *size/description* (adjectives: thick, heavy, massive, large, arctic), *victims of the disaster* (locals, mountaineer, immigrant, son, eyewitness, people, refugees, children, driver, traveller, residents, civilians, kayakers, victims); *organisations or political actors, officials* (minister, police); *other actors involved in the disaster* (commuters, coastguard, ambulance, volunteers, paramedics, army, evacuees) and *observation/analysis* (expert, forecasters, meteorologist); *the consequences of the disaster* (disaster, chaos, damage, mud, debris); *emotional perception of the disaster* (alarm, alert, threat, fear, risk, danger, alarmed, terrified, fearful), *actions* (drop, force, leave, move, block, trigger, halt, batter, collapse, damage, destroy, devastate, disrupt, kill, ruin, strike, warn).

The second thematic group "Climate" is composed of the words denoting results and consequences of climate problems, their interrelation and impact on weather conditions. The next Table 2 presents the words that form this thematic group.

Table 2
Thematic group "Climate"

Subgroups	Keywords
1) the effects of climate change	warming, flooding, greenhouse, ozone, hole, drought, deficit, fire, glacier, thawing
2) the causes of climate change	urbanization
3) scientific terminology	ElNino, research, study, science, scientist, emissions, fossil, climate
4) general vocabulary	impact, adaptation, conclusion, character, effects, assessment, committee
5) descriptive characteristics	disastrous, climatic, man-made, annual, fossil, human-induced, global, vulnerable
6) cause	exacerbate, pollute, emit, alter, modify, affect, devastate
	Total quantity: 41

The keywords, the meaning of which thematically reflects the impact of weather conditions on people's comfortable living environment, their safety and health, we refer to the third thematic group "Weather and people". The keywords which are included into this group are shown in Table 3.

Table 3
Thematic group "Weather and people"

Subgroups	Keywords
1) emotions	anger, grim, laugh, like, distressed, deranged, glee, happy, worried, devastated, shocked, hopeless, mad
2) transport and traffic situation	train, car, boat, vehicles, lights, tailback, sign, driving, speed, delays, diversions, cancellation
3) places	roads, motorway, parks, station
4) people	passenger, the elderly, commuter, children, driver, sun-lovers, sunbathers, sunseekers, children, adults
5) tourism/holidays	bookies, holidays, Christmas, Easter, football, vacations, weekend, match, holidaymakers, camper, traveler, festival-goer, beach, barbeque
6) health/safety	inhaler, death, dehydration, cardiovascular, illness, respiratory, diabetes, discomfort, faint, coughing, hypothermia, cramp, dizziness, chronic, irritation, sensitive
7) location	country, homes, businesses, region, village, area, town
	Total quantity: 77

Thematic group "Weather and nature" is characterized by the use of nominations of different species of plants and animals, as well as natural processes which are directly influenced by the weather. Table 4 shows the keywords that refer to this group.

Table 4

Thematic group "Weather and nature"

Subgroups	Keywords
1) plants/flora	eucalupts, pines, sycamore, oak, ash, beech
2) animals/fauna	chiffchaffs, dog, frigate, seabirds, fish, sparrow, starling, chaffinches, greenfinches, blackbird, woodpigeon, dove, tit, puffin
3) vegetable products	apples, strawberry, blackberry, fruit
4) agriculture	crops, harvest, agriculture
5) natural processes	breeding, flowering, leafing, ripening, budburst, growing, fruiting
6) other words related to the field of nature	wildlife, habitat, species, population, birds
	Total quantity: 40

The fifth thematic group "Weather phenomena" includes key lexical units denoting different weather phenomena and their manifestations. The lexical composition of this group is given in Table 5.

Table 5

Thematic group "Weather phenomena"

Subgroups	Keywords
1) atmosphere phenomena	rain, flurry, thunder, snow, winds, heat, temperature, humid, fog, tempest, gale, sunshine
2) meteorological terms	ElNino, Met, mercury, updated
3) description of weather phenomena	heavy, strong, clear, scorching, cold, conditions, humidity
4) evaluation of weather phenomena	unseasonable, awful, terrible, glorious, topsy-turvy, freak, crazy, driving, ropy, cooler
5) action and movement	rise, blow, batter, hit, smite, move, soar, bask, plummet, grip, batter, swoop, slam, loom, pummel, stop, finish, end, begin, start
	Total quantity: 54

Some of the keywords (morning, today, spells, pm) do not fit any of the groups and their number is not enough to group them separately, that is why there were not classified. Quantitative composition of each group is presented in the pie-chart diagram (Figure 1).
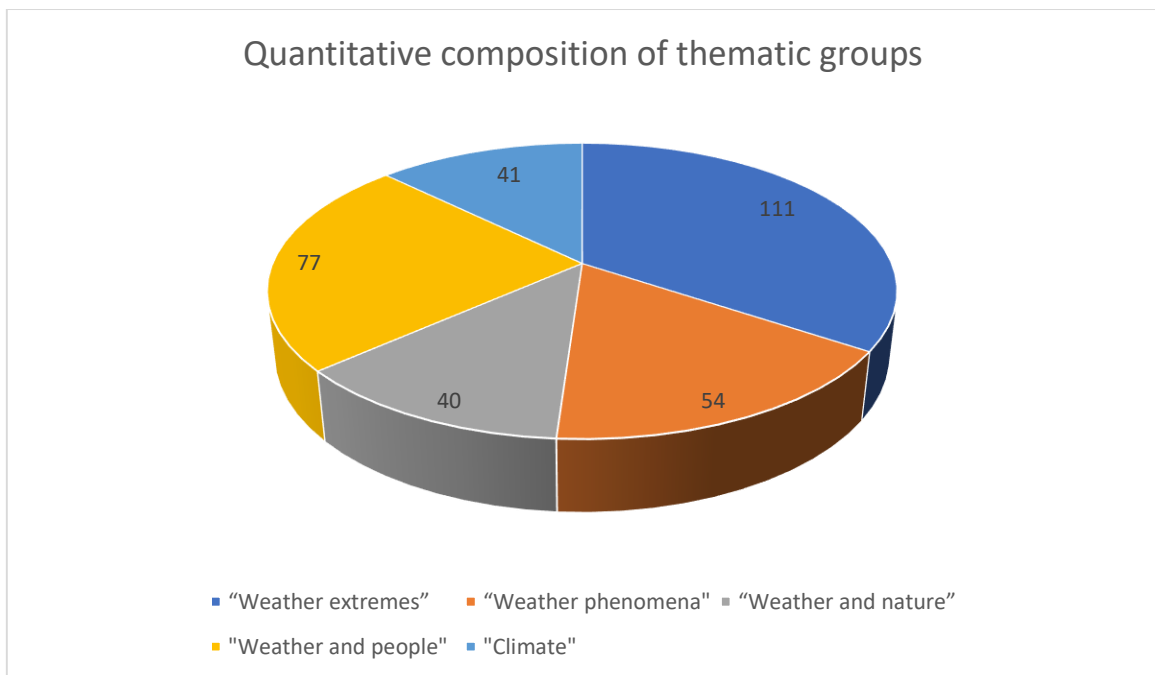
Figure 1: Quantitative composition of thematic groups

6. Discussion

Keywords appear as effective tools for the analysis of thematic focus/foci of the text or corpus. While the analysis of individual keywords does not provide a consistent and exhaustive text analysis, their grouping into categories according to the themes they represent establishes a link between the lexical level of the text and its themes.

A quantitative analysis of the keywords representing the weather allowed us to identify dominant thematic groups and, therefore, to recognize the themes that prevail in weather news of British online press. The identification of lexical-semantic groups [5] and related themes enables a more in-depth understanding of the texts of weather news. This survey paves the way for the development of more rigorous methodology for the analysis of relationship between keywords and other words in contexts. It would also be interesting to closer examine weather vocabulary through keywords extracted using other statistical tests (e.g. t-test, the Wilcoxon-Mann-Whitney test) or software packages and compare their results, as well as to find out whether dominant themes related to the representation of weather differ from newspaper to newspaper or particular season.

7. Conclusions

Having analysed the obtained data, we may conclude that weather news has anthropocentric character, since dominant thematic groups are "Weather extremes" and "Weather and people". Thus, within the corpus of the research two thematic lines can be outlined: (1) the depiction of adverse weather conditions affecting people's daily life; (2) depicting the effects of weather disasters on people and their environment. The overall study reveals the following conclusion: everything that happens in the sphere of weather, in any case influences people, their physical, moral, psychological and emotional state.

The method presented in this research provides more options for further collocational analysis of weather-related vocabulary in weather news of British online press on the basis of concordances and might be used as a cornerstone for the study of other vocabulary through keywords in different

discourses. Further scrutiny can be combined with a more interpretative approach to the analysis of weather news in line with stylistic research.

8. References

[1] J. Egbert, D. Biber, Incorporating text dispersion into keyword analyses. Corpora 14/1: 77–104, 2019

[2] M.Bondi, M. Scott (eds.), Keyness in Texts. Amsterdam: John Benjamins, 2010.

[3] D. Dedrick, R. E., MacLaury, G. V. Paramei, Anthropology of Color: Interdisciplinary multilevel modeling. (OAPEN (Open Access Publishing in European Networks).) John Benjamins Publishing Company, 2018.

[4]M. Thelwall, K. Buckley, G. Paltoglou, D. Cai, A. Kappas, Sentiment strength detection in short informal text. *Journal of the American Society for Information Science and Technology*, *61*(12), 2544-2558, 2010.

[5] N. Bondarchuk, I. Bekhta, Quantitative characteristics of lexical-semantic groups representing weather in weather news stories (based on British online press) In: Computational Linguistics and Intelligent Systems, COLINS, CEUR workshop proceedings, Vol 2870, 799-810, 2021.

[6] Q. Jun, N. Zhendong, Sh. Chongyang, Sentiment Analysis Model on Weather Related Tweets with Deep Neural Network. In Proceedings of the 2018 10th International Conference on Machine Learning and Computing (ICMLC 2018). Association for Computing Machinery, New York, NY, USA, (2018) 31–35. DOI:https://doi.org/10.1145/3195106.3195111

[7] I. Żołnowska, Weather as the source domain for metaphorical expressions. AVANT. Pismo Awangardy Filozoficzno-Naukowej, (1), 165-179, 2011.

[8] A. E. Stewart, J. K. Lazo, R. E. Morss, J. L. Demuth, The Relationship of Weather Salience with the Perceptions and Uses of Weather Information in a Nationwide Sample of the United States. *Weather, Climate & Society*, *4*(3), 2012.

[9] G. Matoré, La méthode en lexicologie. Domaine français. Paris: Marcel Didier, 1953

[10] R. Williams, Keywords. London: Fontana, [1976] 1983

[11] M. Scott, WordSmith Tools, Version 5. Liverpool: Lexical Analysis Software, 2008

[12] M. Scott, PC analysis of key words – and key key words. System, 25(2). 233-245, 1997.

[13] M. Scott, C. Tribble, Textual patterns: Key words and corpus analysis in language education. Amsterdam: Benjamins, 2006

[14] P. Baker, Querying keywords: questions of difference, frequency and sense in keyword analysis. Journal of English Linguistics 32(4): 346-359, 2004

[15] P. Baker, The question is, how cruel is it? Keywords, fox hunting and the house of commons. What"s in a word-list? Investigating word frequency and keyword extraction / ed. by D. Archer. Farnham: Ashgate. P. 125–136, 2009.

[16] T. Berber Sardinha, Using key words in text analysis: Practical aspects. DIRECT Papers 42, LAEL, Catholic University of Sao Paulo, 1999. 9 p

[17] A. Wierzbicka, Understanding Cultures Through Their Key Words: English, Russian, Polish, German, and Japanese. New York: Oxford University Press, 317 p, 1997.

[18] L. Jeffries, B.Walker, Key words in the press, English Text Construction 5(2): 208-29, 2012

[19] J. M. Sinclair, Document Relativity. Tuscany, Italy: Tuscan Word Centre, 2005

[20] J. Sinclair, The search for units of meaning. Textus. No. 9 (1). P. 75–106, 1996.

[21]J. Firth, Papers in Linguistics 1934–1951. London: Oxford University Press, 233 p., 1957.

[22] N. Fairclough, New Labour. New Language? London: Routledge, 2000.

[23] S. Th. Gries.„Dispersions and adjusted frequencies in corpora", International Journal of Corpus Linguistics 13: 403-37, 2008.

[24] A. Wilson, In press. „Embracing Bayes Factors for key item analysis in corpus linguistics", in A. Koll-Stobbe and M. Bieswanger (eds.), New approaches to the study of linguistic variability. Frankfurt: Peter Lang, 2013.

[25] M. Stubbs, „Three concepts of keywords", in M. Bondi and M. Scott (eds.), pp. 21-42, 2010

[26] M. Stubbs, Text and Corpus Analysis. Blackwell, Oxford, 288 p, 1996.

[27] M. Nelson, A corpus based study of business English teaching materials. Unpublished PhD Thesis. University of Manchester, 2000.

[28] G. Leech, P. Rayson, A. Wilson, Word frequencies in written and spoken English: Based on the British National Corpus. London: Longman, 2001.

[29] M. Mahlberg, „Clusters, key clusters and local textual functions in Dickens", Corpora 2(1): 1-31. Republished 2012 in D. Biber and R. Reppen (eds.), Corpus Linguistics, Part 3: Phraseology. London: Sage, 2007.

[30] J. Culpeper, Keyness: Words, parts-of-speech and semantic categories in the character-talk of Shakespeare"s Romeo and Juliet, International Journal of Corpus Linguistics 14(1): 29-59, 2009.

[31] T. McEnery, Keywords and moral panics: Mary Whitehouse and media censorship, in D. Archer (ed.), pp. 93-124, 2009.

[32] G. Wilcock, Introduction to linguistic annotation and text analytics. (Introduction to linguistic annotation and text analytics.) San Rafael, Calif.: Morgan & Claypool, 2009.

[33] . Dilai, M. Dilai, Automatic Extraction of Keywords in Political Speeches, 2020 IEEE 15th International Scientific and Technical Conference on Computer Sciences and Information Technologies, CSIT 2020 - Proceedings, 2020, 1, art. no. 9322011, pp. 291-294.

[34] D. Biber, R. Reppen, The Cambridge handbook of English corpus linguistics, 2020.

[35] M. A. K. Halliday, An Introduction to Functional Grammar. 2nd edn. London: Edward Arnold, 1994.

[36] M. Hoey, The discourse colony; a preliminary study of a neglected discourse type. Talking about Text. Discourse Analysis Monograph 13, English Language Research / University of Birmingham. 1986. P. 1–26.

[37] M. J. Toolan, Narrative progression in the short story: A corpus stylistic approach. Amsterdam: John Benjamins Pub. Co, 2009.

[38] The Guardian, 2014–2017 [Electronic resource]. URL: https://www.theguardian.com/uk.

[39] The Daily Mail, 2014–2017 [Electronic resource]. URL: https://www.thedailymail.co.uk.

[40] The Sun, 2014–2017 [Electronic resource]. URL: https://www.thesun.co.uk.

[41] The Times, 2014–2017 [Electronic resource]. URL: https://www.thetimes.co.uk.

[42] BNC. URL: https://www.english-corpora.org/bnc/

[43] T. Dunning, Accurate methods for the statistics of surprise and coincidence, Computational Linguistics 19(1): 61-74, 1993.

[44] B. Fischer-Starcke, Corpus linguistics in literary analysis: Jane Austen and her contemporaries. London: Continuum, 2010.

[45] L. Mastropiero, Corpus stylistics in heart of darkness and its Italian translations: Bloomsbury Publishing, 248 p., 2017.

[46] J. Morris, G. Hirst, Lexical cohesion computed by thesaural relations as an indicator of the structure of text. Computational Linguistics, 17(1), 21–48, 1991

[47] G. P. Cantos, Statistical methods in language and linguistic research. Oakville, CT: Equinox Pub. Ltd., 2012.

[48] N. Hrytsiv, T. Shestakevych, J. Shyyka, Quantitative Parameters of Lucy Montgomery"s Literary Style, In Computational Linguistics and Intelligent Systems. Proceedings of the 5th International Conference on COLINS 2021. Volume I: Workshop. Kharkiv, Ukraine, April 22-23, 2021, CEUR-WS.org, pp. 670-684, 2021.

[49]L. Jeffries, B. Walker, Choice is the Word of the Hour. In Keywords in the Press: The New Labour Years (pp. 67–92). London: Bloomsbury Academic, 2018.