

CORPUS ANALYSIS OF COLLOCATIONS WITH *HEALTH* IN THE MEDIA TEXTS

Olha Sivaieva,

Polissia National University
7 Staryi Bulvar St, Zhytomyr, 10008, Ukraine
olgasivaeva1987@gmail.com
ORCID iD 0000-0003-0123-9446

The media is an influential tool in shaping public's opinion about HEALTH and its basic components. As this topic has been of great importance lately, the corpus study of media texts with HEALTH can reveal verbal means of how this lemma is depicted by journalists as well as what urgent social concerns are connected with HEALTH and what issues readers are aware of.

The research is aimed at studying collocations with HEALTH in The Guardian and The Mirror newspapers, focusing on the comparative analysis of them presented in the broadsheet and tabloid. Sketch Engine has been used to investigate the lemma HEALTH in both newspapers, which helps to disclose the linguistic means used to outline the concept HEALTH.

The findings of the study prove that despite the use of modifiers and verbs with HEALTH common for both newspapers (e.g., mental, physical, public; improve, protect, affect), The Mirror presents a wider choice of collocations with HEALTH compared to The Guardian, whereas the lexeme HEALTH is more frequently used in the latter — 2,367.84 per million as to 1,615.61 per million in the first one. Furthermore, the tabloid presents a larger range of health subjects while the broadsheet displays a narrower area of the topic with a more conservative point of view.

Key words: HEALTH, collocations, corpus analysis, media, The Guardian, The Mirror.

Сіваєва О. С.

Корпусний аналіз словосполучень з *HEALTH* у медіа текстах

Стаття присвячена корпусному аналізу концепту HEALTH в газетах The Guardian та The Mirror. Головна увага зосереджена на порівняльному аналізі словосполучень з лексемою HEALTH, представлених у британських суспільно-політичній пресі й таблоїдах. Для дослідження лексеми HEALTH в обох газетах створено корпуси текстів, які проаналізовано за допомогою комп'ютерної програми Sketch Engine. Виокремлені мовні засоби відображають спільні та відмінні аспекти концепту HEALTH, представленого у двох діаметрально протилежних газетах.

Ключові слова: HEALTH, словосполучення, корпусний аналіз, медіа, The Guardian, The Mirror.

Introduction

Corpus research enables language peculiarities analysis through using computer technologies. Our research is aimed at studying the corpus of media texts containing *HEALTH* collocations.

The aim of the research is to determine specific linguistic features of the lemma *HEALTH* in British broadsheets and tabloids. Considering the aim of the study, the following tasks have been set forward:

1) to create corpora of newspaper articles containing language units of the lemma *HEALTH* in broadsheets and tabloids;

2) to determine peculiarities of the collocations containing the lemma *HEALTH* in British broadsheets and tabloids;

3) to compare two corpora analyzing the collocations containing the lemma *HEALTH* in British broadsheets and tabloids and determining their similarities and peculiarities.

The object of the study is newspaper discourse containing the lemma *HEALTH*.

The subject of the study is the collocations with the lemma *HEALTH* in the newspaper discourse.

Theoretical background

Media discourse is defined as interactions through a spoken or written broadcast platform, the discourse being oriented to a non-present viewer, listener or reader. Media discourse is understood as a public form of interaction being manufactured and on-record. The language of newspapers was studied by Rhea analyzing word order, the role of newspaper audiences and the influence of lexical and syntactic choice on creating and manipulating the meaning [13, 312–331]. Modern mass media pay great attention to health issues having an impact on the reader's health care interests. Researchers all over the world are interested in media influence on individual

and public health, the purpose of these effects and consequences of this influence.

Corpus linguistics is the study of linguistic phenomena based on large collections of texts read by machines; they are called corpora [10, 217–232]. The methods used in corpus linguistics came into use in the 1960s whereas the term itself appeared in the 1980s [11, 149–172]. The focus of corpus linguistics is a group of methods or procedures aimed at studying languages. On the other hand, corpus linguistics is seen as an independent theory. Sandra Kübler and Heike Zinsmeister comment in their book “Corpus Linguistics and Linguistically Annotated Corpora” applied [7, 248–260] that “the answer to the question whether corpus linguistics is a theory or a tool is simply that it can be both. It depends on how corpus linguistics is.” Corpus linguistics has considerably developed due to the possibility to process languages with the help of computers. Thus data can be obtained quickly and easily, being the format good for the analysis. However, corpus linguistics is not to get the data, but to analyze them [3, 143–145]. The major schools of Corpus Linguistics are in Lancaster, Birmingham, Nottingham, and Arizona. Different approaches considered [9, 1], corpus linguistics is dealing with some set of machine-readable texts which is deemed an appropriate basis on which to study a specific set of research questions.

Studying collocations began in the 1950s. According to Firthian, collocations are empirical statements connected with the predictability of word combinations, they measure the “mutual expectancy” possible between words and their statistical influence on their surroundings [5, 757–759]. On the other hand, the term “collocations” can be used in phraseology for lexically determined word combinations [14, 103]. The phraseological notion of collocations and the general empirical notion have got some similarities as well as some substantial differences.

Moreover, the term “collocations” functions in computational linguistics where it is defined as any lexicalized word combination having special syntactic and semantic qualities, which is given specific treatment in some natural language processing system [4, 139]. Popular textbooks [8, 184] define collocations in terms of non-compositionality, non-modifiability and non-substitutability.

Methods

The material of the research is the corpora of newspaper articles from the The Guardian broadsheet and the The Mirror tabloid containing collocations with the lemma *HEALTH*. The analysis comprises four stages.

First, the two corpora are created. Second, modifiers and verbs used with the lemma *HEALTH* are established in each corpus. Third, comparative analysis of the data from two corpora

is done. The last stage is interpreting the results of the comparative analysis. While analyzing the lemma *HEALTH* in the media discourse, a complex of research methods is used, *discourse critical analysis* [1, 275–278]. It is a qualitative analytical approach used for describing, interpreting, and explaining the ways in which discourse can construct and maintain social inequalities. In this research *discourse critical analysis* is used to study the lemma *HEALTH* in British broadsheets and tabloids.

The next method is *corpus analysis* [6, 249]. It can be used to analyze corpora of the newspaper articles taken from The Guardian and The Mirror focusing on the collocations with *HEALTH*.

The next method used is *contextual-interpretative analysis*. This method is involved to define the context the lemma *HEALTH* is used in and to interpret the influence of modifiers and verbs on the lemma *HEALTH* [2, 119]. This method helps to explain what influence the lemma *HEALTH* has on the reader of both newspapers.

So this research involves a set of methods to study the lemma *HEALTH* presented in both newspapers and to do comparative research establishing the peculiarities of the lemma studied.

Results and Discussion

At the beginning of the research two corpora were created with the help of *Sketch Engine* program. The corpora present the data received from the newspaper articles taken from the broadsheet *The Guardian* and the tabloid *The Mirror*. The focus of attention is the collocations with the lemma *HEALTH* found in the newspapers mentioned.

The data given by the *Sketch Engine* present various information, showing similarities and differences between these certain collocations found in newspaper discourse. The following n-grams demonstrate collocations in both newspapers.



Figure 1. The n-gram of collocations with *HEALTH* in *The Guardian*



Figure 2. The n-gram of collocations with HEALTH in The Mirror

There are four groups of the collocations:

- ✓ nouns modified by HEALTH, for example *health service*;
- ✓ modifiers of HEALTH, for example *good health*;
- ✓ verbs with HEALTH as subject, for example *health is*;
- ✓ verbs with HEALTH as object, for example *improve health*.

It should be mentioned that both n-grams show a great variety of lemmas in each sector, the number of them being different anyway. *The Guardian* n-gram shows that the largest sector is represented by the nouns modified by HEALTH collocations. The second biggest segment of the pie-chart is presented by the modifiers of HEALTH. Next in number is the sector with verbs used with HEALTH as a subject. The smallest one comprises verbs with HEALTH as an object.

The *Mirror* n-gram shows that the biggest sector is modifiers with HEALTH. Then go nouns modified by HEALTH. Position number three is verbs with HEALTH as an object and the smallest sector is verbs with HEALTH as a subject.

The comparison of both n-grams demonstrates that the number of collocations in each specific

groups in both newspapers is almost the same. There are 42 nouns modified by HEALTH in *The Guardian* and in *The Mirror*, modifiers of HEALTH — 14 and 15 respectively, verbs with HEALTH as subject — 5 in each newspaper, while verbs with HEALTH as object — 5 in *The Guardian* and 10 in *The Mirror*. Whereas, each newspaper discourse has its own amount of certain collocation components used with the lemma HEALTH, in this way revealing their different approaches to problems concerning HEALTH in the society.

The next stage of the research is connected with analyzing the verbs and modifiers used with the lemma HEALTH. M. Stubbs distinguishes three categories of discourse prosody: positive, negative, and neutral, underlying that “there are always semantic relations between node and collocates, and among the collocates themselves [12, 225]. The interpretation of the results obtained has proved that images created in both newspapers concerning health are either positive or negative. Thus, the verbs and modifiers have been divided into *positive and negative ones*. The positive or negative semantic prosody has been based on the definitions of these particular verbs or modifiers. It should be mentioned that these positive and negative collocation components make the collocations with the lemma HEALTH positive and negative correspondingly.

The table given below demonstrates the number of positive and negative verbs and modifiers used with HEALTH in *The Guardian*.

All of things we want to do to protect us against climate change will improve our health, whether it's active transport — walking, cycling — eating healthier sustainable local diets or cutting air pollution (*The Guardian*).

Positive verb *improve* used with the lemma HEALTH in this sentence shows the society's concern being into healthy way of living, which is reflected in the newspaper discourse.

Table 1

Positive / negative collocates with HEALTH in The Guardian

The Guardian			
<i>Positive verbs</i>	<i>Positive modifiers</i>	<i>Negative verbs</i>	<i>Negative modifiers</i>
have — 6	service — 112	deteriorate — 5	problem — 39
improve — 5	secretary — 79	affect — 4	condition — 23
be — 5	minister — 43	underlie — 2	issue — 23
protect — 5	public — 33		inequality — 7
put — 4	professional — 32		ill — 4
mean — 3	team — 11		poor — 3
own — 3	physical — 7		
	good — 4		
	UK — 1		

Six small changes you can make right now to improve your health and wellbeing (The Mirror).

Positive verb *improve* used with the lemma *HEALTH* in this tabloid shows the media are concerned about the society, trying to prove that even small changes in people's lives can lead to general progress and real wellbeing. The newspaper hints at the fact that changing your life is quite easy, which is supportive as well.

If comparing these two sentences with the positive verb *improve* used with the lemma *HEALTH*, taken from two different newspapers, we can see that The Guardian does not only mention that improving health is important; but also shows the way how to achieve the aim. *The Mirror* expresses the same idea, suggesting rather easy ways to be fitter.

The negative verb *deteriorate* used with the lemma *HEALTH*, uncovers worries about people's health.

Sources have also confirmed Yang's health has deteriorated over the past month (The Guardian).

Negative verb *deteriorate* used with the lemma *HEALTH* turns the news presented into the negative one. It shows the newspaper concern about what is going on.

Negative verb *deteriorate* used with the lemma *HEALTH* makes the negative connotation possible in the following example.

The eighties' icon's health has been deteriorating since 2009 when he injured a vertebrae in his upper neck while performing (The Mirror).

In this sentence pessimism and sadness are conveyed as well. *The Mirror*, as different from *The Guardian*, except for expressing their concern, immediately shows the reason of this deterioration giving as much information as possible, feeling freer in presenting the news.

The number of positive and negatives verbs and modifiers with *HEALTH* in *The Mirror* is given below.

For example, when positive modifier *professional* is used with the lemma *HEALTH*, the picture of security and protection is established. However, when negative *problem* is used with the lemma *HEALTH*, the picture of nervousness and scaredness is constructed.

Comparing positive and negative verbs in both newspapers helps us see the differences as well as similarities. Thus, both newspapers use the verbs *improve* and *protect* in collocations with the lemma *HEALTH*, which shows that both newspapers care about the reader, writing about how to improve their health and how they can protect themselves from any harm. On the other hand, the verbs used in *The Mirror* have stronger positive connotation than the verbs offered by *The Guardian*, thus aiming at boosting, promoting and supporting the reader. Besides, it should be mentioned that the set of the verbs chosen by each newspaper is determined by the fact that *The Guardian* is a broadsheet whereas *The Mirror* is a tabloid, both newspapers having their own style of communicating with the reader.

Ban entire pesticide class to protect children's health, experts say (The Guardian).

The verb *protect* in the collocation with the lemma *HEALTH* implies that the experts are attracting attention to children, their health in particular. They hope to make the society be concerned about children's lives and their future.

Cranberries are new weapon in battle to boost health of your brain, gut, heart and immune system (The Mirror).

The verb *boost* in the collocation with the lemma *HEALTH* used in *The Mirror* implies emotional involvement, meaning encouragement, positive attitude, real care and sincere excitement.

Such study is meant to promote health and science, not the opposite, by 'training' unqualified people in subjects that could easily lead to endangering a vulnerable public (The Mirror).

Table 2

Positive / negative collocates with *HEALTH* in *The Mirror*

The Mirror			
Positive verbs	Positive modifiers	Negative verbs	Negative modifiers
improve — 31	professional — 53	affect — 2	problem — 43
protect — 12	physical — 41		condition — 26
support — 4	expert — 32		concern — 21
promote — 3	benefit — 22		issue — 14
boost — 1			
benefit — 1			
maintain — 1			
find — 1			

The verb *promote* in the collocation with the lemma *HEALTH* used in *The Mirror* shows that the society truly cares about all people, implying that everybody matters. People are not the only ones who need promotion, health needs promotion as well. This collocation adds excitement to the message as well as motivation.

As for the negative verbs, one verb common for both newspapers was identified. This is the verb *affect*, which means that the newspapers in spite of the differences in communication style show the reader what can harmfully influence their lives. This verb is the only one used in *The Mirror* whereas *The Guardian* frequently uses two more verbs, such as *underlie* and *deteriorate*, which means that the register of communication is more serious and the reader is offered more serious information, can be even made worried and scared.

How does it affect your health? (The Mirror).

The three housing problems that most affect your health (The Guardian).

The verb *affect* in the collocations with the lemma *HEALTH* in both newspapers implies genuine concern and worry, meaning to demonstrate that society is filled with fear, anxious and is even terrified. The purpose of the media is to attract attention to the problem that is real.

The next stage of the study is comparing the collocation components used with the lemma *HEALTH* in both newspapers. In each group there are verbs and modifiers that are used in *The Guardian* as well as in *The Mirror*. Both newspapers communicate with their readers using such nouns in collocations as *problem*, *issue*, *professional*, *condition*, *service*, *visitor*, accentuating what the reader is really concerned about, either the reader reading a broadsheet, or a tabloid. Besides, the number of the nouns used in the collocations with the lemma *HEALTH* (1,910) in *The Mirror* is bigger than the one in *The Guardian* (1,855).

Noun modified by *HEALTH*:

A health visitor is a qualified nurse or midwife, who has undergone further training (The Mirror).

The health visitor arranged for her to get the appropriate help (The Guardian).

When the noun in these examples is modified by *HEALTH*, the collocation implies hope and optimism, showing that the way out the society needs is absolutely possible.

HEALTH as object:

A switch to ecological farming will benefit health and environment (The Guardian).

Miriam Stoppard. Teenage exercise can benefit your health later in your life (The Mirror).

The verb *benefit* + *HEALTH* as an object show the excitement about the possibility to have a better future, the media revealing encouragement and hope as for good chances the society might have.

HEALTH as a subject:

Our health is all we have (The Guardian).

Physical health is high on agenda - but you might be neglecting one obvious part (The Mirror).

The lemma *HEALTH* used as a subject with the verb *is* accentuates the importance of health for the society, its primary position in the modern world.

Modifiers of *HEALTH*:

Detained British-Australian academic Dr Kylie Moore-Gilbert is in good health and has access to food (The Guardian).

Link between cycling and good health has been clear for years (The Mirror).

The lemma *HEALTH* used with the modifier *good* in both newspapers demonstrates that the society is truly concerned about health, wishing it to be always good only.

As for the verbs with the lemma *HEALTH* as an object, the ones in common are *protect*, *affect* and *improve*, which shows that the two newspapers, though different, appeal to the reader in the same way, showing what can affect them and, what is really important. The other verbs used in *The Mirror* add optimistic coloring whereas the ones in *The Guardian* are of explanatory kind with no positive emotions.

It should be also mentioned that most of the modifiers in the collocations with *HEALTH* are the same in both newspapers, e.g. *good*, *own*, *physical*. The modifiers mentioned help two different kinds of newspapers touch the reader to have the greatest impact.

Conclusions

The research revealed similarities and differences in using the collocations with *HEALTH* in *The Guardian* and *The Mirror* newspapers.

The four groups of collocations (nouns modified by *HEALTH*, modifiers of *HEALTH*, verbs with *HEALTH* as subject and verbs with *HEALTH* as object) are represented by different amounts of collocations in each newspaper, either same (*improve*, *protect*, *own*, *good*) or different (*underlie*, *deteriorate* — *The Guardian*), (*benefit*, *boost*, *promote* — *The Mirror*).

The verbs and modifiers used with the lemma *HEALTH* can be *positive* (*protect*, *expert*, *benefit*) and *negative* (*problem poor*, *issue*).

In *The Guardian* there are more positive modifiers than positive verbs, more negative modifiers than negative verbs. In *The Mirror* there are more positive verbs than positive modifiers, but more negative modifiers than negative verbs.

There are positive verbs that are used in collocations by both newspapers (*improve* and *protect*) and negative verbs (*affect*) whereas some verbs are characteristic only of *The Guardian* (positive: *put*, *have*, *mean*; negative: *underline*, *deteriorate*) or

The Mirror (positive: *boost, benefit, promote*; no special negative verb found).

Thus, the collocations with the lemma *HEALTH* used in *The Guardian* describe situations or facts, no imperative connotation is expressed. On the other hand, the collocations with the lemma *HEALTH* used in *The Mirror* are not only descriptive, but imperative,

too, encouraging the reader to do something or warning the reader not to.

The further research will be aimed at investigating the most frequently used collocations with *HEALTH* considering their implications and analyzing the differences and similarities conveyed in broadsheets and tabloids.

REFERENCES

1. Baker, P., Costas, G., & McEnery, T. (2013). Sketching Health: a Corpus Driven Analysis of Representations Around the Word 'health' in the British Press 1998–2009. *Applied Linguistics* 34(3), 255–278. Retrieved from <https://www.amazon.co.uk/Discourse-Analysis-Media-Attitudes-Representation/dp/11070088242>
2. Bednarek, M. (2008). Semantic Preference and Semantic Prosody Re-examined. *Corpus Linguistics and Linguistic Theory*, 4.2, 119–139. Retrieved from <https://www.semanticscholar.org/paper/Semantic-preference-and-semantic-prosody-Bednarek/77989bf41f20f4290f7cae0bf4b18602df115e63>
3. Blum-Kulka, Sh., & Hamo, M. (2011). Discourse Pragmatics. In Teun A. van Dijk (Ed.) *Discourse Studies: a Multidisciplinary Introduction*, 143–164. <https://www.edisoportal.org/attachments/article/274/Blum-Kulka%20Shoshana.pdf>
4. Brezina, V., McEnery, T., & Wattam, S. (2015). Collocations in Context: a New Perspective on Collocation Networks. *International Journal of Corpus Linguistics* 20(2), 139–173. <https://benjamins.com/catalog/ijcl.20.2.01bre>
5. Firth, A., & Wagner, J. (2007). On Discourse, Communication, and (some) Fundamental Concepts in SLA Research. *The Modern Language Journal*, 91(s1), 757–772. https://www.researchgate.net/publication/227684972_On_Discourse_Communication_and_Some_Fundamental_Concepts_in_SLA_Research
6. Hunston, S. (2007). Semantic Prosody Revisited. *International Journal of Corpus Linguistics*, 122(2), 249–268. https://www.researchgate.net/publication/228929277_Semantic_prosody_revisited
7. Kübler, S., & Zinsmeister, H. (2015). *Corpus Linguistics and Linguistically Annotated Corpora*. London: Bloomsbury Academic. Retrieved from <https://www.twirpx.com/file/1583647/>
8. Manning, Ch., & Schütze, H. (1999). *Foundations of Statistical Natural Language Processing*, Cambridge: MIT Press. Retrieved from <https://nlp.stanford.edu/fsnlp/>
9. McEnery, T., & Hardie, A. (2011). *Corpus Linguistics: Method, Theory and Practice*. Cambridge: Cambridge University Press. <https://www.cambridge.org/ua/academic/subjects/languages-linguistics/applied-linguistics-and-second-language-acquisition/corpus-linguistics-method-theory-and-practice?format=PB&isbn=9780521547369>
10. McEnery, T., & Wilson, A. (2011). *Corpus Linguistics: An Introduction*. Edinburgh: Edinburgh University Press. Retrieved from <https://www.twirpx.com/file/893556/>
11. Stubbs, M. (2001). Text, corpora and problems of interpretation: a response to Widdowson. *Applied Linguistics*, 22, 149–172. Retrieved from <https://www.semanticscholar.org/paper/Texts%2C-corpora%2C-and-problems-of-interpretation%3A-A-Stubbs/eeab3b8a1c9dbc35f2ecc24e8d6dc1d7879de02>
12. Stubbs, M. (2002). Two quantitative methods of studying phraseology in English. *International Journal of Corpus Linguistics*, 7(2), 215–244. Retrieved from http://www.cantab.net/users/michael.stubbs/articles/stubbs-2002-two_quantitative_methods.pdf
13. Wilson, J. M., Marin, P. J., Rhea, M. R., Wilson, S. M., Loenneke, J. P., & Anderson, J. C. (2012). Concurrent training: a meta-analysis examining interference of aerobic and resistance exercises. *The Journal of Strength & Conditioning Research*, 26(8). Retrieved from <https://scholar.google.com/citations?user=tnhk9iMAAAAJ&hl=en>
14. Xiao, R. Z. & McEnery, A. M. (2006). Collocation, semantic prosody and near synonymy: a cross-linguistic perspective. *Applied Linguistics*, 27(1), 103–129. https://www.academia.edu/2986741/Xiao_R_and_McEnery_T_2006_Collocation_semantic_prosody_and_near_synonymy_A_cross_linguistic_perspective
15. *The Guardian*. Great Britain. Retrieved from <https://www.theguardian.com/society/health> (access date: 15.01.21)
16. *The Mirror*. Great Britain. Retrieved from <https://www.mirror.co.uk/lifestyle/health/> (access date: 21.01.21)

Дата надходження статті до редакції: 04.03.2021 р.

Прийнято до друку: 20.04.2021 р.