

---

# Using the Google Books Ngram Corpus to Study Social Evolution

---

**Valery Solovyev**

*Kazan Federal University, Russia*

## ABSTRACT

*This article briefly summarizes primary publications that use Google Books Ngram (GBN) to study societal change. GBN is the most extensive tagged diachronic corpus available. Trends in societal evolution can be studied using year-by-year word frequency statistics. The development of individualism, changes in emotions and happiness, social psychology, and some other topics are among those examined in this article as research areas that have attracted the most interest. This paper discusses the specific findings and the research methodology, particularly its limitations. There are some examples of how GBN can be used to test existing scientific theories. New, unexpected, and scientifically significant findings are possible with GBN that would not be possible with other approaches.*

**Keywords:** *social evolution, language changes, digital humanities, corpora, Google Books Ngram.*

## INTRODUCTION

Extra-large diachronic text corpora have recently become a popular research tool, offering new information on the development of society and social change. This discussion focuses primarily on Google Books, a fundamental resource created by Google. It was created by continuously scanning books from the world's largest libraries and contains 30 million digitized books. Based on this, Google Books Ngram (<https://books.google.com/ngrams/>), a publicly-accessible corpus of n-long word strings ( $n < 6$ ) in texts, was developed in 2009. It

Recommended citation: Solovyev, V. Using the Google Books Ngram Corpus to Study Social Evolution. *Social Evolution & History*, Vol. 23 No. 2, September 2024, pp. 144–164. DOI: 10.30884/seh/2024.02.06.

© 2024 'Uchitel' Publishing House

includes subcorpora for English (separately British and American dialects and a fiction corpus), German, French, Russian, Spanish, Italian, Hebrew, and Chinese. The corpus is constantly growing, and there are currently three versions: 2009, 2012, and 2019. Data quality and quantity differ between the versions, and the character recognition system and metadata have been improved. The parameters of the third version are as follows. The English subcorpus has more than 2,000 billion words (or tokens), while the Russian subcorpus has 90 billion, which is 60 times more than the Russian National Corpus. The corpus is diachronic and covers the years 1470 to 2019. The dependency grammar in the corpus allows for part-of-speech tagging and syntactic tagging. It is the world's largest tagged diachronic corpus; hence, it qualifies as both Big Data and Long Data. Since the corpus is free to download, researchers can perform any data transformations. There is a useful feature that shows a clear graphical representation of n-gram frequencies for each year.

Thousands of studies have used the corpus in its almost 14 years of existence. The Google Academy system has more than 5,000 publications that mention GBN. GBN was first described in (Michel *et al.* 2011), which has since been referenced more than 2,500 times and is considered a seminal work in the field. It postulates that 'the most robust historical trends are associated with frequent n-grams.' The second version of GBN is described in detail in (Lin *et al.* 2012). The main difference between the first and second versions is the addition of syntactic tags. A Russian translation of the monograph is also available (Aiden and Michel 2016). The Culturomics Lab (<http://www.culturomics.org/>) was founded at Harvard with the mission of 'the quantitative study of human culture across societies and across centuries.' Book texts capture the social climate at a given time, and textual analysis provides insights into social dynamics that are not available through other methods, such as archeological research, sociological surveys, or psychological tests. This article aims to provide a concise overview of GBN-based sociological research.

### **THE DIGITAL FOOTPRINT OF SOCIAL CHANGE**

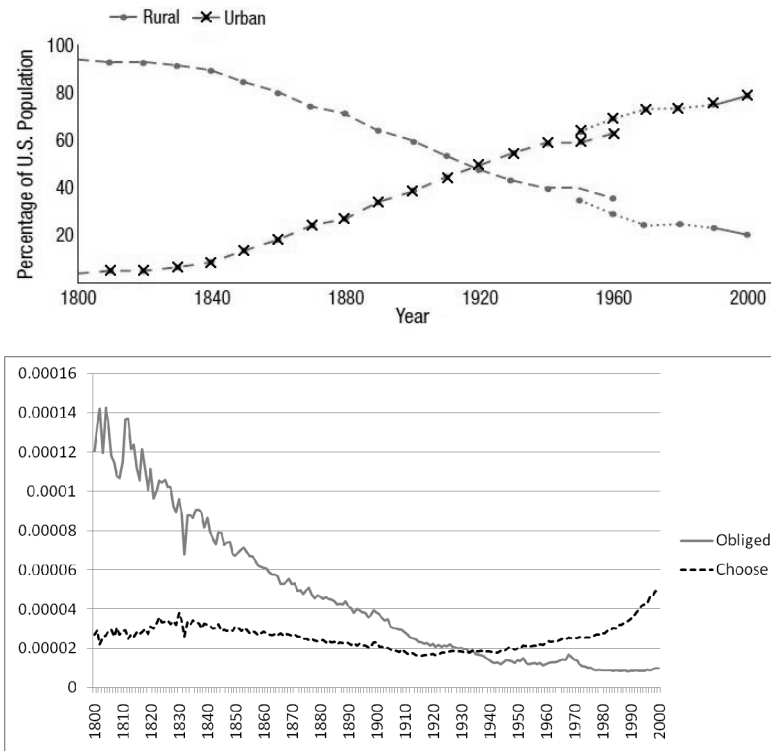
GBN research has been carried out in many different areas, but the most important ones are the following: the development of individualism; cultural change; gender studies; the dynamics of emotion and the sense of happiness; the economy and the functional systems of society; and social psychology. However, the boundaries drawn here are

largely arbitrary. All the research is interdisciplinary and uses computer technologies to combine sociology, psychology, and linguistics. Let us describe the main results obtained.

### **Collectivism vs Individualism**

This is perhaps the area of research that has received the most attention. Greenfield's (2009) Theory of Social Change and Human Development discusses the shift from collectivism to individualism in psychology due to socio-demographic factors like urbanization and the spread of education. Individualism, characterized by qualities like independence, personal preference, personal growth, and materialism, is associated with the urban way of life. Collectivist characteristics such as the connection of community members (*Gemeinschaft*), mutual responsibilities, reverence for elders, and religiosity are typical of the rural (patriarchal) lifestyles. GBN immediately caught the author's attention and she used it to test her theory in (Greenfield 2013). The following describes the general research methodology used in this and more recent work. Sociological and lexical parameters are compared in terms of how they change over time. Here is an example from the publication mentioned above. The sociological indicator is the proportion of urban residents (with urban communities considered to have a population of more than 2,500). The lexical parameters used are words indicating collectivism and individualism like 'obliged' and 'choose'. The decline of the rural population and the use of the lexeme 'obliged' are clearly related.

In addition to notional lexemes, changes in the use of personal pronouns were also taken into account. Individualism correlates with an increased frequency of first-person singular pronouns and a decreased frequency of first-person plural pronouns. For instance, between 1960 and 2008, the frequency of *I, me* pronouns increased by 42 per cent, while the frequency of *we, us* pronouns decreased by 10 per cent in American English (Twenge, Campbell, and Gentile 2013). Some papers discuss similar studies for other GBN-supported languages (Yu *et al.* 2016; Uz 2014; Hamamura and Xu 2015).



**Fig. 1. Changes in the US urban population and the frequency of the lexemes 'obliged' and 'choose'**

Source: Greenfield (2013), Figures 1 and 2 (adapted for black and white printing)

Later, similar studies with more words and sociological parameters (number of divorces, availability of higher education, *etc.*) were conducted for Chinese (Zeng and Greenfield 2015), Japanese (Ogihara 2017), Russian (Velichkovsky *et al.* 2019; Skrebyte, Garnett, and Kendal 2016), German (Younes and Reips 2018), and some other languages/countries. The paper (Velichkovsky *et al.* 2019) discusses the following words in Russian: freedom/duty, compete/distribute, acquire/share, private/public, talent/effort, uniqueness/promptness, personality/relationship. The general pattern shows that China, Japan, and Russia have been slower to embrace individualism, and that these countries are also better at preserving traditional values. China's unique features attract more interest, and the study of the development

of individualism in China continues, with various aspects of this process being clarified (Li 2022; Ogihara 2023). It has been found that Germany experienced the following phenomenon: despite the widespread rise of individualism during World War II, an opposite trend was observed in Germany. This line of research has also yielded very surprising results, such as the link between collectivism and corruption (Li *et al.* 2019).

### **Culture**

Researchers' attention has been drawn to the drastic changes that China has undergone in recent years, making China a model object for the study of socio-cultural change (in the broad sense). Xu and Hamamura (2014) investigated how the Chinese population perceives the changes in the country. GBN data were compared with sociological survey data to show the importance of such concepts as materialism, freedom, democracy and human rights, family relations, and friendship in the worldview of Chinese residents. The paper (Zhang, and Weng 2017) analyzed data from the Chinese GBN subcorpus since 1980 and drew attention to the preservation of traditional cultural values in China despite ongoing global changes. The paper (Wang, Liu, and Huang 2022) analyzed the use of Chinese modal verbs from 1901 to 2009. It was previously found that the use of modal verbs in English declined between 1961 and 1992, which correlates with a decline in authoritarianism in society. The situation is more complicated in the Chinese language. Over the century, the frequency of modal verbs has both increased and decreased due to the combined effects of several factors, one of which was the completion and finalization of the simplification of the Chinese writing system.

The paper (Younes, and Reips 2019b) studied the changes in the frequency of religious terms. It was found that the frequency of religious terms increased in years of crisis, such as World War II.

Liu (2016) introduced a new concept – cultural complexity – expressed numerically by Shannon entropy, and showed that the complexity of American culture has steadily increased throughout the twentieth century. The complexity of British and Chinese culture has also generally increased, but less steadily. Further exploration of this concept has been done in (Zhu and Lei 2018).

### **Emotionality and Happiness**

According to (Morin and Acerbi 2017), there has been a decline in the use of emotive language in English-language fiction over the past 200 years. Furthermore, the first words to be used less frequently are

those that express positive emotions. This is shown to depend on demographic shifts, changes in language vocabulary, and the popularity of different genres.

This phenomenon is supported in (Brand, Acerbi, and Mesoudi 2019) for lyrical song texts and in (Wodarz and Harris 2022) for the Corpus of Historical American English materials. Valery Solovyev with co-authors (Solovyev, Bochkarev, and Bayrasheva 2016) published the results for the six major European languages (represented in GBN) somewhat earlier. However, this rule does not apply everywhere. The situation in China is quite different. Since the end of the so-called Cultural Revolution in the late 1980s, the frequency of emotion use has increased, albeit not significantly, with an increase in the frequency of positive emotions in the first place (Younes and Reips 2019b). This is supported by (Peng and Luo 2022), who show that, since 2006, the frequency of positive emotive words has increased significantly, whereas the frequency of negative emotive words has remained essentially stable.

Rumen Iliev *et al.* (2016) found similar results for all words in the language (English), not just the emotive ones. This study investigated the phenomenon of linguistic positivity bias, or the propensity of people to use words with positive rather than negative connotations. A diachronic study showed that the magnitude of this effect diminishes over time, with the trend well approximated by a linear law. The study also shows a relationship between linguistic positivity bias and national happiness levels.

Shigehiro Oishi *et al.* (2013) examine how the concept of happiness has changed over time. After reviewing various sources, the author draws two main lines of research. 1) First, in the 1800s, the phrase *happy nation* was used 14 times more often than *happy person in the American English subcorpus*. The situation changed in the 1920s when the phrase *happy person* became more common. In 2008, the phrase *happy person* was used 22 times more often than *happy nation*. In 2019, the same ratio was reported by the third version of the GBN. The phrase *happy nation* triumphed during a period of low income and low migration rates, while the *happy person* triumphed during a period of high income and high migration rates. Second, the conception of happiness as good fortune, prevalent in American culture until the 1920s, was replaced by the notion of happiness as a psychological state of well-being.

The paper (Qiu, Lu, and Chiu 2014) investigates a different aspect of happiness. Curious about how eudaemonism and hedonism interact, the authors have shown that over two centuries, the frequency of hedonism-related keywords in English has decreased, while that of eu-

daemonism has increased. The authors have established a link between this phenomenon and rising living standards. People have no time for eudemonism when their standard of living is low (as it was in the nineteenth century).

Changes in well-being in different languages and countries were taken into account. The paper (Hills, Proto, and Sgroi 2016) examines how different events, such as economic depression, affect the level of well-being in six different countries. By examining the corresponding GBN words, Han-Wu-Shuang Bao, Huajian Cai, and Zihang Huang (2022) demonstrated an improvement in China's living standards. These results were also obtained using other methods (sociological surveys [Su and Liu 2020]), thus validating the GBN results.

### **Gender Studies**

Gender studies are one area of research that actively employs GBN. The paper (Twenge, Campbell, and Gentile 2012) examines the shift in women's status in the United States using the ratio of masculine to feminine pronouns. In the first half of the twentieth century, masculine pronouns were used 3.5 times more often than their feminine counterparts, and this ratio increased to 4.5 in the 1960s. However, with the rise of feminism after 1968, the use of feminine pronouns increased dramatically, and masculine pronouns are now only used twice as often. This paper highlights the correlation between this indicator and social measures of women's educational attainment, labor force participation, *etc.*

The paper (Ye *et al.* 2018) investigates the evolution of adjectives describing men and women. The Big Five model is employed, which includes the following characteristics: agreeableness, extraversion, conscientiousness, neuroticism/emotional stability, and openness to experience/intellect. Several dozen adjectives associated with each parameter were selected. The following major findings were obtained: 1) adjectives corresponding to the parameter agreeableness are the most commonly used; 2) positive characteristics are used more frequently than negative ones; 3) all parameters except openness are used more frequently to characterize men; and 4) gender differences decrease over time.

Kesebir (2017) looks at the order in which the words *men*, *women* are used in conjunctive structures. The male-first *men and women* is used more frequently, indicating a greater relevance to the context of men and representing the expression and support of pre-existing gender stereotypes.

Marco Del Giudice (2017) investigates how blue came to be associated with boys and pink with girls. The author engages in a debate with a previous publication (Paoletti 2012), which claimed that this happened only after 1940. Del Giudice claims that blue was previously associated with boys and provides GBN data to support his claim.

### **Functional Systems of Society**

Politics, economy, culture, religion, education, and other societal structures, referred to as functional systems of society, were all examined by Steffen Roth in a series of articles (Roth 2013, 2014; Roth *et al.* 2017a, 2017b). These articles focused on the interconnections and functions of these different structures. There are two opposing viewpoints on how they are connected. One suggests that all these functional systems are fundamentally independent. However, according to the economization of society hypothesis, the economy plays a dominant role, 'increasing influence of economic factors and values on the political agenda and other areas of society' (Blumler and Kavanagh 1999: 210). Roth compares the frequency of words corresponding to functional systems and concludes that the hypothesis of societal economization is irrelevant. The research is done in-depth using a large amount of data from six different languages. The author attaches high value on the results, believing that if they are true, 'this will indeed change the face of modern society' (Roth 2014: 52). This series additionally examines the development trends of all the functional systems highlighted.

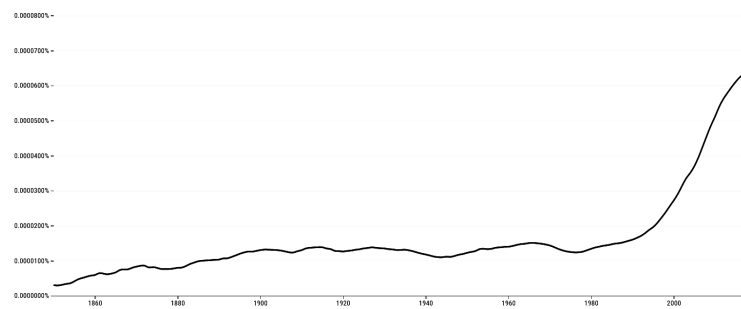
Carlton Clark, Lei Zhang, and Steffen Roth (2022) conducted a similar study on Chinese. Notable differences were found between the Chinese and European subcorpus data. Compared to books written in European languages, Chinese books focus more on science and economics than on politics.

A similar study in (Bochkarev, Shevlyakova, and Solovyev 2015) demonstrates that the frequency of words like *development*, *education*, and *production*, which reflect the key ideas of social development, increased in the English language for more than a century and a half. Then, surprisingly, their frequency began to fall dramatically in the 1980s and has not recovered yet. Nowadays, the word *education* is used less frequently than it was a century ago. This apparently reflects a change in society's priorities, but in-depth, multidimensional research is needed to pinpoint new priorities and trends in societal development.

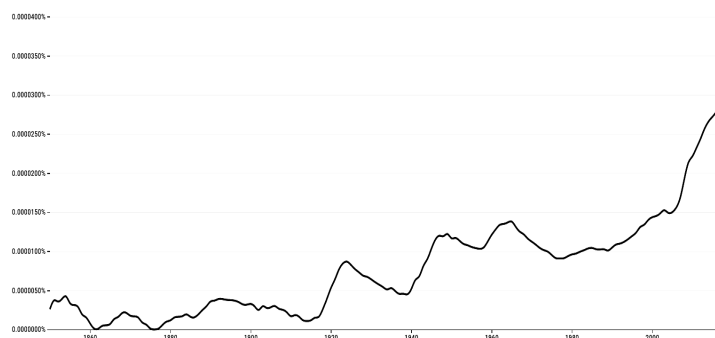


### Social Psychology

A well-established cultural narrative in the United States is that the Millennial generation values purpose in life more than previous generations (Sheahan 2005). Grant (2017) puts this proposition to the test using GBN data. Since the 1990s, the frequency of the phrase ‘purpose in life’ has risen sharply, as if to confirm the popular belief, as shown in Figure 2. However, this paper acknowledges that the situation is different for other languages/cultures. While the frequency of the corresponding phrase has increased in French and Spanish over this period, it has not increased to the same extent as in American English. It is fascinating to see how close the Russian and American frequency charts are for the phrase ‘purpose in life’ (Figure 3). Starting from the same level, 0.00001 % (the proportion of this phrase in the whole corpus), the growth in both languages begins almost simultaneously since the late 1970s. Despite the differences in the political system, culture, and economic relations, it appears that our countries share a great deal in common.



**Fig. 2.** Frequency of the Phrase *purpose in life* in American English



**Fig. 3.** Frequency of the Phrase *purpose in life* in Russian

The words used in different semantic fields explain differences in psychology. William H. Zywiak and Gao Niu (2021) investigate the frequency changes of key character traits identified in (Peterson and Seligman 2004): Love, Hope, Perspective, and Leadership. The paper shows that the frequency of Love, Perspective, and Leadership has been increasing in the United States since 1920, and the frequency of Hope has been increasing since 1987. Alberto Acerbi and Pier Luigi Sacco (2022) examine the other four important concepts – *indulge*, *want*, *restrain*, and *must* – within the tightness-looseness paradigm. The frequencies of the words ‘indulge’ and ‘want’ are shown to be U-shaped, with a minimum in the 1970s. Throughout the twentieth century, the frequency of the words ‘restrain’ and ‘must’ increased.

Attitudes in society towards different social groups (Black, White, Asian, Irish, Hispanic, Native American, male, female, old, young, fat, thin, rich, poor) have been studied over 200 years in (Charlesworth, Çalışkan, and Banaji 2022). Positive and negative attitudes towards a group are shown to be highly stable, while the vocabulary used to describe the group changes over time. The research methodology is what makes this article interesting. In addition to pure frequency analysis, the word embedding method is used, which consists in building a vector of compatibility of the studied word in a large corpus of texts (Mikolov *et al.* 2013), followed by a study of the changes in the built vectors over time within the approach (Hamilton, Leskovec, and Jurafsky 2016).

### **Miscellaneous**

This section contains some papers that are difficult to relate to the topics discussed above.

Shai Ophir (2016) investigates the frequency pattern of the word ‘truth’ and finds that it correlates with the frequency pattern of the word ‘love’. Similarities between the author's findings and those of Peterim Sorokin (1937) are highlighted.

Yunsong Chen and Fei Yan (2018) found a relationship between the amount of foreign investment in a province in China and the frequency with which the GNB mentions that province.

Studies (Dechesne and Bandt-Law 2019) were carried out in the context of terror management theory. Changes in the concept of ‘depth’ were shown to correlate with changes in the concept of ‘fear’, but not ‘anxiety’. An unexpected correlation with romantic relationships was also found.

Based on data on the frequency changes of many words, including emotive ones, (Popescu and Strapparava 2018) attempted to categorize modern history into epochs, characterizing each epoch and highlighting the society's interests during these periods. A strict algorithmic procedure was proposed. Furthermore, it was claimed that this method allows both accounting for the past and predicting the future.

The paper (Zywiak, Bobroff, and Niu 2021) proposed the intriguing idea of studying the Black Swan years in different languages/countries. The years 1799, 1865, 1917, 1945, and 1948 were chosen as significant for the French, American, Russian, German, and Hebrew languages respectively. It is shown that changes in the number of references to these years in the corresponding GBN subcorpora are very different from changes in the number of references to regular years.

According to (Bochkarev, Solovyev, and Wichmann 2014), the American and British dialects of English diverged (at the vocabulary level) during the nineteenth and first half of the twentieth centuries and then began to converge. The impact of globalization, which has been felt since the 1950s as a result of the growth of the media (transnational media corporations, widespread television), tourism (the introduction of jet air travel, the tourist industry), and trade (the signing of the General Agreement on Tariffs and Trade (GATT)), is a natural explanation for this.

#### **Problems, Limitations, and Recommendations for GBN Use**

Some scholars have criticized the use of GBN and highlighted the limitations of this approach.

Several papers (Pechenick *et al.* 2015; Belikov 2016; Koplénig 2017) have drawn attention to the presence of errors in GBN, such as spelling errors due to inaccurate character recognition and metadata errors. The balance of the corpus has also been questioned. It should be noted that Google promptly addressed these criticisms and significantly improved the quality of corpus in the second and third versions. (Solovyev, Bochkarev, and Akhtyamova 2020) provide a comprehensive analysis of corpus errors and their impact on statistical analysis results. The paper also deals with GBN balance. The concept of diachronic balance is introduced, which refers to the overall corpus balance and the balance of each year with arguments presented for both synchronic and diachronic GBN balances.

Zięba (2018) found that there is sometimes a discrepancy between the frequency of use of a word and the corresponding phenomenon in

everyday life. For instance, both the frequency of the word *family* and the number of divorces increase. An increase in the frequency of the word *family* only indicates that more people talk about it, not that the number of families has increased. The paper also covers the mediatisation of society – changes in communication brought about by the growth of mass media.

According to (Madsen and Slåtten 2022), while books have been the primary means of preserving and disseminating knowledge for centuries, the Internet has recently emerged as a powerful resource. This may mean that different data sets will be needed to keep up with the latest changes in society.

To obtain more reliable data from corpora, the paper (Younes 2019a) suggests using (1) different language corpora, (2) cross-checking on different corpora from the same language, (3) word inflections, and (4) synonyms.

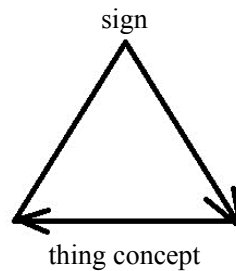
Yanlei Ge (2022) says that statistical analysis methods like the Pearson' correlation coefficient should be used to process data correctly.

## DISCUSSION

In this section we will summarize and discuss the types of social trends and the language changes they cause. To ensure effective communication in changing living conditions, language must inevitably change. Probably, almost all noticeable changes in society, not to mention global macro trends, have an impact on language. The article provides a number of examples of social trends that affect language use. These include urbanization, globalization, changes in the political system of society, higher living standards, changing status of social groups, new priorities in life and so on.

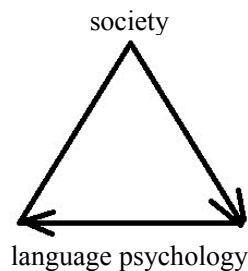
Let us specify what kind of language changes are analyzed in this article. Technological progress leads to the emergence of new facts that require new words to describe them (car, airplane) or changes in the meaning of existing words (mouse). As a result of globalization, cultures are mutually enriched, supported by borrowing words from other languages (*sushi, pizza, tequila*). These language changes are trivial, they are obvious, but this article is not about them. Social dynamics also cause completely non-obvious shifts in the frequency of use of the most commonly used words of the language. These changes can only be traced using computational linguistic methods based on large text corpora.

The external conditions of life usually affect language not directly, but through changes in psychology. Similar to the well-known Frege triangle in philosophy:



**Fig. 4. The classic Frege triangle**

we can propose another triangle that visualizes the connection between society, psychology and language at a higher level than the level of individual words:



**Fig. 5. Triangle Society–Psychology–Language**

For example, as mentioned above, changes in society (urbanization) lead to changes in psychology (individualism), which are reflected in the language (the frequency of the use of the pronoun 'I' increases, *etc.*). Thus, the frequency of the words, whose semantics directly correlate with the meanings of the changes, alters: collectivism / individualism  $\Leftrightarrow$  public / own.

When analyzing different cases of language change, three types of such changes can be distinguished. The first type includes almost all the examples described in this paper, when publications confirm changes in the frequency of use of only a small number of words in response to changes in society.

The second type. However, situations in which whole layers of vocabulary change are also possible. This refers to the above-mentio-

ned convergence of the American and British dialects of English under the influence of globalization. Here is another example – the preliminary results of a study currently being conducted by our team and not yet fully completed. An interesting pattern was revealed when replicating the work published 50 years ago to create the famous associative dictionary of the Russian language by Yuri N. Karaulov (Karaulov *et al.* 1994, 1996, 1998). Although many associations remained unchanged, the use of adjectives as associations sharply reduced by a factor of 2.5 (Solovyev and Vol'skaya 2023). More than 600 thousand associations have been analyzed. Such a large number of associations does not allow us to attribute this result to some kind of coincidence, it requires an explanation.

The probable explanation for this phenomenon is as follows. The emergence of new means of communication – SMS, social networks (especially with a limited volume of messages), as well as a general increase in the pace of life, led to the need for compact messages, containing only the most important information. This has probably resulted in the gradual vanishing of adjectives, which are not always essential for understanding the meaning of the message. According to GBN, the proportion of adjectives in the Russian language is really falling. There have been several stages of this in the last decades – since 2000, then since 2008 and especially sharply since 2015. It should be mentioned that the SMS-service for mobile phones appeared in 2000, the Odnoklassniki and VKontakte networks have been gaining popularity since 2006, the Russian version of Facebook (forbidden in Russian Federation) appeared in 2008, and Telegram and Instagram (forbidden in Russian Federation) have become increasingly popular since around 2015. Thus, these technological innovations lead to serious changes in the use of the language as a whole. The change is not superficial, but manifests itself in the corpus of texts and psychological experiments.

Let us analyze the third type of change. So far, we have been talking about trends in the development of society that lead to permanent changes. However, the language can also be influenced by major significant events, such as wars and revolutions, which occupy short periods of time. An unexpected effect was found in the work (Solovyev *et al.* 2011), namely, that during the French Revolution and the subsequent Napoleonic Wars, the use of words describing colors sharply increased. The reason for this phenomenon remains unclear, as there is no direct link between the revolution and the color. It could be explained by the influence of some deep psychological associations. One can also recall the Reds (the Red Guards) and the Whites (the White

Guards) during the Civil War in Russia, the Orange Revolution, the Brown Plague. The connection between social events and language can be hidden.

## CONCLUSION

Texts naturally reflect changes in society. Therefore, analyzing the content of texts from different periods helps us to understand the main priorities and trends of social change. Undoubtedly, this opportunity has always existed for researchers, but gaining access to early records is very challenging. The advances in information technology open up new avenues for the study of social evolution. The focus is on diachronic digital corpora of texts, which are extremely large. Search technologies enable users to quickly find the information they need. The GBN corpus, developed almost 15 years ago, has proven particularly useful because it contains a convenient frequency table for all words and phrases printed in books for the last 500 years. Other diachronic corpora are also used to confirm the results obtained via GBN.

The hypothesis that major historical trends are reflected in the frequency of n-grams was first put forward in the early years and has since been supported by a large number of studies. Roth (Roth *et al.* 2017b) coined the amusing term *socioencephalography* for frequency charts. Researchers now have a fundamentally new tool at their disposal that can be metaphorically compared to other scientific instruments such as the *encephalograph* and the telescope.

So many papers have now been published that it would be impossible to give an exhaustive overview of them all. Inevitably, the selection of papers reviewed here is somewhat subjective. However, we made an effort to focus on the most vital aspects of the field, including the most recent and widely cited publications. In particular, this overview includes all of the research fields cited by Giménez (2018).

We can conclude from the review that GBN can be used to study social changes. According to Michael Pettit, ‘technological changes have made forms of historical research appear more empirical and objective’ (Pettit 2016). Nevertheless, it is important to remember that this is a new tool, and more research is needed to determine how it can be used more effectively.

## ACKNOWLEDGMENT

The research was supported by the Russian Science Foundation (project No. 20-18-00206).

## REFERENCES

- Acerbi, A., and Sacco, P. L. 2022. The Self-Control vs. Self-Indulgence Dilemma: A Culturomic Analysis of 20<sup>th</sup> Century Trends. *Journal of Behavioral and Experimental Economics* 101: 101946.
- Aiden, E., and Michel, J. B. 2016. *Uncharted: Big Data as a Lens on Human Culture*. Moscow: AST. *Original in Russian* (Эйден Э., Мишель Ж.-Б. Неизведанная территория. М.: АСТ).
- Bao, H.-W.-S., Cai, H., and Huang, Z. 2022. Discerning Cultural Shifts in China? Commentary on Hamamura *et al.* *American Psychologist* 77 (6): 786–788. <https://doi.org/10.1037/amp0001013>.
- Belikov, V.I. 2016. What and How can a Linguist Get from Digitized Texts? *Siberian J. Philol.* 3: 17–34. *Original in Russian* (Беликов В. И. Что и как может получить лингвист из оцифрованных текстов. *Сибирский филологический журнал* 3: 17–34).
- Blumler, J. G., and Kavanagh, D. 1999. The Third Age of Political Communication: Influences and Features. *Political Communication* 16 (3): 209–230.
- Bochkarev, V. V., Shevlyakova, A. V., and Solovyev, V. D. 2015. The Average Word Length Dynamics as an Indicator of Cultural Changes in Society. *Social Evolution and History* 14 (2): 153–175.
- Bochkarev, V., Solovyev, V., and Wichmann, S. 2014. Universals versus Historical Contingencies in Lexical Evolution. *Journal of the Royal Society Interface* 11 (101): 20140841.
- Brand, C. O., Acerbi, A., and Mesoudi, A. 2019. Cultural Evolution of Emotional Expression in 50 Years of Song Lyrics. *Evolutionary Human Sciences* 1, e11: 1–14. <https://doi.org/10.1017/ehs.2019.11>.
- Charlesworth, T. E., Caliskan, A., and Banaji, M. R. 2022. Historical Representations of Social Groups across 200 Years of Word Embeddings from Google Books. *Proceedings of the National Academy of Sciences* 119 (28), e2121798119.
- Chen, Y., and Yan, F. 2018. International Visibility as Determinants of Foreign Direct Investment: An Empirical Study of Chinese Provinces. *Social Science Research* 76: 23–29.
- Clark, C., Zhang, L., and Roth, S. 2022. What's Trending in the Chinese Google Books Corpus? In Fiormonte, D., Chaudhuri, S., and Ricaurte, P. (eds.), *Global Debates in the Digital Humanities* (pp. 151–169). University of Minnesota Press.
- Dechesne, M., and Bandt-Law, B. 2019. Terror in Time: Extending Culturomics to Address Basic Terror Management Mechanisms. *Cognition and emotion* 33 (3): 492–511.
- Del Giudice, M. 2017. Pink, Blue, and Gender: An Update. *Archives of Sexual Behavior* 46 (6): 1555–1563.



- Ge, Y. 2022. The Linguocultural Concept Based on Word Frequency: Correlation, Differentiation, and Cross-Cultural Comparison. *Interdisciplinary Science Reviews* 47 (1): 3–17.
- Giménez, A. L. 2018. *The Ngram Viewer Tool and its Use in the Social Sciences: A Review*. URL: [Repositori.uji.es](https://repositori.uji.es).
- Grant, G. B. 2017. Exploring the Possibility of Peak Individualism, Humanity's Existential Crisis, and an Emerging Age of Purpose. *Frontiers in Psychology* 8: 1478.
- Greenfield, P. M. 2009. Linking Social Change and Developmental Change: Shifting Pathways of Human Development. *Developmental Psychology* 45 (2): 401–418.
- Greenfield, P. M. 2013. The Changing Psychology of Culture from 1800 through 2000. *Psychological science* 24 (9): 1722–1731.
- Hamamura, T., and Xu, Y. 2015. Changes in Chinese Culture as Examined through Changes in Personal Pronoun Usage. *Journal of Cross-Cultural Psychology* 46 (7): 930–941.
- Hamilton, W. L., Leskovec, J., and Jurafsky, D. 2016. Diachronic Word Embeddings Reveal Statistical Laws of Semantic Change. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics* 1: 1489–1501.
- Hills, T., Proto, E., and Sgroi, D. 2016. Historical Analysis of National Subjective Wellbeing Using Millions of Digitized Books: Introducing the HPS Index. *The CAGE Background Briefing Series* 6.
- Iliev, R., Hoover, J., Dehghani, M., and Axelrod, R. 2016. Linguistic Positivity in Historical Texts Reflects Dynamic Environmental and Psychological Factors. *Proceedings of the National Academy of Sciences* 113 (49): E7871-E7879.
- Karaulov, Yu. N., Sorokin Yu. A., Tarasov, Ye. F., Ufimceva, N. V., Cherkasova, G. A. (ed.). 1994, 1996, 1998. *Russian Associative Dictionary. Associative Thesaurus of the Modern Russian Language*. 3 vols. Moscow: AST-Astrel'. *Original in Russian* (Карaulов, Ю. Н., Сорокин Ю. А., Тарасов Е. Ф., Уфимцева Н.В., Черкасова Г.А. (ред.). *Русский ассоциативный словарь. Ассоциативный тезаурус современного русского языка*. М.: АСТ–Астрель).
- Kesebir, S. 2017. Word Order Denotes Relevance Differences: The Case of Conjoined Phrases with Lexical Gender. *Journal of Personality and Social Psychology* 113 (2): 262–279.
- Koplenig, A. 2017. The Impact of Lacking Metadata for the Measurement of Cultural and Linguistic Change Using the Google Ngram Data Sets – Reconstructing the Composition of the German Corpus in Times of WWII. *Digital Scholarship in the Humanities* 32 (1): 169–188.

- Li, Y. 2022. *Temporal Changes in Individualism and Collectivism in Modern China: Evidence from Google N-gram Viewer and Sina Weibo*. University of Chicago. <https://doi.org/10.6082/uchicago.3746>.
- Li, Y., Tan, X., Huang, Z., and Liu, L. 2019. Relationship between Collectivism and Corruption in American and Chinese books: A historical perspective. *International Journal of Psychology* 54 (2): 180–187.
- Lin, Y., Michel, J. B., Lieberman, E. A., Orwant, J., Brockman, W., and Petrov, S. 2012. Syntactic Annotations for the Google Books Ngram Corpus. In *Proceedings of the ACL 2012 System Demonstrations* (pp. 169–174).
- Liu, Z. 2016. A Diachronic Study on British and Chinese Cultural Complexity with Google Books Ngrams. *Journal of Quantitative Linguistics* 23 (4): 361–373.
- Madsen, D. Ø., and Slåtten, K. 2022. The Possibilities and Limitations of Using Google Books Ngram Viewer in Research on Management Fashions. *Societies* 12 (6): 171. <https://doi.org/10.3390/soc12060171>.
- Michel, J. *et al.* 2011. Quantitative Analysis of Culture Using Millions of Digitized Books. *Science* 331 (6014): 176–182.
- Mikolov, T., Sutskever, I., Chen, K., Corrado, G., and Dean, J. 2013. Distributed Representations of Words and Phrases and Their Compositionality. URL: <https://arxiv.org/abs/1310.4546>.
- Morin, O., and Acerbi, A. 2017. Birth of the Cool: A Two-Centuries Decline in Emotional Expression in Anglophone Fiction. *Cognition and Emotion* 31 (8): 1663–1675.
- Ogihara, Y. 2017. Temporal Changes in Individualism and Their Ramification in Japan: Rising Individualism and Conflicts with Persisting Collectivism. *Frontiers in psychology* 8: 695.
- Ogihara, Y. 2023. Chinese Culture Became More Individualistic: Evidence from Family Structure, 1953–2017. *F1000Research* 12 (10): 10.
- Oishi, S., Graham, J., Kesebir, S., and Galinha, I. C. 2013. Concepts of Happiness across Time and Cultures. *Personality and Social Psychology Bulletin* 39 (5): 559–577.
- Ophir, S. 2016. Big Data for the Humanities Using Google Ngrams: Discovering Hidden Patterns of Conceptual Trends. *First Monday*. URL: <https://journals.uic.edu/ojs/index.php/fm/article/download/5567/5535>.
- Paoletti, J. B. 2012. *Pink and blue: Telling the Boys from the Girls in America*. Bloomington: Indiana University Press.
- Pechenick, E. A., Danforth, C., Dodds, P., and Barrat, A. 2015. Characterizing the Google Books Corpus: Strong Limits to Inferences of Socio-Cultural and Linguistic Evolution. *PLoS ONE* 10 (10), e0137041.

- Peng, L., and Luo, S. 2022. Impact of Social Economic Development on Positive and Negative Affect among Chinese College Students: A Cross-Temporal Meta-Analysis, 2001–2016. *Culture and Brain* 10 (2): 1–18.
- Peterson, C., and Seligman, M. E. P. 2004. *Character Strengths and Virtues: A Handbook of Classification*. Oxford University Press.
- Pettit, M. 2016. Historical Time in the Age of Big Data: Cultural Psychology, Historical Change, and the Google Books Ngram Viewer. *History of Psychology* 19 (2): 141–153.
- Popescu, O., and Strapparava, C. 2014. Time Corpora: Epochs, Opinions and Changes. *Knowledge-Based Systems* 69: 3–13.
- Qiu, L., Lu, J., and Chiu, C. Y. 2014. Detecting the Needs for Happiness and Meaning in Life from Google Books. In *2014 International Conference on Orange Technologies* (pp. 145–148). IEEE Xplore.
- Roth, S. 2013. The Fairly Good Economy: Testing the Economization of Society Hypothesis against a Google Ngram View of Trends in Functional Differentiation (1800–2000). *Journal of Applied Business Research* 29 (5): 1495–1500.
- Roth, S. 2014. Fashionable Functions: A Google Ngram View of Trends in Functional Differentiation (1800–2000). *International Journal of Technology and Human Interaction* 10 (2): 35–58.
- Roth, S., Clark, C., and Berkel, J. 2017a. The Fashionable Functions Reloaded: An Updated Google Ngram View of Trends in Functional Differentiation (1800–2000). In *Research paradigms and contemporary perspectives on human-technology interaction* (pp. 236–265). IGI Global.
- Roth, S., Clark, C., Trofimov, N., Mkrtichyan, A., Heidingsfelder, M., Apignanesi, L., ... and Kaivo-Oja, J. 2017b. Futures of a Distributed Memory. A Global Brain Wave Measurement (1800–2000). *Technological Forecasting and Social Change* 118: 307–323.
- Sheahan, P. 2005. *Generation Y: Thriving and Surviving with Generation Y at Work*. Prahran: Hardy Grant.
- Skrebyte, A., Garnett, P., and Kendal, J. R. 2016. Temporal Relationships between Individualism–Collectivism and the Economy in Soviet Russia: A Word Frequency Analysis Using the Google Ngram Corpus. *Journal of Cross-Cultural Psychology* 47 (9): 1217–1235.
- Solovyev, V. D., Ahtyamov, R. B., Bajrasheva, V. R. 2011. Evolution of the Frequency of Names of Colors. *Uchenye zapiski Kazanskogo Universiteta. Gumanitarnye nauki* 153 (5): 102–109. *Original in Russian* (Соловьев В. Д., Ахтямов Р. Б., Баджрашева В. Р. Эволюция частотности наименований цветов. *Ученые записки Казанского университета. Гуманитарные науки* 153 (5): 102–109).
- Solovyev, V. D., Bochkarev, V. V., and Akhtyamova, S. S. 2020. Google Books Ngram: Problems of Representativeness and Data Reliability.

*Communications in Computer and Information Science* 1223: 147–162.  
[https://doi.org/10.1007/978-3-030-51913-1\\_10](https://doi.org/10.1007/978-3-030-51913-1_10).

- Solovyev, V., Bochkarev, V., and Bayrasheva V. 2016. Dynamics of Emotions in European Languages. *The 7th International Conference on Cognitive Science* (pp. 71–72). Institute of Psychology of RAS.
- Solovyev, V. D., and Vol'skaya, Yu. A. 2023. Towards a New Dictionary of Associations in the Russian language. In *Language as it is* (pp. 408–411). Moscow: Buki Vedi. *Original in Russian* (Соловьев В. Д., Вольская Ю. В. К новому словарю ассоциаций в русском языке. / *Язык как он есть. Сборник статей к 60-летию Андрея Александровича Кибрика*. М.: Буки-Веди).
- Sorokin, P. 1937. *Social and Cultural Dynamics. Volume 2: Fluctuation of Systems of Truth, Ethics, and Law*. New York: American Book Co.
- Su Q., and Liu, G. 2020. Birth Cohort Changes in the Subjective Well-Being of Chinese College Students: A Cross-Temporal Meta-Analysis, 2002–2017. *Frontiers in psychology* 11: 1011. doi: 10.3389/fpsyg.2020.01011.
- Twenge, J. M., Campbell, W. K., and Gentile, B. 2012. Male and Female Pronoun Use in U.S. Books Reflects Women's Status, 1900–2008. *Sex Roles* 67 (9–10): 488–493.
- Twenge, J. M., Campbell, W. K., and Gentile, B. 2013. Changes in Pronoun Use in American Books and the Rise of Individualism, 1960–2008. *Journal of Cross-Cultural Psychology* 44 (3): 406–415.
- Uz, I. 2014. Individualism and First Person Pronoun Use in Written Texts across Languages. *Journal of Cross-Cultural Psychology* 45 (10): 1671–1678.
- Velichkovsky, B. B., Solovyev, V. D., Bochkarev, V. V., and Ishkineeva, F. F. 2019. Transition to Market Economy Promotes Individualistic Values: Analysing Changes in Frequencies of Russian Words from 1980 to 2008. *International Journal of Psychology* 54 (1): 23–32.
- Wang, S., Liu, R., and Huang, C-R. 2022. Social Changes through the Lens of Language: A Big Data Study of Chinese Modal Verbs. *PLoS ONE* 17 (1): e0260210.
- Wodarz, P. F., and Harris, I. G. 2022. Historical and Contemporary Patterns of Emotional Expression in Written Texts. *Research Square*. DOI: <https://doi.org/10.21203/rs.3.rs-1982012/v1>.
- Xu, Y., and Hamamura, T. 2014. Folk Beliefs of Cultural Changes in China. *Frontiers in Psychology* 5, Article ID 1066.
- Ye, S., Cai, S., Chen, C., Wan, Q., and Qian, X. 2018. How have Males and Females been Described over the Past Two Centuries? An Analysis of Big-Five Personality-Related Adjectives in the Google English Books. *Journal of Research in Personality* 76: 6–16.

- Younes, N., and Reips, U. D. 2018. The Changing Psychology of Culture in German Speaking Countries: A Google Ngram Study. *International Journal of Psychology* 53: 53–62.
- Younes, N. 2019a. *State-of-the-Art Research Using the Google Books Ngram Viewer: Improving the Method and Investigating Cultural Change* (Doctoral dissertation). Universitat Konstanz.
- Younes, N., and Reips, U. D. 2019b. Guideline for Improving the Reliability of Google Ngram Studies: Evidence from Religious Terms. *PloS one* 14 (3): e0213554.
- Yu, F., Peng, T., Peng, K., Tang, S., Chen, C. S., Qian, X., ... Chai, F. 2016. Cultural Value Shifting in Pronoun Use. *Journal of Cross-Cultural Psychology* 47 (2): 310–316.
- Zeng, R., and Greenfield, P. M. 2015. Cultural Evolution over the Last 40 Years in China: Using the Google Ngram Viewer to Study Implications of Social and Political Change for Cultural Values. *International Journal of Psychology* 50: 47–55. <https://doi.org/10.1002/ijop.12125>.
- Zhang, R., and Weng, L. 2017. Not all Cultural Values are Created Equal: Cultural Change in China Re-examined through Google Books. *International Journal of Psychology* 54 (1): 144–154. <https://doi.org/10.1002/ijop.12436>.
- Zhu, H., and Lei, L. 2018. British Cultural Complexity: An Entropy-based Approach. *Journal of Quantitative Linguistics* 25 (2): 190–205.
- Zięba, A. 2018. Google Books Ngram Viewer in Socio-Cultural Research. *Research in Language* 16 (3): 357–375.
- Zywiak, W. H., Bobroff, R. P., and Niu, G. 2021. Black Swan Years in American English, French, German, Hebrew, and Russian: Years That Reverberate in Ngram Viewer. *Advances in Historical Studies* 10 (3): 208–214.
- Zywiak, W. H., and Niu, G. 2021. Love, Hope, Perspective, and Leadership in the Ngram Database: Solace for Modern Times. *Open Journal of Social Sciences* 9: 159–166.