

The Ngram viewer tool and its use in the social sciences: a review.

Autora: Lidia Giménez Avellaneda (53786804Q)
Tutor: Julio Gonzáles Álvarez
Curso: 2017-2018

Introduction

In 2010 Google Books presented the Ngram Viewer program whose database consists of more than 5 million digitized books, which represents the 4% of every book ever published. It is a very useful tool for linguistic researches that require lexical analysis with Big Data (Michel et al., 2011). Ngram Viewer performs n-grams searches, finally showing by means of a graph the words or phrases formed by n parts and their frequency of use per year (Figure 1). It also offers the possibility of adjusting the search criteria, letting you choose among different options such as the language, year of search, search with wildcards, gives the option of review the sources and to adjust the statistical data. The objective of the present revision work is to analyse, through scientific literature, the use of this tool in the field of psychology, and by extension, in social sciences.

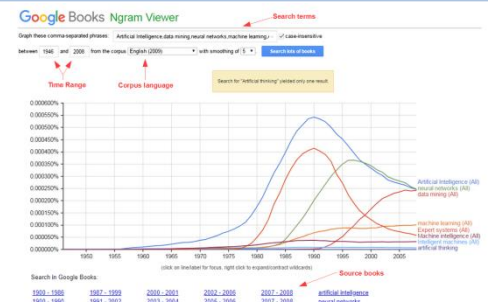
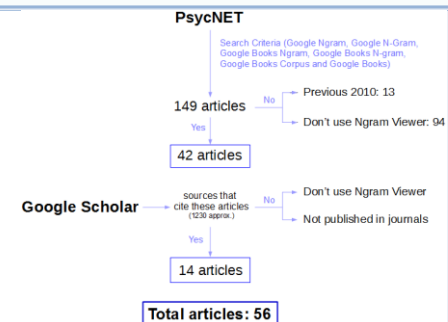


Figure 1. The Google Books bigram percentage of AI-related terms using the English 2009 corpus (Omar, Mehmood, Choi & Park, 2017)

Method

The databases used in this bibliometric review were PsycNET and Google Scholar. Firstly, a PsycNet search was performed using the terms *Google Ngram*, *Google N-Gram*, *Google Books Ngram*, *Google Books N-gram*, *Google Books Corpus* and *Google Books*. From the articles obtained, those prior to 2010 (release date of Ngram Viewer) and those that did not use this tool in their methodology were eliminated, finally obtaining 42 articles. After that, through Google Scholar, all the sources that quoted the articles previously obtained were reviewed. The sources that met the following conditions were selected: a) articles published in scientific journals; b) Ngram Viewer used in their methodology; and c) not included in the previous set. As a result, 14 additional articles were obtained.



Results

The 56 articles were classified, according to their content, in the following categories.

Género	Género (G)	4
Lenguaje	Lingüística (Li)	6
	Psicolingüística (PL)	6
	Personalidad (P)	3
Psicología	Inteligencia (I)	3
	Aprendizaje (A)	3
	Emoción (E)	4
	Historia de la psicología (HP)	4
	Desarrollo cultural (DC)	7
Socio-cultural	Individualismo/Colectivismo (IC)	10
	Ecología y cultura (EC)	4
	Otros (O)	2

- EC Arias, M., & Dewas, S. (2015). Nature as relationship partner: An old frame revisited. *Environmental Education Research*, 21(6), 1656-1678.
- LI Buchkann, V., Solovyev, V., & Wichmann, S. (2014). Universalist versus historical contingencies in lexical evidence. *Journal of the Royal Society Interface*, 11(161), 20140864.
- PL Bybee, M., Buchheimer, M., Conrad, M., Jacobs, A. M., Billo, J., & Bohn, A. (2011). The word frequency effect: A review of recent developments and implications for the choice of frequency estimates in German. *Experimental Psychology*, 58(5), 414-424.
- PL Bybee, M., Keilman, E., & New, S. (2011). Assessing the usefulness of Google Books' word frequencies for psycholinguistic research on word processing. *Frontiers in Psychology*, 2, Article ID 27.
- CD Cai, H., Zhu, Y., Liu, Y., & Jing, Y. (2015). Increasing neural responses for Chinese in contemporary China: Empirical evidence. *Frontiers in Psychology*, 6.
- CD Cai, H., Zhu, Y., Liu, Y., & Jing, Y. (2015). Increasing neural responses for Chinese in contemporary China: Empirical evidence. *Frontiers in Psychology*, 6.
- PL Coltrane, S. (2013). The processing of metaphor uses in sentence reading and self-pretention using eye movement data: Implications for theories of MVUs. *Dissertation Abstracts International Database*, 17434-17434.
- E Dehaene, M., & Brannon, B. (2018). Terror in time: Entering culturomics to address basic terror management theories. *Evolutionary Psychology*, 16(4), 1555-1563.
- GD DiGirolamo, M. (2017). Pink, blue, and gender: An update. *Archives of Sexual Behavior*, 46(6), 1555-1563.
- CD Fink, R., & McClos, J. M. (2017). Reviewing myths of bias: Charting cultural and religious trends with Google's Ngram Viewer. In R. Fink & C. D. Baker (Eds.), *Faithful measures: New methods in the measurement of religion* (pp. 237-216). New York, NY: New York University Press.
- HP Garavito, J. E. C. (2015). Interest in astrology and phrenology over two centuries: A Google Ngram study. *Psychological Reports*, 117(3), 540-542.
- IC Grant, O. B. (2017). Exploring the possibility of peak individualism, humanity's existential crisis, and an emerging egalitarian. *Frontiers in Psychology*, 8, Article ID 1478.
- EC Greenfield, P. M. (2013). The changing psychology of culture from 1800 through 2000. *Psychological Science*, 24(9), 1722-1731.
- IC Hanumanth, T., & Xu, Y. (2015). Changes in Chinese culture as examined through changes in personal pronoun usage. *Journal of Cross-Cultural Psychology*, 46(7), 850-861.
- Hoffman, M. J., Barnham, C., & Rensu, S. (2017). Benchmarking n-grams, topic models and recurrent neural networks by cross-complexity, EEGs and eye movements: In *Cognitive Approach to Natural Language Processing* (pp. 197-215).
- LI Johnson, P. (2015). The use of semantics, pre-emption and slow in linguistic distribution: The case of the uncorrection. *Frontiers in Psychology*, 6, Article ID 985.
- LI Johnson, P., Hovav, J., Delgado, M., & Jaeger, R. (2016). Linguistic position in historical texts reflects dynamic environmental and psychological factors. *PLoS Proceedings of the National Academy of Sciences of the United States of America*, 13(40), E078170.
- G Keiser, S. (2017). Word order denotes valence differences: The case of conjoined phrases with lexical gender. *Journal of Personality and Social Psychology*, 113(2), 302-379.
- La Lee, V., Yuan, M., Ye, L., & Reicher, M. (2016). Reconstructing the influence on and focus of the learning experience from the field's published conference proceedings. In M. A. Chase, M. J. Packer, & R. K. Sawyer (Eds.), *Current perspectives in social and behavioral sciences: Reflections on the learning sciences* (pp. 105-122). New York, NY: US: Cambridge University Press.
- IC Li, Y., Tan, X., Huang, Z., & Liu, L. (2017). Relationship between collectivism and corruption in American and Chinese books: A historical perspective. *International Journal of Psychology: Advance online publication*.
- Liu, L., Mowtow, J., & Assi, G. S. (2013). Generating example contexts to help children learn word meaning. *Natural Language Engineering*, 19(5), 519-522.
- Liu, Z. (2015). A diachronic study on British and Chinese cultural complexity with Google Books Ngrams. *Journal of Quantitative Linguistics*, 22(4), 381-372.
- Mason, S. E., Kurtz, C. V., & McGill, C. M. (2015). Oldsters and Ngrams: Age stereotypes across time. *C Psychological Reports*, 117(3), 540-542.
- I Marie, M. A., Fernandez, H., & José Figueredo, A., & Malsberg, G. (2015). By their words ye shall know them: Evidence of general intelligence and contextual environmental enrichment in vocabulary usage since the mid 19th century. *Frontiers in Psychology*, 6, 361.
- CD Moshir, C., Li, J., & Li, J. (2015). The evolution of social identity terms in lay and academic sources: Implications for research and public policy. *Analyses of Social Issues and Public Policy (ASAP)*. Advance online publication.
- P Montagna, M., & Morgan, M. (2013). Drugs on the Internet: part 4. Google's Ngram Viewer analysis tool applied to drug literature. *Substance Use and Misuse*, 48(5), 414-419.
- E Morin, O., & Acosta, A. (2017). Birth of the cool: A neo-constructivist decline in emotional expression in anglophone fiction. *Cognition*, 158(1), 160-167.
- L Mowtow, J., Gales, D., Elson, R., & Gouvan, R. (2015). Automatic Identification of Nouns: Contexts for Learning Vocabulary Words. *International Educational Data Mining Society*.
- PL Mouton, A., & Richard, C. (2015). Lexical influences on spoken spontaneous word recognition in hearing-impaired patients. *Frontiers in Neuroscience*, 9, 476.
- HP O'Sullivan, D. P., Duffy, R. M., & Kelly, D. B. (2017). Culturomics and the history of psychiatry: testing the Google Ngram model. *World Journal of Psychological Medicine*, 8, 5.
- IC Ogilvie, Y. (2017). Temporal changes in individualism and human religiosity in Japan: rising individualism and conflicts with parenting collectivism. *Frontiers in Psychology*, 8, 695.
- E Oishi, S., Graham, J., Kessler, S., & Graham, I. C. (2013). Conceptions of happiness across time and cultures. *Personality and Social Psychology Bulletin*, 39(5), 559-577.
- I Omar, M., Mehmood, A., Choi, G. S., & Park, H. W. (2017). Global mapping of artificial intelligence in Google and Books. *Scientometrics*, 113(3), 1549-1559.
- LI Pechenik, E. A., Darfom, C. M., & Doodis, P. S. (2015). Characterizing the Google books corpus: Strong links to increases of socio-cultural and linguistic evolution. *PLoS ONE*, 10(10), Article ID e0137044.
- HP Pianta, M. (2015). Historical trends in the age of hip data: Cultural, psychological, historical, change, and the Google Books Ngram Viewer. *History of Psychology*, 18(1), 141-152.
- E Popescu, O., & Szepesvari, C. (2014). Time corpus: Epochs, opinions and changes. *Knowledge-Based Systems*, 63, 3-18.
- LI Rial, R. H., Islam, A., & Miles, E. (2018). Improving text readability by incorporating phrase-level stress with word readability. *Computational Intelligence: Advance online publication*.
- I Rothermel, E. (2014). Changes in word usage frequency may hamper intergenerational comparisons of vocabulary skills: An ngram analysis of Webster's 1918 and 1913-14 forms. *Journal of Psycholinguistic Assessment*, 20(1), 83-87.
- P Rothermel, E. (2015). Personality differences in Twitter tweets and in the Google books corpus: An analysis of the latent structure of the openness factor of personality. *Current Psychology: A Journal for Diverse Perspectives on Health, Psychological Issues*, 34(4), 461-465.
- CD Roth, R. (2014). Fashionable functions: A Google Ngram view of trends in functional differentiation (1800-2000). *International Journal of Technology and Design Education*, 22(2), 35-58.
- O Roth, R., Clark, C., Trifunovic, N., Mochizuki, A., Haidichkin, M., Agapinovic, L., & Kivoo-Oja, J. (2017). Future of a distributed network: A global brain wave measurement (1800-2000). *Technological Forecasting and Social Change*, 118, 307-323.
- IC Savelbergh, A., Ganser, P., & Kordel, 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), 1217-1226.
- LI Solovyev, V., & Ivanov, V. (2016). Knowledge-driven event extraction in Russian: Corpus-based linguistic resources. *Computational Intelligence and Neuroscience*, 20(6), Article ID 437820.
- PL Thirteenth, A. Age, E., Johnson, A., Magill, M. Z., Baraballo, T., & Neuman, A. J. (2016). What the internet told us about sexual and gender observation. *The Medical Journal*, 11(1), 115-160.
- G Twenge, J. M., Campbell, W. K., & Gentile, B. (2012). Male and female pronoun use in U.S. books reflects century's trends, 1900-2008. *Sex Roles*, 67(10), 488-493.
- IC Twenge, J. M., Campbell, W. K., & 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.
- IC Uz, I. (2014). Individualism and first person pronoun use in written texts across languages. *Journal of Cross-Cultural Psychology*, 45(10), 1671-1676.
- IC Volkovskiy, B. B., Solovyev, V. D., Sotnikov, V. V., & Ishkhanova, F. F. (2017). Transition to market economy promotes individualistic values: Analyzing changes in frequencies of function words from 1980 to 2008. *International Journal of Psychology: Advance online publication*.
- HP Wang, C., & Peng, J. (2015). History of "behavior" and "mind": Use of behavioral and cognitive terms in the 20th century. *The Psychological Record*, 65(1), 23-30.
- CD Xu, Y., & Hanumanth, T. (2015). Risk related cultural changes in China. *Frontiers in Psychology*, 6, Article ID 1066.
- CD Xu, Y., Cai, S., Chen, C., Wang, Q., & Qian, X. (2016). Have new media and Internet been described over the past two centuries? An analysis of Big Five personality-related adjectives in the Google English Books. *Journal of Quantitative Linguistics*, 23(4), 381-372.
- IC Younan, N., & Rios, L. D. (2017). The changing psychology of culture in german-speaking countries: A Google ngram analysis. *International Journal of Psychology: Advance online publication*.
- IC Yu, F., Peng, T., Peng, K., Tang, S., Chen, C. S., Qian, X., ... & Chai, F. (2016). Cultural value shifting in Chinese. *Journal of Personality and Social Psychology*, 111(2), 316-318.
- EC Zhang, R., & Greenfield, P. M. (2015). Cultural evolution over the last 40 years in China: Using the Google Ngram Viewer to track implications of history and political change for cultural values. *International Journal of Psychology*, 50(1), 47-55.
- CD Zhu, Y., & Li, J. (2015). Not all cultural values are created equal: Cultural change in china reexamined through google books. *International Journal of Psychology: Advance online publication*.
- CD Zhu, Y., & Li, J. (2015). Global Cultural Complexity: An Entropy-based Approach. *Journal of Quantitative Linguistics*, 22(2), 190-205.

Conclusions

The review of these articles suggests a greater use of Ngram Viewer to reflect cultural and social changes through modifications in the frequency of use of representative terms of concepts or values (e.g., the cultural and historical variations of the concept of happiness). In psychology, it has been used in different fields such as the study of the diachronic evolution of emotional terms or the vocabulary subtests of intelligence tests.

On the other hand, there are works dedicated to linguistics and psycholinguistics; in this last, word frequencies of Ngram Viewer are used in the investigation of word recognition and processing.

Moreover, Ngram Viewer shows important results in the field of gender studies. The application of this tool provides an observation of the evolution of social gender concepts, the representation of genres in written literature and the consequently associated stereotypes (e.g., the prevalence of masculine pronouns versus feminine pronouns in number and order of appearance, or the changing association of pink and blue colours to masculine and feminine gender).

In addition, we can also appreciate some sociocultural and anthropological studies that identify important changes in society through concepts such as individualism versus collectivism or establish a relationship between ecological and cultural changes.

Finally, Ngram Viewer is a young tool that so far has not been widely used, as we can see from the few number of publications online, but it presents new opportunities and opens up other future lines of research, for example, it offers the "big data" perspective on the cultural trends over the decades.

Referencias:

Michel, J. B., Shen, Y. K., Aiden, A. P., Veres, A., Gray, M. K., The Google Team, ... & Aiden, E. L. (2011). Quantitative analysis of culture using millions of digitized books. *Science*, 331, 176-182