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.
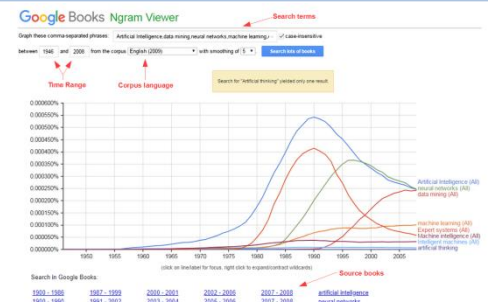

Figure 1. The Google Books bigram percentage of AI-related terms using the English 2009 corpus (Omar, Mehmood, Choi & Park, 2017)

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.
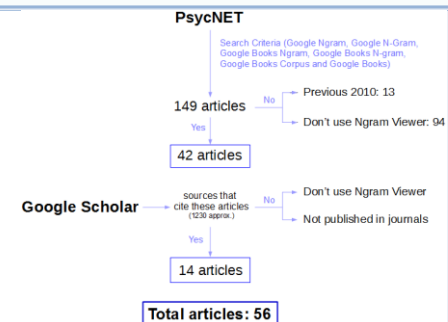
Results

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

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

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