

USE OF GOOGLE TRENDS IN DETERMINING ROMANIAN PUBLIC OPINION TOWARDS ENGLISH LANGUAGE COURSES

Gyongyver Măduța*

Abstract

Public interest towards foreign language courses is a major factor in influencing consumer behaviour, university policy and even public policy towards a more educated population. By using a “big data” approach, freely available through the Google Trends service, an assessment was made with regards to Romanian public interest towards English language courses available on the market, as well as the regions with the highest demand for English as a second language over the past five years. The results are relevant to academia, language centres, cultural centres, English schools and Universities with English- taught programmes.

Keywords: *Consumer behaviour, Google Trends, English language, public interest, language courses*

JEL Classification: D83, P36, P46, Z11

1. Introduction

The present research involves an interpretive approach to a current issue, respectively Romanian public interest towards foreign language courses by using the Google Trends service. There are two types of Google Trends data that can be accessed: the first type consists of real-time data over the last seven days of queries, and the second type consists of non-real-time data starting in 2004 and leading up to the past 72 hours before the data inquiry. The sample is representative because although the Google Trends service only uses a sample of Google user searches, Google features billions of searches daily. The full data set is even too great for any artificial intelligence to access and process it quickly. To this end, data sampling is used to analyze a representative set of Google searches and obtain reliable statistics which are also actionable even within minutes of the occurrence of a real-world event. For easy comparisons between terms, search data is normalized based on the time and location of a query.

It is worth mentioning that Google Trends data differs from Google AdWords Search Terms report which is built for monthly searches and features average search volume statistics which are specifically designed for advertisers, while Google Trends focuses on real-time data.

Ginsberg et al. (2009) were among the first to include Google Trends in research studies, where they demonstrated Google Trends tracking and possibility of prediction

* Gyongyver Măduța is Assitant Professor at the Romanian American University in Bucharest. E-mail: gyongyver.maduta@rau.ro

for the spread of influenza even earlier than Centers for Disease Control and Prevention (CDC).

As consumers act in their information seeking stage to reduce uncertainties and avoid risks in the purchase process, be they psychological or financial, the trend is to adopt information seeking behaviors by searching the web, causing increases of Internet accessibility. Online searches will inevitably lead to online purchases, hence the positive correlation between a user's willingness to search online and purchase a language service online [Goel et al., (2010)]. Search is thus an act that accompanies a purchase or the acceptance of an innovation. Statistics on search activities are thus useful resources for monitoring activities, analysis and even predictions on acceptance of a service or a new technology by consumers or the general public [Jun et al., (2019)], hence the proliferation of studies using search engine data even before the appearance of the Google Trends service.

Web data sources, such as search engine terms' traffic, web traffic, and customer feedback oWith the increase in availability of 'big data' sources, they are naturally used for various demand predictions and public interest evaluations towards given topics.

English has remained the working language of many companies with foreign capital in Romania and, moreover, in the company jargon, whether it is banking, retail or especially IT, many English terms abound.

2. Methods

The purpose of the study is to identify the public interest towards English language learning in Romania by using the Google Trends service as a research tool.

The Google Trends data is available as a time series for each individual keyword starting from 2004. Each individual value is proportional to the frequency of the search term. Google normalizes each time series between 0 and 100, with 100 corresponding to the maximum of the time series. Up to 5 multiple search keywords can be selected; in this case all series are normalized with respect to the maximum value of the multiple keywords. While Google makes available only normalized data, which makes the frequency data immune to changes in total search volume, other effects are not taken into consideration. One of the most important factors which can affect the data is the change in Romanian speaking Google users, as a fraction of total search data. In order to account for this variable, the Google search data frequency for "este" ("it is") and "are"("it has") were used as a baseline.

The specific objectives of the current study were stated as follows:

- Identifying the changes occurring on language learning searches in Romania up to the end of 2022.
- Identifying the potential need to adapt the offer and subjects taught in order to facilitate an efficient allocation of language learning resources.
- Identifying the main changes in the evolution of demand for English language courses as a result of digitization.

3. Results

The raw data provided by Google Trends for 5 keywords related to English language ("cursuri engleza" translated as English courses, "scoala engleza" translated as English school, "engleza incepatori" translated as beginner English, "limba engleza"

translated as English language, “invata engleza” translated as learn English) is presented in Figure 1. The “cursuri engleza” and “limba engleza” keywords show a period of maximum use during 2004 -2007 time range, followed by a gradual decline in search frequency. The other keywords show increased use after 2008.

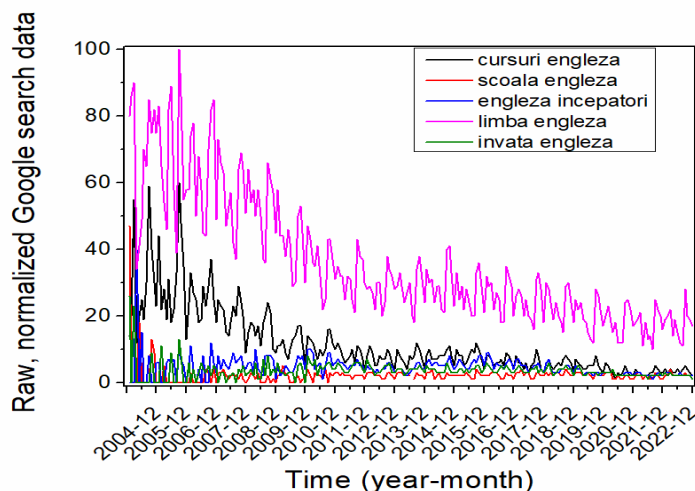


Figure 1. Raw, normalized Google search data frequency for the 5 keywords presented as monthly data

The baseline data shows an increasing trend over the considered time range (Figure no. 2). This indicates an increased use of Google searches using the Romanian language, not necessarily an increase in total user counts.

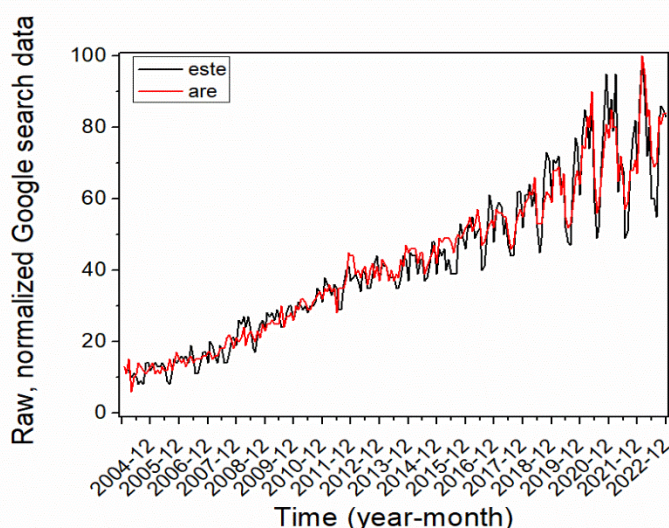


Figure 2. Raw, normalized Google search data frequency for the 2 baseline keywords presented as monthly data

Both the 5 keywords as well as the baseline data exhibit significant month-by-month variation. The average monthly search frequency values were computed for each year starting from 2004 and presented in Figure no 3.

The current language training market is evolving by adapting to technological progress, learning methods are becoming more flexible, personalized and more participatory with the involvement of the learner himself in managing his own training. The current offers of the foreign language centers that are also active in Romania are oriented towards the comfort of the learner, by overcoming the impediments regarding the place, the schedule, the learning activities. In addition to these, it is worth noting offers of foreign language learning services focused on training projects that emphasize the learner's acquisition of learning objectives, contents, tasks and activities, as well as exclusively online offers, through the possibilities offered by applications for distance learning, e-learning, multimedia, etc. that open new perspectives for the development of innovative formulas for training credits or internationally recognized accreditations.

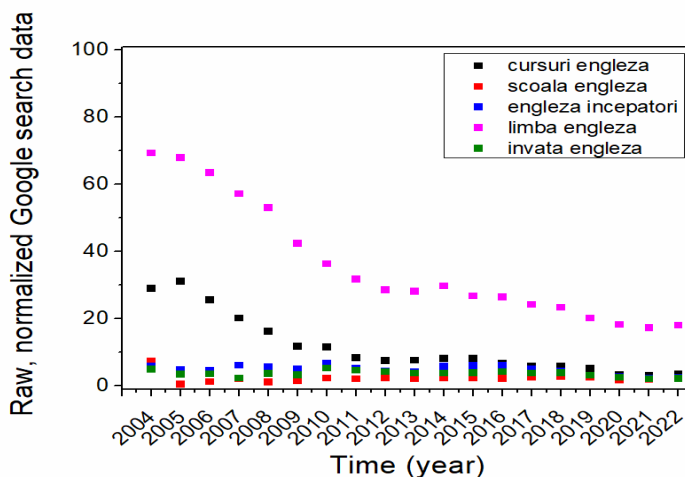


Figure 3. Google search data frequency for the 5 keywords presented as the yearly average frequency data

Google Trends also offers regional query search data. The regional breakdown of the 5 investigated keywords is presented in Figure no. 4. Interestingly, only the “limba engleza” keyword has been searched for in all of Romanian counties. The other 4 keywords only appear in significant numbers in the most urban regions of the country: Bucharest-Ilfov, Cluj County, Timiș County and Iași County. These regions correspond to the capital and the 3 largest cities in Romania.

According to the evaluation, English remains at the top of students' preferences, occupying around 80% of the search terms on the foreign language courses market. The demand remained constant in this segment, especially due to a need for better defined courses, specialized in business and negotiations.

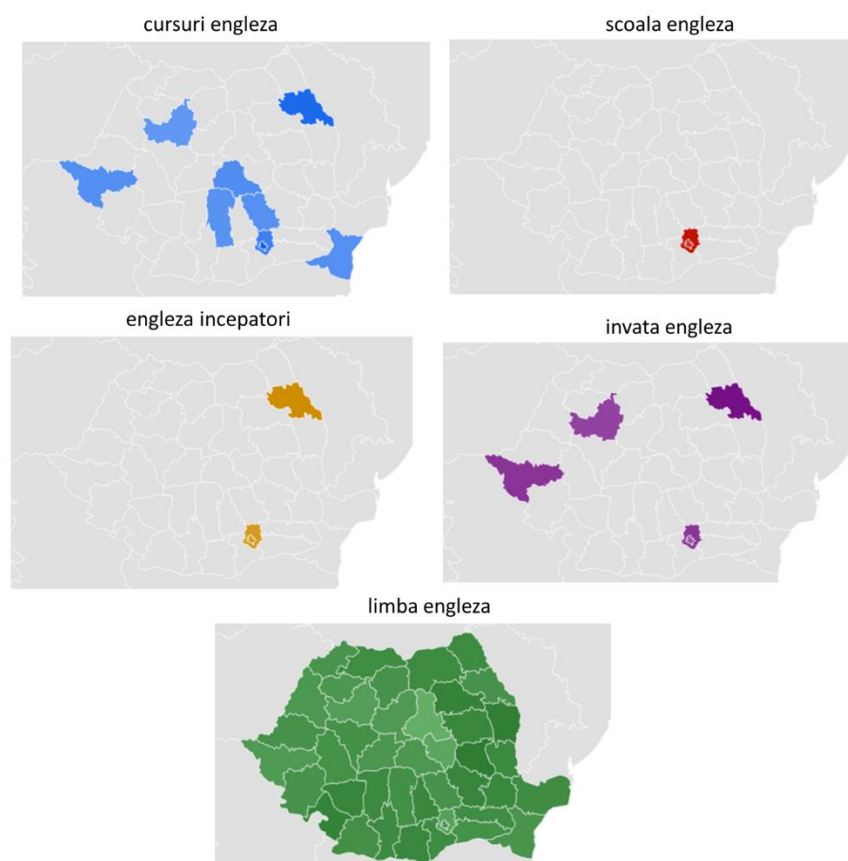


Figure 4. Regional distribution of the keyword search frequency.
Darker colours indicate higher frequency

Note from Figure no. 4 that English is everywhere in the country, probably because of the subject being taught in school. The rest of the keywords have a large spread only in Bucharest, Cluj, Iasi and Timisoara. Another key word that is also used in Arges and Prahova counties – likely because these regions also feature the Dacia factories, the oil refineries Arpechim and Petrobrazi, and due to the intense tourism specific to the Prahova Valley. It is worth mentioning that in Romania, the oldest native foreign language centers (Fides, Lexis) were established in 1993. The 90s are the decade when the privatization of state enterprises in Romania began, with the possibility of being bought by foreign investors, however, in lack of a legislative framework conducive to foreign investments, this was achieved hesitantly. In addition, economic, legislative and political play an important role in learning foreign languages.

4. Conclusion

Our work provides extensions to previous research featuring Google Trends data by methodology improvement and focus on country-specific search queries on language learning data, where such analyses have not been previously performed. By use of previously determined search queries, the analysis was extended by using

search terms specific for national regions of interest. The most relevant queries which would suggest strong relevancy for language learning queries and related terms were also compared to baseline terms, for a more accurate representation. The search query volume that was geographically restricted to Romania shows the possibility of improvements in predictions of demand occurrences in Romania and the existing supply.

The present study's results support the value of indirect sources use for the forecasting of demand for language courses in Romania. Such studies constitute an added benefit for Romanian public officials as they can anticipate changes in demand occurrence, indicating when action could be applied for better public and private education and a better prepared Romanian labor force. Hence, this research has created a novel tool for improving current education policies in Romania.

References

1. Jun, S. P., Yoo, H. S., & Choi, S. (2018). *Ten years of research change using Google Trends: From the perspective of big data utilizations and applications*. Technological forecasting and social change, 130, 69-87.
2. Ginsberg, J., Mohebbi, M. H., Patel, R. S., Brammer, L., Smolinski, M. S., & Brilliant, L. (2009). *Detecting influenza epidemics using search engine query data*. Nature, 457(7232), 1012-1014.
3. Goel, S., Hofman, J. M., Lahaie, S., Pennock, D. M., & Watts, D. J. (2010). *Predicting consumer behavior with Web search*. Proceedings of the National academy of sciences, 107(41), 17486-17490.
4. Barros, J. M., Melia, R., Francis, K., Bogue, J., O'Sullivan, M., Young, K., ... & Duggan, J. (2019). *The validity of Google Trends search volumes for behavioral forecasting of national suicide rates in Ireland*. International journal of environmental research and public health, 16(17), 3201.
5. Petropoulos, A., Siakoulis, V., Stavroulakis, E., Lazaris, P., & Vlachogiannakis, N. (2021). *Employing Google Trends and Deep Learning in Forecasting Financial Market Turbulence*. Journal of Behavioral Finance, 1-13.
6. Höpken, W., Eberle, T., Fuchs, M., & Lexhagen, M. (2019). *Google Trends data for analysing tourists' online search behaviour and improving demand forecasting: the case of Åre, Sweden*. Information Technology & Tourism, 21(1), 45-62.