

Corpus linguistics for the acquisition of English lexicon

Fausto Luna, César Aguilar

Universidad Veracruzana, Mexico

<https://doi.org/10.36505/ExLing-2024/15/0017/000642>

Abstract

This work is a continuation of the research made by Luna & Aguilar (2022), and Aguilar & Luna (2023), focused on analysing what impact the use of linguistic corpora has on the teaching of English, specifically on students who have not had any previous approach to corpus linguistics. Our theoretical framework is in line with the proposals made by Granger (1998, 2004) regarding the use of corpora for learning English, as well as that of Boulton (2009, 2010) on language learning based on data recognition, via the use of digital corpora. Likewise, studies carried out by and Moon (2010), Szudarski (2018) or Friginal (2018) regarding the employ of corpus for vocabulary acquisition can also be mentioned here. What we present here is a quasi-experimental study, applied to a population of 36 participants, divided into two groups: one will be the intervention group and the other will be the control group. The results of this study might allow us to observe and understand the impact that teaching techniques and resources from corpus linguistics has on the acquisition of advanced vocabulary in English as well as the differences between the skills that English learners have to increase their lexicon, whether or not they are able to use computer tools such as electronic corpora. And, to recognise if the design of a course model for learning advanced English -which can be an important support for successfully passing a certification exam such as the CAE.

Keywords: CAE, corpus linguistics, English, learning, lexicon

Introduction

The use of Corpus linguistics (CL) has turned up as an increasingly relevant research topic in recent years, in particular, regarding its application in language teaching. Conspicuously, Sylviane Granger (2004) stands out as one of the key figures and pioneers in the field of corpus linguistics for foreign language learning. In this scope, corpus linguistics has proven to be a highly valuable tool for teaching foreign languages, particularly English. Consequently, this methodology, as defined by scholars such as McEnery (1996), Granger et al. (2002), and Bolaños (2015), can be employed not only for general English instruction but also to enhance specific language skills that students require to develop as part of their training.

Research problem

The research problem highlights that students at the BA Lengua Inglesa at the Languages School who are taking or have already concluded the Educational Experience (EE) Advanced English, as well as the graduates from the same educational program are not able to reach the language level desired on account of several circumstances even when they accomplished all the English levels offered by such program. As a result, several students do not feel confidence, and in turn, they find taking a C1 certification test quite complex.

Hypothesis and null hypothesis

The hypothesis presented in this paper states that the use of a CL made of sample tests of CAE might be employed as an essential tool which can allow students to acquire advanced English lexicon in order to pass such English certification successfully. Furthermore, the learners might increase their lexicon significantly. Notwithstanding, while it is true that the utilization of CL might lead to an increase in the range of vocabulary students and graduates possess, there is also the possibility of stating a null hypothesis. This is to say that, using CL could not produce significant improvements in lexicon acquisition or certification outcomes, suggesting that other factors could influence such results.

Research question and objective

This study aims to find out and asses the effect of corpus linguistic utilization on the development of advanced English lexical competence among English Language undergraduates and graduates. A pre- and post-test design is also employed, comparing an intervention group (receiving corpus-based instruction) with a control group, so as to examine the influence of the use of CL on lexical proficiency. In this regard, the research question in this work is stated as follows: What is the influence of using the CL on the lexical competence in English of the participants of the Bachelor's Program in English Language?

Background

CL for foreign language teaching	CL for English teaching	CL for lexicon teaching
- Beggs (2011)	- Ramón (1999)	- Terrazo (2006)
- Granger (2012)	- López (2008)	- Pérez-Álto (2006)
- Zichensky (2014)	- Vito (2008)	- Cabré (2000)
- Bellas (2015)	- Rice (2012)	
- Douai (2018)		

Figure 1. The state of the art.

Methodology

This paper involves a multi-faceted research approach, combining exploratory, correlational, comparative, and quasi-experimental designs to achieve a comprehensive understanding of the research phenomenon.

Research design

It is imperative to highlight that, a mixed-method approach was adopted, integrating both quantitative and qualitative methods to collect and analyse data. The quantitative component involves the collection of numerical data through surveys and the units of analysis, while the qualitative component consists of the collection of textual data through semi-structured interviews.

Context and informants

This research paper has been developed with informants who are current students or graduates from the BA Lengua Inglesa from the Languages School of the Universidad Veracruzana. Such higher education institution is located in the capital city Xalapa, in Veracruz, Mexico. The intervention group consisted of 18 participants, comprising 12 undergraduate students and 6 graduates. All participants were enrolled in both the Bachelor of Arts in Lengua Inglesa (12 students) and the Bachelor of Arts in Enseñanza del Inglés (1 virtual student). A notable characteristic of this group was that all participants had completed an advanced English course, demonstrating a high level of proficiency in the language. Demographically, the participants' ages ranged from 21 to 47 years old. In terms of language proficiency, all participants had achieved a level of B2-B2+ according to the Common European Framework of Reference for Languages (CEFR). This homogeneity in language proficiency ensured that the intervention was tailored to the participants' needs and abilities.

Data collection methods

This research work employed a multi-method approach to collect data. A demographic questionnaire has been used to gather information about participants' background and characteristics. A satisfaction survey is administered to assess participants' perceptions and satisfaction with the intervention. Semi-structured interviews are conducted to gather in-depth information about participants' experiences and opinions. Furthermore, participant observation is definitely used so as to gather data on informants' behaviour and interactions during the intervention.

Units of analysis

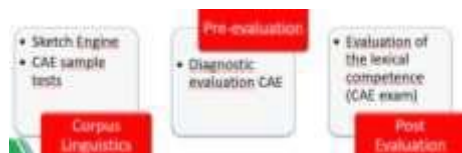


Figure 2. Units of analysis.

Preliminary results, discussion and conclusion

First of all, the informants' language level at the beginning of the course ranged from B2 to B2+. In terms of certification background, the participants reported holding various certificates, including TOEFL ITP (14.3%), PET (9.5%), EXAVER 3 (9.5%), while 19% do not hold any certificate. Remarkably, all participants (100%) agreed on the relevance of Lexicon in a certification test. Aside from that, preliminary findings from this ongoing research indicate that participants initially struggled with the CAE diagnostic test, obtaining low grades and citing the test's complexity. Additionally, informants demonstrated limited familiarity with advanced lexicon, admitting to numerous unknown terms in an online test. Nonetheless, participants in the intervention group have shown keen interest in utilizing the Sketch Engine platform and Corpus Linguistics tools to improve their vocabulary register. With regular sessions and instructor guidance, participants are making progress in handling the platform and understanding new lexicon. While only preliminary conclusions can be drawn at this stage, the study aims to yield positive and significant results that support its main hypothesis. Further results will be reported upon completion of the research.

References

- Aguilar, C., Luna, F. 2023. Today is another day to find you: Using Corpus Linguistics For Language Learning. In Gómez, L.; González, Ch., Lees, J. (eds.), ICERI2023 Proceedings 16th International Conference of Education, Research and Innovation, IATED Academy, Valencia, Spain: 8042-8050.
- Boulton, A. 2009. Data-driven learning: reasonable fears and rational reassurance. *Indian Journal of Applied Linguistics*, 35(1): 81-106.
- Boulton, A. 2010. Data-Driven Learning: taking the computer out of the equation. *Language Learning*, 60(3): 534-572.
- Friginal, E. (2018): *Corpus Linguistics for English Teachers*, London, Routledge.
- Granger, S. (1998): *Learner English on Computer*, London, UK, Longman.
- Granger, S. 2004. Computer learner corpus research: Current status and future prospects. In Connor U., Upton, T. (eds.), *Applied Corpus Linguistics: a Multidimensional Perspective*. Brill, Amsterdam: 123-145.
- Luna, F., Aguilar, C. 2022. Corpus linguistic tools and terminology translation tasks. In Botinis, A. (ed.), *Proceedings of 13 th International Conference of Experimental Linguistics*, International Society of Experimental Linguistics, Paris, France: 121-124.