The influence of reading on the vocabulary in mother tongue and foreign language

Inmaculada-Clotilde Santos-Díaz
Universidad de Málaga

Abstract
The aim of this article is to observe the influence of reading frequency on the vocabulary in Spanish and in a foreign language (English and French). The sample comprises 171 students of the Master’s Degree in Teaching Training at the University of Málaga. The data has been gathered through a test of lexical availability and a test to identify technical terms (in English, Spanish and French), as well as through a sociological questionnaire. The results confirm the initial hypothesis that reading reinforces the lexical competence and especially the passive vocabulary. Reading takes on a main role in foreign language, mostly in English. It favours not only an increase of updated words but also a better identification of technical words. Nevertheless, reading in the mother tongue has an influence only on the passive vocabulary. The academic level attained by informants standardises the lexical competence, which is partially improved by reading frequency.

Resumen
Este artículo tiene como objetivo conocer la incidencia de la frecuencia de lectura en la competencia léxica en español y en lengua extranjera (inglés y francés) de una muestra de 171 estudiantes del Máster Universitario en Profesorado de la Universidad de Málaga. Los datos se han recabado a través de una prueba de disponibilidad léxica, de un test de identificación de tecnicismos (en español, inglés y francés) y de un cuestionario sociológico. Los resultados corroboran la hipótesis de partida que postula que a través de la lectura el informante puede reforzar su caudal léxico y, especialmente, el vocabulario pasivo. En lengua extranjera, sobre todo en inglés, la lectura cobra un papel principal favoreciendo tanto la actualización de un mayor número de palabras como el reconocimiento de tecnicismos. En cambio, en lengua materna, el efecto de la lectura es significativo tan solo en el vocabulario pasivo. El nivel de estudios alcanzado por los informantes estandariza la competencia léxica que se ve mejorada parcialmente por la frecuencia de lectura.
Introduction

There are many studies that demonstrate the key role played by reading when acquiring and learning a second and a foreign language (Brown, Waring & Donkaewbua, 2008; Horst, 2005; Paribakht & Weshe, 1999; Pigada & Schmitt, 2006; Schmitt, 1998; Waring & Takaki, 2003; Zahar, Cobb & Spada, 2001). Students are able to learn a large number of lexical units incidentally through reading (Coady, 2000; Elley & Mangubhai, 1983). Reading, seen as an educational objective by the Ley Orgánica para la Mejora de la Calidad Educativa (Spanish Organic Act on Improvement of Education's Quality, LOMCE) is analysed in depth by researchers such as Larrañaga & Yubero (2005) and Yubero & Larrañaga (2010), who analyse reading behaviours and their link to reading habits. Nevertheless, the influence of reading is a variable that is not commonly used in the studies on lexical availability.

This area of research appeared in France in the middle of the twentieth century and aimed at creating a basic vocabulary list to help immigrants arriving in France and the people of the French Union to learn French. To that end, Gougenheim, Michéa, Rivenc & Sauvageot (1956) presented a basic vocabulary list created on the basis of frequently used and available vocabulary. The starting point for further research focuses on the idea of “available vocabulary” defined by Michéa (1953, p. 340) as a word that is not particularly frequent but is always available and instantly and naturally comes to mind when necessary.

The data collection method of the available vocabulary proposed by Gougenheim et al. (1956, 1964) consisted on asking the students to write twenty words related to centres of interest linked to everyday life. Michéa (1953, p. 39) refers to traditional pedagogy to explain that centres of interest are varied groupings of a certain kind of objects or places that allow making associations in different directions. He proposes a total of sixteen centres aiming at collecting all the basic vocabulary of a given language.

This method was modified by Dimitrijević (1969), who introduced the open-list system so that informants can make contributions with all the words they can in each centre during a determined period of time. Nevertheless, the most widely used methodological criteria in this field emerged after the Proyecto Panhispánico de Disponibilidad Léxica (Pan-Hispanic Project of Lexical Availability) developed in the 90s. The data collection process was performed in line with the open-list system and allocated two minutes to each centre of interest of the total of sixteen envisaged in the French work. The objective of the Pan-Hispanic project, coordinated by López-Morales, was to identify the available vocabulary of the different Spanish-speaking countries through availability dictionaries created following the same principles.


There is less research in the case of English and French. In the case of English, we should mention the studies by Dimitrijević (1969), Germany & Cartes (2000), Ferreira (2006), Zubanov (2006), Gallardo-del-Puerto & Martínez-Adrián (2014) & Fuentes-Páez (2014). In the case of French, in addition to the pioneering work by Gougenheim et al. (1956, 1964), we should refer to the studies by Mackey, (1971), Fall (1976), Njock (1978), Gontier (1986),
Santos-Díaz, I. C.
The influence of reading on the vocabulary in mother tongue and foreign language


The first study on available vocabulary that includes a variable referring to reading was carried out by Mackey (1971), which analyses the readings made by the informants in Canada regarding the vocabulary provided. The next study, performed by Ávila-Muñoz (2007), links the number of books read by the informants each year to the spelling mistakes made in their mother tongue in an exploratory study with second-year students of the Degree in Translation at the University of Malaga. Ávila-Muñoz (2007, p. 45) concludes that the higher the socio-economic status of the informant is and the more books a year he/she reads, the fewer spelling mistakes he/she makes in the lexical availability test.

Ávila-Muñoz & Villena-Ponsoda (2010) carried out a study of the vocabulary available on a representative sample in the city of Malaga, with a uniform allocation of 72 speakers stratified by age, sex and level of education. They analyse the impact of reading frequency on the vocabulary repertory but, unlike the previous study by Ávila-Muñoz (2007), reading is not a significant variable regarding lexical availability.

Riffo-Ocares, Reyes-Reyes, Novoa-Lagos, Véliz-de-Vos & Castro-Yáñez (2014) analyse the reading comprehension, the lexical competence and the academic performance of 32 students in their last year of high school in Chile. The results show that a greater lexical knowledge has an impact both on their reading comprehension and on their overall academic performance. In the same spirit, Cepeda et al. (2017) evaluated the reading comprehension through a PISA standardised test and the lexical availability on six centres of interest of 9 first-year students from a rural technical professional context. They concluded that those students with more available vocabulary were those who obtained better results in the reading comprehension tests.

After a short bibliographical review, they found that the variable referring to reading frequency, understood as the number of books the informants read each year, is not present or plays a secondary role. On the other hand, in this article, reading is seen as the independent variable that helps to explain the behaviour of the available vocabulary and to identify specialised terms.

**Method**

**Objectives and assumptions**

This study aims at evaluating the influence of an informant’s reading frequency on the amount of words he/she provides in the available vocabulary test in Spanish and in a foreign language and on his/her ability to link technical terms and the subject they refer to. On the one hand, the assumption is made that reading contributes to improving the receptive knowledge of vocabulary. It is also expected that those students who read more are able to recognize specialized terms and link them to the appropriate centre of interest more easily because in previous studies (Cepeda et al. 2017; Riffo-Ocares et al. 2014) those readers with a richer vocabulary were also those who had a greater reading comprehension.

On the other hand, the nature of the language –mother tongue or foreign language– is differentiated. This way, a greater influence of reading in a foreign language compared to reading in their mother tongue is found, because the informants have spent years reading different types of texts in Spanish but not in the foreign languages analysed.

**Data collection and processing**

The sample of the study is made up by students of the Master’s Degree for Teachers of Secondary Education and “A” Levels, Vocational Training and Languages in the academic year 2010/2011 of the University of Malaga. Of a total of 385 students enrolled, 171 students from ten majors3 participated in the study, 117 were women and 54 were men. The informants’ age ranges from 22 to 49 and the average is 28.73 years old.
The data of the available vocabulary were collected and lemmatized following the recommendations of the Pan-Hispanic Project of Lexical Availability contained in the work by Samper-Padilla (1998). Since the objective of this study is to know the available vocabulary in Spanish as their mother tongue and in a foreign language (English or French), nine centres were chosen so that the test lasted for a similar time than in most studies. A new centre of interest related to new technologies was included because it is interesting from the point of view of teaching, and eight traditional centres of interest to compare between samples: parts of the body; clothes; foods and drinks; the school; furniture and materials; the city; means of transport; games and leisure; jobs and professions; computers and the Internet.

In order to complete the results of the available vocabulary, a specific test focusing on identifying specialized terms was performed, using the previous research performed by Ávila-Muñoz & Villena-Ponsoda (2010) as a reference. The test consisted on allocating two terms per centre of interest, totalling eighteen specialized terms per language of the test. For example, the technical terms of the topic “city” they had to link were “conurbation” and “megalopolis”.

The results of the specialized vocabulary test in the study by Ávila-Muñoz & Villena-Ponsoda (2010, p. 217) were shown through the Pident defined as the individual's ability to identify specialised lexical units. This rate is obtained by adding all the correct answers given by the informants, regardless of the test's implicit chance. In order to correct the technicism identification rate, in this study the incorrect answers were deducted from the addition of correct answers. The former were assigned a value of 0.56 and the latter 0.7, calculated depending on probability distribution.

Reading frequency was measured according to four intervals: informants who do not read any book during the year (0); some books (1); more than five books (2); more than ten books (3) and more than twenty books (4). This variable differentiates between books in Spanish or in their mother tongue (English or French, depending on the option picked by each informant).

**Results**

In first place, the overall results showing the number of books read by the students each year are presented; this variable shall hereinafter be called «reading frequency». The average of words provided by the informants depending on the number of books they read each year is then presented, together with the bivariate correlation. In last place, the relation between the ability to identify technicisms compared to the reading frequency will be analysed.

**General data**

Graph 1 shows that only 2.34% of the informants assures he/she does not read any book at all during the year. Nevertheless, most of them assure they read some books (between one and five), which corresponds to 53.80% of the population. 28.07% assure they read between six and ten books, 10.53% between eleven and nineteen and 5.25% more than twenty.

Graph 1. Books in Spanish read by the informant during the year
Source: Prepared by the authors

In general terms, the number of books written in a foreign language read by the students each year drops considerably. Graph 2 shows that 47.95% of the informants assure that they do not read any book during the year and 46.78%
Influence of reading frequency on the available vocabulary

This section aims at showing the relation between the number of words provided by the informants in Spanish and in a foreign language compared to the reading frequency during the year. Graph 4 shows that the average number of lexical units in Spanish is more homogeneous in the five groups reading frequency is divided into. Those informants who do not read any book during the year provided 188 words, those who read some books provided 189.09 words, those who read more than five books provided 193.4 words, those who read more than ten books provided 202.72 words and those who read more than twenty books provided 203.21 words.

In the case of foreign languages, the number of average words increases considerably as the number of books read by each informant during the year increases. Therefore, those informants who assure they do not read any book only provided 117.84 words, while those who read up to five books provided an average of 145.59 words. Nevertheless, these results become more homogeneous among those students who assure they read more than five and ten books, who provide an average of 176.2 and 177.1 words, respectively.

In English, the average number of words provided amounts to 131.3. In the case of those informants who do not read any book, this number drops to 116.69 words; in the case of those who read some books increases to 142.63 and in the case of those who read more than five and ten books the average number is similar, 178.33 and 180.5 words more specifically. In French, the average number of words is greater, amounting to 150.52 words. There is a big difference between those informants who do not read any books, with 130.14 words, compared to those who read some books, with 150.52 words, and those who read more than five books, with 159 words.
In order to compare from a statistical point of view the assumption that the more books readers read the broader their available vocabulary is, a bivariate correlation analysis is performed. In the case of Spanish, this correlation is not significant and has a Pearson correlation coefficient of 0.078. As far as foreign languages are concerned, this correlation is significant at the 0.01 level, the Pearson correlation coefficient being 0.428. This correlation is even more significant in the case of English, with a correlation coefficient of 0.443 and a significance level of 0.01, while in the case of French this coefficient falls to 0.308 and the relation is not significant. Since we are dealing with postgraduate students, it is assumed that all of them have sufficient command of Spanish that is influenced by reading frequency. In the case of English, reading frequency has greater influence and conditions the average number of words provided by the informants to a certain extent.

Influence of reading frequency on the ability to identify technicisms

This analysis is based on the assumption that informants can broaden their vocabulary repertory -notably their passive vocabulary- through reading. In first place, it was found that the sample's distribution is homogeneous and that the average rate increases when it comes to identifying technicisms compared to the number of books read by each informant during the year both in Spanish and in a foreign language.

The average corrected rate when it comes to identifying technicisms by those readers who do not read any book in Spanish is the lowest one (2.87), followed by those readers who read up to five books (4.54). It is also found that those informants who read more than five books during the year have a similar specialised vocabulary, ranging from 5.88 to 6.40.

The bivariate correlation was calculated below in order to know if there is any linear relation between both variables. The Pearson correlation coefficient is 0.162 and the significance level is below 0.05, 0.035 more specifically. On the other hand, the variation test shows that the populations are very similar, since F value is 1.578 at 0.182 significance level.

The relation between the number of books read by each informant during the year compared to the rate that measures the ability to identify technicisms (ICIT, as per its Spanish acronym) follows a similar pattern both for foreign languages and the mother tongue, as graph 5 shows. Those informants who do not read any book have the lowest rate, 2.47 more specifically; in the case of those who read some

Graph 4. Average number of available words depending on reading frequency in Spanish, English and French
Source: Prepared by the authors

Graph 5. ICIT average depending on reading frequency in Spanish, English and French
Source: Prepared by the authors
books, it amounts to 3.59; in the case of those who read more than five books, it amounts to 2.65 and, in last place, those who read more than ten books have the highest rate, 5.25 more specifically. As far as English is concerned, those informants who do not read any book have a rate amounting to 2.28; in the case of those who read some books it amounts to 3.27; in the case of those who read more than five books it amounts to 3.36 and those who read more than ten books have a rate amounting to 5.25. In French, the results are higher and more homogeneous than in English, probably because it is another Romance language having words that are more similar to those in Spanish. Those informants who do not read any book have a rate amounting to 3.51; in the case of those who read some books it amounts to 4.76 and in the case of those who read more than five books, to 4.89.

The analysis of the bivariate correlation between reading frequency and the ability to identify technicisms (ICIT) in a foreign language shows that it is significant at 0.01 bilateral. The Pearson correlation coefficient is 0.314 and the relation is thus positive: the higher the number of books read by each student during the year is, the greater his/her ability to identify technicisms is. On the other hand, that relation is maintained only in English, with a coefficient above 0.379 and a significance level amounting to 0.01, while in French the Pearson correlation coefficient amounts to 0.188 only and the relation is significant at 0.05 level.

Conclusions

The results of the study show that the correlation between active and passive vocabulary and reading frequency is higher in the case of foreign languages than in Spanish. In Spanish, the average number of words provided in the available vocabulary test does not show any significant relations from a statistical point of view because we are dealing with postgraduate students who have a similar level of education and this thus somehow standardizes the active vocabulary. On the other hand, the variable referred to the ability to identify vocabulary has a significant relation, with a Pearson correlation coefficient of 0.162 and a significance level below 0.05. This shows that the more books students read the greater their ability to identify technicisms is.

Regarding foreign languages, a distinction must be made between English and French because the results are different. The relation between the results of the vocabulary tests and reading frequency in French and Spanish are similar, while in English there is greater influence of reading frequency on linguistic variables. Regarding the available vocabulary, the average number of words in English increases as reading frequency increases and the relation between both variables is significant, while it is not in French. Regarding identifying technicisms, the results in French and Spanish are very similar, with a significant relation at 0.5 level. There is greater correlation in English than in Spanish and French, with a significance level also below 0.01.

It was found that university studies homogenise the vocabulary in Spanish and that the impact of reading frequency is secondary. On the contrary, as far as foreign languages—notably English—are concerned, the number of books read by the informants during the year have greater importance. Although all the informants have a B1 level (independent user level) according to the Common European Framework of Reference for Language, those students who provided more words in foreign languages have pursued university studies in this field (Philology or Translating and Interpreting) and they are also those who assure to read more books during the year.

In short, this study shows the influence of reading on vocabulary, both in the mother tongue and in foreign languages. Using a bivariate correlation, it has been demonstrated that those readers who read more often in their mother tongue also read more in a foreign language. This way, it is found that reading makes updating a greater number of words
easier, as well as identifying technicisms in a foreign language. It therefore contributes to improving the command on lexical competence.

Reading could be analysed in greater depth in further research. Despite the fact that many studies highlight the importance of vocabulary repertory in reading comprehension (Galloso-Camacho, 2004) and in the overall learning process (Lin, 2012), only a few studies on available vocabulary analyse the influence of reading. This variable could be approached in greater detail based on the way to measure the reading behaviour proposed by Yubero & Larrañaga (2010) through the number of books read during the last year, the number of compulsory readings at school and the number of books read during vacation periods. Similarly, the relation between pleasure in reading and the results obtained in the vocabulary tests could also be analysed because, as demonstrated by Molina-Villaseñor (2006), pleasure in reading is even associated to the students’ academic performance.

Lastly, these results initiate new discussion threads about the importance of reading in the mother tongue when learning a foreign language. It is also intended to create a precedent in the studies on lexical availability by analysing in detail the influence of a variable that has been very rare until now, such as the informants’ reading regularity. Therefore, this means a strengthening of the role played by available vocabulary in applied linguistics to language teaching, attached by their precursors Gougenheim et al. (1956).

Notes
2 http://www.dispolex.com/info/el-proyecto-panhispanico

3 Foreign Language: English; Foreign Language: French; Language and Literature. Latin and Greek; Geography and History; Technology, Computer Science and Industrial Processes (TIPI, as per its Spanish acronym); Mathematics; Drawing, Image and Visual Arts (DIAP, as per its Spanish acronym); Music, Educational Guidance; Job Training and Guidance (FOL, as per its Spanish acronym); Hotels and Tourism.

References
to stories. *Reading in a Foreign Language*, 20(2), 136-163.


Yubero, S., & Larrañaga, E. (2010). El valor de la lectura en relación con el comportamiento lector. Un estudio sobre los hábitos lectores y el estilo de vida en niños. Ocnos, 6, 7-20. doi: http://dx.doi.org/10.18239/ocnos_2010.06.01
