Lexical Availability and L2 Vocabulary Acquisition

Marjana Šifrar Kalan

University of Ljubljana, Slovenia

Abstract

Vocabulary research has followed a different path in English and in Spanish applied linguistics. Spanish applied linguistics has paid more attention to available lexicons of speakers than to word frequency. The measure of lexical availability combines the frequency at which a word is produced as a member of a semantic category (e.g. *dog* in category *Animals*) and the position in the list of associations provided by a group of individuals. It focuses on the words retrieved by speakers in response prompts (word stimulus) related to daily situations.

This paper intends to present some of the aspects of lexical-availability research that are interesting for L2 vocabulary acquisition. It attempts to show the potential of lexical-availability research as an alternate approach for vocabulary planning (the use of L1 lexical-availability measures to select the teaching vocabulary for L2) as well as the study of some psycholinguistic aspects of vocabulary acquisition, such as the organization of learners' mental lexicons, the similarities and the differences between response patterns, the kinds of semantic associations that learners activate in response to prompts (semantic categories), the consideration of the most available words obtained by lexical-availability research as semantic prototypes. Likewise, the study of learners' lexical availability can uncover sociolinguistic and cultural issues. Furthermore, this paper wishes to inspire researchers of languages other than Spanish to apply this methodology to different languages.

All these aspects are hereby presented on the basis of the Slovene learners' available lexicons in Spanish as L2 (N=200) (Šifrar Kalan, 2009; 2012; 2014b) and English as L2 (N=20) (Šifrar Kalan, 2014a).

Keywords: foreign languages, vocabulary acquisition, lexical availability, word associations

Lexical availability

Vocabulary research has followed different paths in English and Spanish applied linguistics. In English applied linguistics, developing word frequency lists from corpora, elaborating dictionaries containing frequency data, designing vocabulary tests and above all, compiling corpora and using it for vocabulary research have been the predominant research concerns. By contrast, Spanish applied linguistics has paid more attention to the available lexicons of speakers than to word frequency, although we could not say that the former has been totally neglected in the recent decades, because the new technologies have enabled the creation of numerous dictionaries and corpora of Spanish language (for more information see Almela et al, 2005; Lavid, 2005). The creation of a PanHispanic dictionary on the basis of available lexicons of speakers from different Spanish-speaking countries and regions has been one of the major projects in Spanish applied linguistics in the last two decades. This paper attempts to present some of the aspects of lexical-availability research that are interesting for L2 vocabulary acquisition and to inspire researchers of languages other than Spanish to apply this methodology to different languages.

Studies on lexical availability have a tradition in the Hispanic world since the 1970's when López Morales conducted the first investigation in Spanish as L1 in Puerto Rico. But the studies began in France in 1950s with the aim of selecting vocabulary for teaching French (in former French colonies) as a complementary approach to basic vocabulary, which at that time meant the most frequent words. When it was observed that some words, well known and used by French speakers, the words with specific semantic content did not appear in the frequency list, the concepts such as *frequent*, *basic* and *usual* vocabulary started to be defined as different notions. It had become clear that some words regarded as common or everyday words were not actually frequent and that their use was conditioned by the discourse theme; they were therefore called thematic words. On the other hand, certain words would almost always appear, regardless of the theme, the so-called nonthematic words. (López Morales, 2014, p. 2) The idea of artificial gathering of associations through word cues known as centres of interest (*Parts of the body, Clothes,* etc.)[1] was borrowed from the empirical psychology of the time. "Lexical availability came to be understood as the vocabulary flow *usable* in a given communicative situation. Behind this concept lies the belief that the mental lexicon includes words that are not realised in practice unless they are needed to communicate specific information. Such words make up the available lexicon." (López Morales, 2014, p. 3) French took the lead in lexical-availability research for several years, both in France and in Canada, but in 1969 the Yugoslav Naum Dimitrijević published the results of his lexical-availability study carried out in English as L1 among Scottish school students. He used open lists for 11 semantic fields, as opposed to former closed list of 20 associations. The same technique was followed closely by American linguist

Bailey (1971), who compared available lexicons of monolinguals and bilinguals of English and Spanish.

During the last two decades, lexical-availability studies have focused almost exclusively on Spanish, mainly as studies of Spanish as L1 within the PanHispanic project, but also with the students of Spanish as L2 in Finland (Carcedo González, 2000), Poland (López González, 2010), China (Jing, 2012), Island (Magnúsdóttir, 2012), Turkey (González Fernández, 2013) and Slovenia (Šifrar Kalan, 2009, 2012, 2014b). Similar studies have been carried out with foreign students in Spain (Samper Hernández, 2002; Sánchez Gómez, 2005; López Rivero, 2008; Pérez Serrano, 2009; Fernández-Merino Gutiérrez, 2011; Sánchez-Saus Laserna, 2011; Jiménez Berrio, 2013 as cited in Šifrar Kalan, 2014b: 64). Some studies have been conducted on English as L1 or English as L2 (Jiménez Catalán, 2014).

The lexical-availability studies are being renewed constantly and will probably trigger new interdisciplinary studies in addition to the main disciplines - sociolinguistics, dialectology, psycholinguistics and ethnolinguistics.

Word frequency versus lexical availability

Both word frequency and lexical availability deal with the assessment of vocabulary knowledge. In the former, the reference is always the written and/or oral text and in the latter, it is the speaker's mental lexicon. In the first case, the words are actually used; in the latter, the words would be hypothetically used in connection with a certain topic. Obviously, the frequency counts as well as the available lexicons have its limitations. A frequency count is only as good as the corpus it is based upon, and every corpus has limitations. No corpus can truly mirror the experience of an individual person and there are some language types that are very difficult to collect. On the other hand, the available lexicons are always based on individual associations that are calculated according to the frequency and the order of appearance of responses to make collective lists. Reliability is questionable because the availability tests would show different results on each occasion for it is not only a linguistic but also a cognitive activity. Corpus presents the most frequent words, which are usually those with grammatical functions; meanwhile, the most available words are usually those that carry content or meaning. A similar distinction between function and content vocabulary has been pointed out by Schmitt (2010, p. 54): "Corpus word counts consistently show that function words are among the most frequent in language, which is not surprising because they are necessary for communicating about any topic, from daily life to astrophysics. This holds true regardless of whether the discourse is general in nature, technical, or academic." All this speaks in favour of the idea that these two methodologies are complementary.

There are many limitations in both measurements of vocabulary, but this should not invalidate the arguments in favour of vocabulary research from these two different methodologies. The concept of word frequency is based on some very basic and important assumptions. First, the most important group of words is the high frequency words of the language. Most lists of high frequency words consist of around 2,000 word families. This number has been generally accepted as high-frequency vocabulary (Nation, 2008, p. 7; Schmitt, 2010, p. 69): "In most texts around 80% or more of the running words are from the most frequent 2000 words of English. In friendly conversation, over 90% of the running words tend to be from the high frequency words of English." (Nation, 2008, p. 8). But Schmitt (2010, p. 69) states that these traditional frequency levels have been called into question by Nation's recent research and that they will need to be reappraised. Second, the most frequent words are acquired before the least frequent words. A greater knowledge of infrequent words is related to lexical richness and therefore higher lexical competence, more comprehension and more language production. (Nation & Waring, 1997). In contrast, "lexical availability research focuses on the words retrieved by speakers in response to prompts related to daily situations; particularly it focuses on an analysis of the positions of words in the ranking of elicited responses." (Jiménez Catalán & Fitzpatrick, 2014, p. 85). The lexical-availability research in L1 and L2 has shown that the first word associations or responses are the most available in the speaker's mental lexicon, but all responses reflect the organization of the speaker's or learner's mental lexicon.

Can these two methodologies be studied in tandem? This novel complementary approach has been proposed by Jiménez Catalan and Fitzpatrick (2014) by applying a word-frequency framework to data produced in English L2 lexical-availability studies. By means of *Lexical Frequency Profile* designed by Laufer and Nation (as cited in Jiménez Catalán and Fitzpatrick, 2014) they measure lexical richness in the words retrieved by 50 Spanish sixth- and eighth-grade learners of English as L2 in response to nine prompts (semantic categories). One of their research hypotheses was that a lexical-availability frequency profile reveals qualitative differences in the output of English L2 learners in the sixth and eighth grades. An increase of low-frequency words would be expected as learners advance in language level, but the results of their study showed that "this increase in learners' word types does not result automatically in a more advanced frequency profile" (2014, p. 96) although the study has also shown that the "EFL learners' lexical availability increases as the course grade increases." (2014, p. 98) According to the authors. a possible interpretation of these results is that vocabulary still has room to grow in the most frequent bands. (2014, p. 97). This study indicates a great potential for further comparative research of lexical availability and word frequency.

Vocabulary selection and lexical availability

Another important area in which word frequency should be complemented with available lexicons is vocabulary selection. There is no doubt that "frequency has long informed the principled selection of vocabulary in L2 teaching pedagogy" (Schmitt & Schmitt, 2012, p. 1). For a long period, 2,000 word families have seemed to be the most cited initial goal for foreign language learners, but Norbert and Diane Schmitt (2012) have now proposed an increase of high-frequency English vocabulary to 3,000 word families: "We suggest that, as a minimum, English language programs emphasize teaching of high-frequency vocabulary up to the 3,000 frequency level." (2012, p. 15) However, frequency is not the only criterion for choosing words to teach explicitly. Another criterion is the words that are particularly useful in a specific topic area, the so-called technical vocabulary recommended to be learned after having mastered the foundation of 5,000 word families. The third category is the words that students want to learn for various reasons. The fourth category, especially important at the beginning of the course or language study, is the classroom management vocabulary. (Schmitt, 2000, p. 144) To these four categories Schmitt and Schmitt (2012) later add the importance of teaching the mid-frequency vocabulary for proficient language use.

In the field of lexical-availability research a slightly different approach was established on the assumption that "the fundamental vocabulary of a given community consists of basic lexicon and the available lexicon. The identification of this available lexicon is an essential underpinning for any planning related to the lexicon" (López Morales, 2014, p. 7). From this perspective, the frequency words, among which the most frequent are the grammatical words, are completed with specific thematic words needed to address certain themes in daily life. The studies of available lexicons among native speakers provide us with vocabulary that they would potentially use in connection with a certain topic. Consequently these are the words that foreign speakers should also know. For example, if in the category "Food and drink", one of the most frequent words and first associations of Spanish native speakers is *garbanzo* (chickpea) and *lenteja* (lentil), we should include these words in explicit teaching regardless of their position in frequency list.[2] These two words are not included among the 10,000 most frequent words in the Corpus of Royal Spanish Academy. Another example to illustrate the necessity of taking into account the criterion of available lexicon is taken from the category of "Professions and jobs": *ganadero* (rancher) and *barrendero* (sweeper). The latter is not listed in the first 10,000 Spanish word frequency list (CREA); and *ganadero* occupies the 8,566th position on the same list, while they are both among the most available words in available Spanish lexicons. These two occupations are also very rarely presented in Spanish L2 textbooks. How can a learner get to know these words if they are neither included in the category of frequent words nor in the specific topic category, and are

not the words that a learner specifically wishes to learn? In this case, the category of available lexicons should be applied. Very often the most available words among native speakers reflect customs and habits. Likewise, the study of learners' lexical availability can uncover many interesting sociolinguistic and cultural issues.

Psycholinguistic aspects of lexical availability

The studies of lexical availability present an important tool for psycholinguistics because they clearly capture the relationships the speakers establish between lexical units. The present paper addresses some issues of the L2 mental lexicon based on the results of the lexical-availability research carried out among Slovene secondary (N=100) and university students (N=100) (Šifrar Kalan, 2009; 2012; 2014b) as well as English as L2 (N=20) (Šifrar Kalan, 2014a). The results of these studies prove that more advanced learners of Spanish and English produce a higher number of association responses to specific domains, but still fall behind the native speakers' production of associations. Similar conclusions were made in other L2 studies (Schmitt, 2000, p. 42). The studies of Slovene students also show that the most available words, both in Spanish and English, and with different levels of proficiency, correspond to the more typical examples of the category. "This tendency seems to confirm the universality of semantic prototypes based on human experience regardless of the language as advocated by Aitchison (1994) and Kleiber (1995)." (Šifrar Kalan, 2014a, p. 134) These studies also confirm the syntagmatic-paradigmatic shift as a person's language matures (Schmitt, 2010, p. 40): Slovene B1 (CEFR level) Spanish students produced more syntagmatic associations than B2 students, and while B2 students produced more paradigmatic associations. The study of individual-association chains of B1 and B2 students shows that words are mainly related by meaning. Nevertheless, there were some form-based responses noted as well. According to these results we agree with Singleton's claim (1999, p. 189) that in L1 and L2 lexical units are increasingly processed by meaning rather than by form as their integration into the mental lexicon progresses. Aitchison (1994) lists three basic findings regarding associations that can offer important insight into the mental lexicon. These findings strongly correspond to Slovene studies. First, the responses are almost always items from the same semantic field, which speaks in favour of creating dictionaries of available lexicons on the basis of semantic domains. Second, adults usually give a response that is the same word class as the prompt word, which is the reason for noun superiority in available lexicons. The introduction of new, experimental semantic category entitled "Actions Carried Out Every Day" (Šifrar Kalan, 2014b), on the other hand, resulted in 72% of verbs. Third, if a stimulus word is part of an obvious pair, the partner word is usually given as the response. The prevalent association pair in our results is in a coordinate relation within the same word class, followed by synonyms, hyperonyms and antonyms, respectively.

Undoubtedly, these findings help us understand the organization of mental lexicons, yet there are still many things to be discovered. "It is important that future studies investigate the similarities as well as the differences between L1 and L2 response patterns, and the differences and similarities within each subject group." (Fitzpatrick as cited in Schmitt, 2010, p. 62)

Conclusion

Originally, lexical-availability research was regarded as an alternative approach for vocabulary selection for teaching French as L2. Some years later it was adopted by Hispanic linguists in order to create a great PanHispanic dictionary of available lexicon. During the years of collecting data for the available lexicons in Spain and Latin America, the linguists started to use the data in L1 for research in sociolingustics and dialectology. During the last two decades the research has refocused on foreign language teaching, namely to vocabulary teaching and research. The available lexicons provide an important tool for vocabulary planning and material designing and should be used together with other tools, such as frequency lists. Likewise, the association responses offer a great potential for research in psycholinguistics. Exploring lexical availability in L1 and L2 in combination with other study areas, such as, vocabulary tests, word frequency, CEFR levels and others, can provide us with valuable data concerning vocabulary acquisition in the future.

1 Traditionally 16 categories or centres of interest are used in lexical-availability studies: *Parts of the human body, Clothing, Parts of the house, House furniture, Food and drink, Objects on the table for the meal, The kitchen and its utensils, School furniture and materials, Heating and lighting, The city, The countryside, Means of transport, Farm and garden work, Animals, Games and entertainment, Jobs and professions.* These categories were already chosen by Gougenheim, a French pioneer of lexical-availability study.

2 The Spanish words *garbanzo* and *lenteja* are among the most available words in the available lexicons of different regions of Spain. For example, *garbanzo* is in Position 6 in Asturias's lexicon, Position 10 in Cádiz and 45 in Valencia. *Lenteja* is in Position 4 in Asturias, 6 in Cádiz and 26 in Valencia. Meanwhile, not even one out of 200 Slovene students of Spanish who participated in the study of lexical-availability failed to write the word *lenteja* as a response. *Garbanzo* is among the available words for Slovene students, but with a very low index of availability. (Šifrar Kalan, 2012)

References

Aitchison, J. (1994). Words in the Mind: An Introduction to the Mental Lexicon. Oxford - New York: Basil Blackwell.

Almela, R., Cantos, P., Sánchez, A., Sarmiento, R., Almela, M. (2005). Frecuencias del español: Diccionario y estudios léxicos y morfológicos. Madrid: Editorial Universitas.

Bailey Victery, J. (1971). A study of lexical availability among monolingual-bilingual speakers of Spanish and English (master thesis). Houston: Rice University.

Carcedo González, A. (2000). Disponibilidad léxica en español como lengua extranjera: el caso finlandés (estudio del nivel preuniversitario y cotejo con tres fases de adquisición). Turku: Universidad de Turku.

González Fernández, J. (2013, August 28). La disponibilidad léxica de los estudiantes turcos de español como lengua extranjera. MarcoELE, 16. Retrieved from http://marcoele.com/disponibilidad-lexica-de-estudiantes-turcos/.

Jiménez Catalán, R. M. (Ed.) (2014). Lexical Availability in English and Spanish as a Second Language. Dordrecht: Springer.

Jiménez Catalán, R.M. & Fitzpatrick, T. (2014). Frequency Profiles of EFL Learners' Lexical Availability. In Jiménez Catalán, R. M. (Ed.), Lexical Availability in English and Spanish as a Second Language (pp. 83-100). Dordrecht: Springer.

Jing, L. (2012, August 28). El estudio de disponibilidad léxica de los estudiantes chinos de español como lengua extranjera. MarcoELE, 14. Retrieved from http://marcoele.com/disponibilidad-lexica-de-estudiantes-chinos/.

Lavid, J. (2005). Lenguaje y nuevas tecnologías: Nuevas perspectivas, métodos y herramientas para el lingüista del siglo XXI. Madrid: Cátedra.

López González, A. M. (2010, August 28). La evaluación del desarrollo de la competencia léxica en L2 por medio de la disponibilidad léxica. Redele, 18. Retrieved from http://www.mecd.gob.es/redele/revistaRedEle/2010/primera.html.

López Morales, H. (2014). Lexical Availability Studies. In Jiménez Catalán, R. M. (Ed.), Lexical Availability in English and Spanish as a Second Language (pp. 1-11). Dordrecht: Springer.

Magnúsdóttir, S. (2012). Disponibilidad léxica en alumnos de español como lengua extranjera. Estudio sobre el léxico disponible en alumnos de ELE en la secundaria en Islandia. Háskóli íslands: Sigillum Universitatis Islandiae. Unpublished master thesis.

Nation, I.S.P. (2008). Teaching Vocabulary: Strategies and Techniques. Boston: Heinle.

Nation, I.S.P. & Waring, R. (1997). Vocabulary seize, text coverage, and word lists. In Schmitt, N. & McCarthy, M. (Eds.), Vocabulary: Description, Acquisition, and Pedagogy (pp. 6-19). Cambridge: Cambridge University Press.

Real Academia Española. Banco de datos (CREA). Corpus de Referencia del Español Actual. Retrieved from http://corpus.rae.es/lfrecuencias.html

Schmitt, N. (2010). Researching vocabulary. A Vocabulary Research Manual. Hampshire: Palgrave Macmillan.

Schmitt, N. & Schmitt, D. (2012). A reassessment of frequency and vocabulary size in L2 vocabulary teaching. Language Teaching, available on CJO2012. doi:10.1017/S0261444812000018.

Schmitt, N. (2000). Vocabulary in Language Teaching. Cambridge: Cambridge University Press.

Singleton, D. (1999). Exploring the Second Language Mental Lexicon. Cambridge: Cambridge University Press.

Šifrar Kalan, M. (2009). Disponibilidad léxica en español como lengua extranjera: el cotejo de las investigaciones en Eslovenia, Salamanca y Finlandia. Verba Hispánica, 17, 165–182.

Šifrar Kalan, M. (2012, September 13). Análisis comparativo de la disponibilidad léxica en español como lengua extranjera (ELE) y lengua materna (ELM). Marco ELE, 15. Retrieved from http://marcoele.com/descargas/15/sifrar-disponibilidad_lexica.pdf.

Šifrar Kalan, M. (2014a). Slovene Student's Lexical Availability in English and Spanish. In Jiménez Catalán, R. M. (Ed.), Lexical Availability in English and Spanish as a Second Language (pp. 125-138). Dordrecht: Springer.

Šifrar Kalan, M. (2014b). Disponibilidad léxica en diferentes niveles de español/lengua extranjera. Studia Romanica Posnaniensia, 41, 63-85.