Educational Linguistics

Rosa María Jiménez Catalán *Editor*

Lexical Availability in English and Spanish as a Second Language

Springer

Lexical Availability in English and Spanish as a Second Language

Educational Linguistics

Volume 17

Educational Linguistics is dedicated to innovative studies of language use and language learning. The series is based on the idea that there is a need for studies that break barriers. Accordingly, it provides a space for research that crosses traditional disciplinary, theoretical, and/or methodological boundaries in ways that advance knowledge about language (in) education. The series focuses on critical and contextualized work that offers alternatives to current approaches as well as practical, substantive ways forward. Contributions explore the dynamic and multi-layered nature of theorypractice relationships, creative applications of linguistic and symbolic resources, individual and societal considerations, and diverse social spaces related to language learning.

The series publishes in-depth studies of educational innovation in contexts throughout the world: issues of linguistic equity and diversity; educational language policy; revalorization of indigenous languages; socially responsible (additional) language teaching; language assessment; first- and additional language literacy; language teacher education; language development and socialization in non-traditional settings; the integration of language across academic subjects; language and technology; and other relevant topics.

The *Educational Linguistics* series invites authors to contact the general editor with suggestions and/or proposals for new monographs or edited volumes. For more information, please contact the publishing editor: Jolanda Voogd, Asssociate Publishing Editor, Springer, Van Godewijckstraat 30, 3300 AA Dordrecht, The Netherlands.

For further volumes:
http://www.springer.com/series/5894

Rosa María Jiménez Catalán
Editor

Lexical Availability in English and Spanish as a Second Language

Springer

Editor
Rosa María Jiménez Catalán
Departamento de Filologías Modernas
Facultad de Letras y Educación
Universidad de La Rioja
Logroño, La Rioja, Spain

ISSN 1572-0292
ISBN 978-94-007-7157-4 ISBN 978-94-007-7158-1 (eBook)
DOI 10.1007/978-94-007-7158-1
Springer Dordrecht Heidelberg New York London

Library of Congress Control Number: 2013947466

Printed on acid-free paper

Springer is part of Springer Science+Business Media (www.springer.com)

Preface

Lexical availability is understood as the words that people have in their minds and that emerge in response to cue words that stand for domains closely related to daily life such as 'Food and drink', 'Animals', 'Politics', or 'Poverty'. Lexical availability is an important dimension of language learners' lexical competence, and in consequence, an essential variable of their communicative competence in the target language; however, in spite of its relevance, little research has been conducted on this issue in second or foreign language education, and practically nothing has been done in the field of vocabulary studies.

Vocabulary research has followed a different path in English applied linguistics and Spanish applied linguistics. In the former, the developing of word frequency lists from corpora, the building up of dictionaries, and the design of vocabulary tests aimed at language learners have been the predominant research concerns in the last two decades. In contrast, within the Spanish applied linguistics tradition, the principal focus has been the creation of a PanHispanic dictionary out of the available lexicons of speakers from different Spanish-speaking regions and countries. Under this influence, second language researchers have focused on the elicitation and description of learners' available lexicons rather than on word frequency. This book attempts to cross a bridge in these two traditions: it contains a collection of original studies written by lexical availability researchers within Spanish applied linguistics and vocabulary researchers within English applied linguists, two communities of practice with shared concerns, but that rarely meet in the same research forums, let alone in the space of a book.

Lexical availability has a great potential to explore psycholinguistic aspects of learners' vocabulary knowledge. Some of them are the study of the organization of learners' lexicon, the comparison of learners' available lexicons to that of native speakers', the nature of the words that learners activate in response to prompts, or the kind of semantic associations that emerge through the patterns of responses at different stages of vocabulary development and different levels of language proficiency. Likewise, the study of learners' lexical availability can provide vocabulary researchers with opportunities to investigate sociolinguistic and cultural issues such as the effect of age, gender, or ethnicity on the words learners retrieve in

response to prompts related to social or cultural issues. Last but not least, lexical availability tasks can be used in combination with other methodologies employed in vocabulary research as for instance corpus techniques and word frequency; in particular, the combination of methodologies has a great potential in the comparison of native speakers' and learners' available lexicons. It has also a great potential in the exploration of learners' lexical output as well as in the study of the vocabulary contained in language learners' course books, reading materials and vocabulary tests. Some of these paths are explored in the chapters included in this book.

The book is preceded by an opening chapter (Chap. 1) by Prof. Humberto López Morales, a narrative of the history of lexical availability studies by the founder of this tradition of studies in Spanish language. The chapter introduces terms, concepts and formulae that will appear later throughout the different chapters in the book. It also defines lexical availability, describes its origin in French applied linguistics, traces its subsequent development into the PanHispanic project, and ends with a summary of present themes and currents of research in lexical availability studies related to foreign language learning and teaching. This opening chapter serves as a framework for the rest of the book which is structured into two parts and a concluding chapter.

The chapters in Part I and Part II all contain empirical studies. The shared concern is lexical availability in second or foreign languages; the focus is on learners rather than on teaching or language teaching materials. Each part comprises research on lexical availability conducted from different perspectives such as sociolinguistics, cognitive psychology, corpus studies or word frequency studies; both parts include research on foreign language learners in primary, secondary and tertiary education, mainly in Spain but also in Chile, Poland and Slovenia. The two parts differ in the mother tongues and target languages observed: English as L1 and L2 in Part I and Spanish as L1 and L2 in Part II.

In Chap. 2, Roberto A. Ferreira Campos honours Prof. Max S. Echeverría Weasson (another great name in lexical availability studies who generously accepted an invitation to contribute to this volume but who sadly could not, as he died at the end of 2010). Ferreira looks at the performance of Chilean university students, advanced English (L2) learners in comparison with English native speakers (L1) in basic ('Body parts', 'Food and drink'), and advanced ('Terrorism and crime', 'Health and medicine') semantic categories. Not surprisingly, L1 speakers outperformed L2 advanced learners in all semantic categories. However, the most significant finding in this study is that both groups retrieved a greater number of words for basic semantic categories than for advanced semantic categories which seem to point to similar patterns in the organization of the available lexicons of L1 and L2 speakers.

In Chap. 3, Rosa María Jiménez Catalán, María Pilar Agustín Llach, Almudena Fernández Fontecha and Andrés Canga Alonso adopt a corpus methodology to compare the lexical availability output of sixth grade primary school children and first year university students, English language learners. The aim was to ascertain whether if, holding language level constant, children and adults would retrieve the same number of word responses as well as similar or different types of words.

The findings suggest the existence of similarities regarding the number of words retrieved by each prompt but also more differences than similarities regarding the specific words activated by the cue words. These results reveal the existence of exclusive vocabularies in the available lexicons of young and adult EFL learners of the same language level.

Age, together with previous exposure to English, is addressed by Francisco Gallardo del Puerto and María Martínez Adrián in Chap. 4. The authors looked at the effect of previous foreign language (English) contact on senior learners' (age 55+) performance in a lexical availability task including 15 prompts, traditionally used in lexical availability studies. The results showed that false beginners outperform true beginners both for the total number of words produced in the lexical availability task and for most of the semantic categories contained in the task. The authors argued that beginners experience similar stages in vocabulary acquisition as a striking similarity is found in the available lexicons of the groups of senior EFL learners in this study and that of young learners examined in other studies. Based on their findings, they also suggest that the ability of the older adult to learn new words is not impaired.

Chapter 5 by María Pilar Agustín Llach and Almudena Fernández Fontecha analyse the effect of gender on words retrieved by the same sample of EFL learners at two points of time: sixth grade and ninth grade. The prompts were: 'Body', 'Food', 'School', 'Town', 'Countryside', 'Transport', 'Animals', 'Sports', and 'Professions'. The study provides evidence of a significant increase of word responses in 9th grade for all cue words and for both groups. This result is relevant for research in lexical availability as well as in vocabulary research as it proves that learners continue learning words within each of the semantic categories represented by the cue words. The study is also relevant for sociolinguistic research on gender and language education as it reveals significant differences in favour of females in six prompts out of nine at the two collection times.

In Chap. 6, Rosa María Jiménez Catalán and Tess Fitzpatrick take a novel approach to the analysis of lexical availability output. They apply a word frequency framework to data produced by 6th and 8th EFL learners in response to nine cue words traditionally used in lexical availability studies. The chapter looks at learner profiles according to the number of words produced in the nine semantic domains, and the proportion of infrequent words to frequent words in each domain. The findings are relevant for lexical availability studies as they open a new line of research in the field. They are also relevant for vocabulary research as they question the assumption of a linear pattern of vocabulary acquisition through frequency bands.

Chapter 7 by Marta Samper Hernández opens Part II and is devoted to studies on lexical availability of learners of Spanish in different learning contexts. In a classical study under PanHispanic research, the author performs detailed descriptive analyses on the lexical availability output of Spanish foreign language learners. These were distributed on the basis of their language proficiency level on Spanish: basic and advanced. In her study, advanced learners produced a larger number of words than learners in the basic group, in practically all the cue words under examination. The exceptions were 'The City' and 'Games and Entertainment', where learners at

the basic level either outperformed learners at the advanced level or behaved in a similar way. Results for these cue words show that a higher language level does not always result in a higher number of words. Other factors such as the kind of instruction, course input, or learners' experiences should be taken into account.

In Chap. 8, Marjana Šifrar Kalan explores the differences and similarities in lexical availability in two foreign languages, English and Spanish. She compares eight semantic categories in a lexical availability task administered to Slovene students, learners of English and Spanish as foreign languages. She describes the most available words in learners' lexical production in these two languages and addresses issues such as prototypicality, language proficiency and years of study of Spanish and English. An important finding in this study is the similarity in the word responses provided by the two groups of language learners on the prompts representing semantic categories. The similarities in learners' responses points to the existence of semantic prototypes in Slovene students' minds, regardless of the target language, or at least as far as English and Spanish are concerned.

Chapter 9 by Antonio María López González compares two bilingual programs in secondary education in Poland. The author conducts a quantitative and qualitative analysis of the lexical availability output of Polish students, learners of Spanish as a foreign language in an intensive and an extensive bilingual program with a similar number of hours of instruction. In addition to providing insights into Polish Spanish learners' available lexicons, the findings of this chapter also have educational implications for bilingual planning as they prove the higher effectiveness of intensive programs over extensive programs.

In Chap. 10, Natividad Hernández Muñoz, Cristina Izura and Carmela Tomé explore cognitive factors influencing lexical availability in Spanish as L1 and L2. This is the first comparative study to date examining these aspects in lexical availability studies. Results showed that the availability of Spanish words, in L1 as much as in L2, is determined by the order at which words are learned and by their typicality. In addition, the degree of cognateness between words in the participant's L1 and L2 was a powerful determinant of lexical availability in L2. An important finding is that lexical availability in Spanish as L1 is not directly comparable with the lexical availability in Spanish as L2. The mere fact of knowing two languages changes the availability of the L2 words.

The concluding section (Chap. 11) by Marta Samper Hernández and Rosa María Jiménez Catalán attempts to unfold the characteristics shared by all the foregoing chapters and to clarify basic terms and concepts in lexical availability research.

This book will be useful for teachers and researchers of Spanish and English as foreign languages. It contains analyses of the words that learners of these languages know and are capable of retrieving when put in an appropriate situation. The lists of the most productive prompts uncover what learners know; but even more interesting are the words that do not appear on the lists since they reveal what learners do not know or are not capable of retrieving. In the same vein, the lists of the most productive prompts representing vocabulary domains are certainly useful, but even more useful are the lists derived from the least productive prompts. These reveal gaps in learners' vocabulary knowledge. Being informed on what learners know and

what they do not know regarding words is extremely important for language teachers but also important for language planning and the design of vocabulary activities for learning and teaching to foreign language learners. Last but not least, information on the words used by learners from different ages, gender, proficiency level, and different target languages such as English and Spanish can provide researchers with invaluable data to investigate the nature and organization of language learners' lexicons.

The editor of this book strongly believes in research as the road to understanding, and to the improvement of things by the application of knowledge. I believe that collaboration and sharing make up the essential luggage in this journey. Hopefully, this book will contribute somehow to narrow the gap between languages, methodologies and traditions: Spanish lexical availability studies and English vocabulary research, two separate research spaces that cast their eyes on the same reality – learners' vocabulary knowledge.

Logroño, Spain Rosa María Jiménez Catalán

Contents

Contributors

María del Pilar Agustín Llach Departamento de Filologías Modernas. Facultad de Letras y Educación, Universidad de La Rioja, Logroño, La Rioja, Spain

Andrés Canga Alonso Departamento de Filologías Modernas. Facultad de Letras y Educación, Universidad de La Rioja, Logroño, La Rioja, Spain

Almudena Fernández Fontecha Departamento de Filologías Modernas. Facultad de Letras y Educación, Universidad de La Rioja, Logroño, La Rioja, Spain

Roberto A. Ferreira Campos Departamento de Lenguas. Facultad de Educación, Universidad Católica de la Santísima Concepción, Concepción, Chile

Tess Fitzpatrick Cardiff School of English, Communication & Philosophy, Cardiff University, Cardiff, UK

Francisco Gallardo del Puerto Departamento de Filología, Universidad de Cantabria, Santander, Spain

Natividad Hernández Muñoz Departamento de Lengua Española, Universidad de Salamanca, Salamanca, Spain

Cristina Izura Department of Psychology, University of Swansea, Swansea, UK

Rosa María Jiménez Catalán Departamento de Filologías Modernas. Facultad de Letras y Educación, Universidad de La Rioja, Logroño, La Rioja, Spain

Humberto López Morales Secretary General of the Association of Academies of Spanish Language (Asociación de las Academias de la Lengua Española), Madrid, Spain

Antonio María López González Associate professor of Spanish linguistics, Katedra Filologii Hiszpanskiej, University of Lódz, Lódz, Poland

María Martínez Adrián Departamento de Filología Inglesa y Alemana. Facultad de Letras, University of the Basque Country, Donostia, Spain

Marta Samper Hernández Departamento de Filología Española. Clásica y Arabe, Universidad de Las Palmas de Gran Canaria, Las Palmas, Spain

Marjana Šifrar Kalan Department of Romance Languages and Literatures, Faculty of Arts, University of Ljubljana, Ljubljana, Slovenia

Carmela Tomé Departamento de Lengua Española, Universidad de Salamanca, Salamanca, Spain

Chapter 1
Lexical Availability Studies

Humberto López Morales

1.1 Introduction

Studies of lexical availability have more than 50 years of history behind them. They were born in France during the first phase of the elaboration of *Le Français Élémentaire*, published in 1954,[1] a work which grew out of a slightly earlier UNESCO initiative.[2]

The primary purpose had changed a lot in those years. Initially, the aim was to teach the French language (graded at several levels of difficulty) to the people that made up the federation of territories known as Union Française. Later, with most of these countries already converted most of them into independent nations, the original project was refocused on ensuring that the citizens of the former colonies, mainly in Africa and Asia, continued to keep bonds with Gallic language and culture. In short, in some of those countries French was maintained as the official language, in others, it was by far the most influential foreign language.

[1] The work was re-edited with several minor changes in 1959 by the new title of Français Fondamental (1er degré) and again, in what was the final Edition, L'élaboration du Français Fondamental (1er degré). Étude sur l'établissement d'un vocabulaire et d'une grammaire de base, in 1964: which I quote. The Institute, created by the French Government to carry out these works, also changed its name after not few heated discussions: from Centre d'Étude du Français Élémentaire to Centre de Recherche et d'étude pour la Difussion du Français, CREDIF.

[2] It is not without significance that the person who made this recommendation to the UNESCO was the representative of France in the International Committee of Linguistics, Profesor of the École Nationale des Langues Orientales Vivantes, M. Aurélien Sauageot, who would be co-author of this project later on.

H. López Morales (✉)
Secretary General of the Association of Academies of Spanish Language (Asociación de las Academias de la Lengua Española), Real Academia, C/Felipe IV,
Madrid, Spain
e-mail: hlopez@ra.es

R.M. Jiménez Catalán (ed.), *Lexical Availability in English and Spanish as a Second Language*, Educational Linguistics 17, DOI 10.1007/978-94-007-7158-1_1,

1.2 Frequent Lexicon, Available Lexicon

In the *Élémentaire* level great importance was attributed to vocabulary. Therefore, out of the vastness of the lexicon of French language, it was necessary to select the words that should be included in that level.[3] Given the state of lexical-statistical knowledge in the early 1950s, the selection criteria that seemed to be more meaningful was that of frequency: the most frequent words were the most useful and also, it was thought, the most used therefore, the ones that should be given priority.[4]

However, in the course of the work,[5] a problem aroused recurrently which needed urgent solution. Some words, well-known and used by French speakers, did not appear in the frequency lists. In other words, although grammatical words, verbs, adjectives or general nouns appeared in the frequency lists (*chose, homme, personne, enfant,* etc.), words whose semantic content was very specific did not (*dents, métro, roi*, etc.).

Concepts that up to that moment had been treated as synonyms -frequent vocabulary, basic vocabulary, and usual vocabulary- started to be defined as different notions. It had become clear that some words regarded as common, even usual were not actually frequent. This infrequency resulted from the fact that part of vocabulary, particularly nouns, was thematic; that is to say, their use was conditioned by the discourse theme.[6] Only if the theme was favorable would certain words be realized in conversation. On the other hand, certain words would almost always appear, regardless of the theme. These were, therefore, non thematic words.

The relationship between frequent vocabulary and non thematic words was soon established; frequency vocabularies included those words with the highest statistical indexes. Working with them involved the selection of a specific type of words, but left other words aside such as those needed to address certain themes in daily life.

---

[3] Fortunately, the idea of selecting the most usual words out of the lexicon repertoire of a common dictionary, for example, Le Petit Larousse was rapidly rejected. In principle, its 50,000 words (in the edition of the time) could be reduced to 8,000 or even 6,000, a very simplistic solution in the view of the authors.

[4] Unlike Basic English, the Français Fondamental envisaged the development of a 2e degré, out of which specialized vocabularies would be prepared, for instance, that of literary studies: Vocabulaire d'initiation à the critique et à l'explication littéraire by CREDIF.

[5] A detailed description of the processes followed to obtain the most frequent vocabulary can be found in the three chapters that make up the second part of L'élaboration du Français Fondamental. Specifically, in "La fréquence" (61–134).

[6] The theme in discourse or speech gives rise to more concrete specifications, such as: (1) the nature of the texts or conversations (historical works, fairy tales, journalistic texts), (2) the characteristics of the countries (the word *roi*, for example, will be more frequently used in countries with monarchies), and (3) the historical circumstances (the word *roi* recurrently appeared in the French press on the occasion of the death of King George V of England), Gougenheim et alii. (1964: 139–140).

Researching the frequent vocabulary consisted of compiling a representative corpus and converting it into electronic data: not a difficult task, even for the rudimentary methods of the time.[7] At least at a first level of analysis, the result was always a hierarchical list of words ranked on the basis of cumulative frequency. That soon proved to be an inappropriate methodology for identifying other lexical units that did not appear in their texts.

At this point, the idea of working with association tasks emerged. This was an artificial way of bringing to the surface the words available for immediate use by a given speaker, or a specific group of speakers. Michéa was the first to make a distinction between 'frequent words' and 'available words'. 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'; its study cannot be undertaken by means of frequency analysis since the 'available lexicon' is pertinent only in the case of actual lexical realizations, not in potential realizations. It was a turning point: it became evident the partiality of a supposedly fundamental lexicon, shaped exclusively on the basis of frequencies.

The available lexicon of a given speech community started to be gathered through word cues known as centers of interest (*centre d'intérêt*) -*'Les parties du corps'* ('Parts of the body'), *'Les vêtements'* ('Clothes'), *'La maison'*('The house'), among others. Given these prompts, informants would produce lists of available lexical units; it was an application of the associative controlled techniques, already used by the empirical psychology of the time. The stimuli would be identical for all speakers and so would be the reaction conditions.

This pioneering and, in many senses exemplary work, was carried out by Georges Gougenheim, René Michéa, Paul Rivenc and Aurélien Sauvageot in *L'élaboration du Français Fondamental* (Gougenheim et al. 1964). For several years, French research took the lead in lexical availability studies. Particularly relevant among them was the great Canadian project directed by William F. Mackey (1971), and realized with the collaboration of Jean Guy Savard and Pierre Ardouin.[8]

A few years prior to the two volumes published by Mackey, the Yugoslav Naum Dimitrijévic (1969) completed his work on the English lexical availability in Scottish school students. This work was apparently unknown to the Canadian researchers. In spite of the many methodological innovations it introduced, the influence of the French model on this work was evident. This influence is also

[7] The most complete information on the computations carried out in these early works is found in Mackey (1971: 61–118).

[8] Also within this line is the study by Njock (1979) on the French and the Basaa of Cameroon African children, and to a lesser extent, that by Azurmendi (1983), who worked with a bilingual population, in this case, students of the area of San Sebastián who speakers of Spanish and Basque (cf. Benitez 1991). With a much more distant relation, it is the research conducted by Bailey (1971), who looked at Spanish and English bilingual speakers in the State of Texas, in the United States. Of pedagogical purpose, all these studies propose to establish the interlinguistic distance between the languages investigated, as well as managing to ascertain the conceptual universes of the communities that they study.

evident in the first Puerto Rican investigation conducted by Humberto López Morales (1973) and in the follow-up studies after this one (López Morales 1978, 1979, 1994).[9]

1.3 Analysis of Lexical Availability

The early studies were assuming that the index of availability words was equivalent to their recorded frequency. Those who followed the steps of the French and Canadian researchers worked in the same manner -on the basis of frequencies. Sometimes the frequencies were absolute as in the case of Dimitrijévic (1969). At other times, they were relative as in Mackey's great work (1971), the work by López Morales (1973), and the projects conducted by their followers at that time. However, working with relative frequencies did little to improve the analysis of lexical availability, even when another very important factor – the number of informants in the test- was taken into consideration. Relative frequencies were usually computed by taking the absolute frequency divided by the number of subjects multiplied by 100.

These analyses were also ignoring another important aspect of lexical availability: the order of appearance of available lexical units both in the individual data and in the group data. This ordering provided information on the degree of availability of a word: highly available words were more likely to appear first in list of responses. Such an exercise would result in a more refined version of reality, since a set of words could reach the same frequency and yet have different index of availability. With a sample of nine Puerto Rican 1st primary school students, López Morales (1983, 1994) demonstrated that if only frequency was considered, discriminating between ranges was essential.[10] Thus, the processing of responses provided by the informants would yield the results shown in the Table 1.1 below.

As shown in the above Table few ranges are produced and the processing is not capable of discriminating among lexical units with identical frequency.

René Michéa (1953) had already spoken about the words that spring readily to mind (*"les mots qui viennent les premières a l'esprit"*). However, he had done this only with reference to nouns and not to other parts of discourse. A year later, coinciding with the first edition of the *L'élaboration du français fondamental*, we can read that the notion of 'degree' in lexical availability refers to the immediate presence of words in our memory ("*la présence plus ou moins immédiate de ces mots dans la memoire*"). In 1973, Charles Müller insisted on the same point, recommending the

[9] The subsequent studies to the Léxico disponible de Puerto Rico followed closely its theoretical and methodological points. The exception was López Chávez (1993), who aimed at a different purpose. See in the literature, the studies conducted by Benítez (1992, 1995), Murillo Rojas (1993, 1994), García Domínguez et al. (1994), Samper (1995), Samper and Hernández (1995, 1997), López Morales and García Marcos (1995), Mateo (1996a, b, 1994, 1997), García Marcos and Mateo (1995), Etxebarria (1996) and González Martínez (1997).

[10] The data were elicited out of the center of interest 06. 'The human body'. See López Morales (1983, 1994) for a full description of the data.

Table 1.1 Ranges of learners' lexical availability

Range	Word	Absolute	Relative
1	*mano* (hand)	6	66.6
1	*pie* (foot)	6	66.6
3	*cabeza* (head)	5	55.5
3	*ojos* (eyes)	4	44.4
3	*pierna* (leg)	43	44.4
4	*cuello* (neck)	3	33.3
4	*nariz* (nose)	3	33.3
5	*brazo* (arm)	2	22.2
5	*dedo* (finger)	2	22.2
5	*hueso* (bone)	2	22.2
5	*oreja* (ear)	2	22.2
5	*pelo* (hair)	2	22.2
6	*barriga* (tummy)	1	11.1
6	*boca* (mouth)	1	11.1
6	*carne* (flesh)	1	11.1
6	*corazón* (heart)	1	11.1
6	*hombro* (shoulder)	1	11.1
6	*ombligo* (belly)	1	11.1
6	*pecho* (breast)	1	11.1
6	*piel* (skin)	1	11.1
6	*rabo* (cock)	1	11.1
6	*rodilla* (knee)	1	11.1

analyses to take into account the rank which words occupied in the lists which –what following Gougenheim et al. (1956) were called 'index of spontaneity'. Muller's words were categorical in this respect: lexical availability should be understood on the basis of both frequency and spontaneity. Nevertheless, these recommendations were never taken up.

It was not until 1983, when formulae capable of weighting factors like frequency and spontaneity began to be developed. In the same year, the Lorán-Lopez Morales formulae appeared (Lorán 1983; Lorán and López Morales 1983), and shortly afterwards that of López Chávez and Strassbuerger (1991). These scholars were starting from the same premise: the need to develop a mathematical formula capable of weighting adequately the frequency achieved by a lexical unit as well as its place in the list. It should not be forgotten that those words that come first to our memory as reaction to a certain stimulus are really the most available. It was necessary therefore to grant a specific range to each of the available words as to determine their degree of availability.

In order to arrange the available words it was necessary to find a mathematical formula capable of providing an index for each lexical unit on the basis of its frequency and position in the list. Among other things, this step would allow discrimination within ranges the lexical units of equal frequency. The respective indexes were obtained by means of formulae, sometimes created out of pre-established properties for the classification, (as in Lorán-Lopez Morales). Other indexes were obtained by a process of trial and error as in Lopez Chávez-Strassbuerger.

The arrangement of a set of words specifies that given two any units of the set, one of them precedes the other or both are in the same position. This is a relationship of *weak order*, with linear and transitive properties. In lexical availability research, the subjects are responsible for producing a number of units in a certain order; since the same word may be retrieved by different subjects, we need to count how many times it occurs in each position in the lists. Consider the following example, taken from Butrón (1987: 23–35):

Subject 1	Subject 2	Subject 3
p^1	p^2	p^1
p^3	p^4	p^2
p^2	p^1	p^4
p^5	p^3	
	p^5	

We have three lists of different sizes; the range of each subject is not constant. Even if it were, in the same position not all the units have identical number of mentions, situation that reflects well the vector of frequencies that can be constructed for each of them:

p^1	2	0	1	0	0
p^2	1	1	1	0	0
p^3	0	1	0	1	0
p^4	0	1	1	0	0
p^5	0	0	0	1	0

The numbers in the matrix indicate that unit p^1 appears twice in the first place of lists, none in the second, one in the third and none in the fourth and fifth places, and so on. This matrix is the basis of mathematical operations that can be applied to the raw data.

Loran-López Morales, for example, weights the raw frequencies by reducing the scores of words which appear anywhere other than first in the list. Items appearing in second place score ¼ of their full value, items appearing in third place score 1/9 of their full value, and so on. In general, a word scores 1/n2 of their full value, where n is their rank order in a list. This reduction does not apply in cases where units appeared in the first place, so the original value remains unchanged.

The proposal started from a statistical framework based on the theory of decisions, comprising five axioms and three theorems, whose explanation in detail can be seen in Lorán (1987) and above all, in Butrón (1987, 1991).

However, in both Puerto Rico and in Chile (Echeverría et al. 1985), as well in Mexico, where researchers had begun working with our program for open lists, it was found that this formula lost its discriminatory power from the 23rd place. Indeed, at this stage, the curve showing cumulative weighted score is virtually flat. This was a problem that had not appeared in the empirical work that was checking the other formula, designed exclusively for lists of equal size. Subsequent revisions of this second formula (Butrón 1987, 1991) managed to control the mismatch somewhat, but failed to eliminate it altogether.

The formula used by López Chávez and Strassbuerger (1991) seems to be superior for linguistic data. Their formula manages to demonstrate a highly plausible descriptive adequacy, in the group as well as in the individual (López Chávez and Strassbuerger 1991).

Each of these two roads seeks to experimentally confirm intuition which tells us that, in a concrete situation, those words that first come to our memory are more readily available in connection with such situation than those others that do not make their appearance immediately. Therefore, the availability index is a measure that links the criteria of frequency and order to a mathematical end in a rigorous axiomatic way. Today those first attempts have resulted in very refined formulae, implemented in two computer programs that facilitate the work of computing enormously.

1.4 Expanding the Focus: From Language Teaching to Cross Multidisciplinary Approaches

The large initial projects, the French and Canadian projects, responded to goals directly related to language teaching: -on the one hand, the learning and teaching of French as a foreign language (Gougenheim et al. 1956, 1964), and on the other, determining lexical availability among the speakers of a bilingual community (Mackey 1971). Similar aims later set out by Bailey (1971), Njock (1979), Azurmendi (1983) or Dimitrijévic (1969), certainly, follow in this same line of research.

Many of the current studies primarily aim to establish the idea of a normally available lexicon for a given speech community. There is no doubt that applied linguistics obtains valuable objective instruments by means of these available lexicons. Together with basic vocabulary, the available lexicon allows vocabulary learning planners and vocabulary tasks designers to conduct a rigorous work both in mother tongue and in foreign languages.

Certainly, pedagogical applications are of paramount importance, as was already noted by Gougenheim (1967) at an early date. It is known that the lexicon of a speaking community is different from the entries comprising the usual dictionaries. The thousands and thousands of words that ordinarily appear in dictionaries are often examples of very specialized vocabularies, words unknown to people who work in other areas. The fundamental vocabulary of a given community consists of the 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 1978; López Chávez and Rodríguez Fonseca 1992; Hernández 1987; Samper and Samper Hernández 2006).

It is true that the statistical nature of these objective instruments of statistical nature needs careful evaluation. They reduce our representation of adults' lexical competence to a series of numbers. This caution is even more necessary when dealing with school children' and adolescents' lexical availability both in first, second

and foreign languages. For these groups it is absolutely necessary for us to be aware of their cognitive development and consequently, the degree of complexity of the semantic structure of the terms (aspects not always captured by numbers). But no matter what kind of weighting is carried out, they constitute the *sine qua non* of any intelligent planning basis. When there is no such programming, or this is flawed (Orama 1990), the lexical learning outcomes cannot be more calamitous (Sanavitis 1992; Lucca 1991, 1995).

In recent decades, studies of lexical availability have broadened their scope considerably to the point of addressing issues hardly envisaged by Gougenheim et al. (1964). A fruitful line of research has been the study of lexical availability in Spanish as L1 under the framework of the PanHispanic project (López-Morales 2012). Born out of the idea of building up a dictionary that could contain the available lexicon of Spanish speakers from different Spanish regions and countries, this project has given rise to an impressive body of research throughout most Spanish speaking areas in both sides of the Atlantic.

Another promising line of research that has come up in the last decade has been the one that looks at lexical availability in second and foreign languages. Within this field, researchers have paid attention to different issues such as the analysis of the vocabulary input contained in L2 textbooks compared to native speakers' available lexicons. (e.g., Benítez 1994, 1995, 1997; Carcedo 1998, 2000; Frey Pereyra 2007; García Marcos and Mateo 1995). With the exception of Chacón (2005), who focuses on English as L2, most analyses have looked at vocabulary in textbooks for learners of Spanish as second or foreign language.

Closely related to the themes addressed in this book, research on the lexical availability of learners of Spanish or English is emerging with force. This research is needed for our understanding of language learning in educational contexts. Among the issues already explored by researchers we find: age/course grade/language level (Carcedo 1998, 2000; Samper Hernández 2002), gender (López-Rivero 2008; Jiménez Catalán and Ojeda Alba 2009a, b; Hernández Muñoz 2010), language exposure (López Rivero 2008; Pérez-Serrano 2009), and type of instruction (Carcedo 1998; Germany and Cartes 2000; Jiménez Catalán and Ojeda Alba 2009a, b).

A different focus, but also one closely related to the content of this volume, is the research conducted on cognitive aspects of learners' lexical availability such as word familiarity, prototypes and structure of the mental lexicon (Ferreira and Echeverría 2010; Hernández Muñoz 2010; Hernández-Muñoz et al. 2006).

1.5 Conclusion

Lexical availability studies are renewed constantly in their search to find new lines of research and different applications. And they do it by means of crossing the borders of disciplines: dialectology, sociolinguistics, psycholinguistics and ethnolinguistics are the main disciplines that provide lexical availability studies with

theoretical frameworks and statistical possibilities, unsuspected in its modest origins. The book that you, gentle reader, now have in your hands is an excellent example of the potential of these studies for the study of learners' available lexicons, as well as for the study of learners' vocabulary knowledge in foreign language education in English and Spanish. The history of lexical availability studies, although not very extensive, has been without doubt a scientifically agile one, and everything seems to indicate that it will remain so in the near future. Watch this space.

References

Azurmendi, M.J. 1983. *Elaboración de un modelo para la descripción sociolingüística del bilingüismo y su aplicación parcial a la Comarca de San Sebastián*. Guipúzcoa: Caja de Ahorros Provincial de Guipüzcoa.

Bailey, V.J. 1971. *A study of lexical availability among monolingual-bilingual speakers of Spanish and English*. Unpublished thesis. Houston: Rice University.

Benitez, P. 1991. Reseña a Azurmendi, 1983. *Revista Española de Lingüística* 21: 355–357.

Benítez, P. 1992. Disponibilidad léxica en la zona metropolitana de Madrid. *Boletín de la Academia Puertorriqueña de la Lengua Española* 1(1): 71–102.

Benítez, P. 1994. Léxico real, léxico irreal en los manuales de enseñanza de español para extranjeros. In *Actas del II Congreso Nacional de la Asociación para la enseñanza del español como lengua extranjera*, E. Montesa Peydro and M.A. Garrido, 325–333. Málaga: ASELE.

Benítez, P. 1995. Disponibilidad léxica en Madrid: análisis cuantitativo. *III Seminario Internacional sobre 'Aportes de la lingüística a la enseñanza de la lengua materna'*. Carolina: Universidad de Puerto Rico, MS.

Benítez, P. 1997. El vocabulario en los manuales de ELE ¿es el adecuado? In *Problemas actuales en la enseñanza del español como lengua extranjera: gramática, pragmática, vocabulario y cultura*, 64–76. Utrecht: UFSIA.

Butrón, G. 1987. *El léxico disponible: índices de disponibilidad*. Tesis doctoral, Río Piedras: Universidad de Puerto Rico.

Butrón, G. 1991. Nuevos índices de disponibilidad léxica, ed. López Morales, 79–89.

Carcedo, González A. 1998. Desarrollo de la competencia léxica en el español LE. *Pragmalingüística* 5–6: 75–94.

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

Chacón, B.R. 2005. Disponibilidad lexica en el aprendizaje del inglés como lengua extranjera por hispanohablantes adultos. In *Palabras, norma, discurso. En memoria de Fernándo Lázaro Carreter*, ed. L. Santos Río et al., 753–765. Salamanca: Ediciones Universidad de Salamanca.

Dimitrijévic, N. 1969. *Lexical availability*. Heidelberg: Julius Gross Verlag.

Echeverría, M., O. Herrera, P. Moreno, and F. Pradeanas. 1985. Disponibilidad léxica en Educación Media. Resultados cuantitativos. *Lingüística Teórica y Aplicada* 25: 55–102.

Etxebarria, M. 1996. Disponibilidad léxica en escolares del País Vasco. Variación sociolingüística y modelos de enseñanza bilingüe. *Revista Española de Lingüística* 26: 301–325.

Ferreira, R.A., and M.S. Echeverría. 2010. Redes semánticas en el léxico disponible de inglés L1 e inglés LE. *Onomazein* 21: 133–153.

Frey Pereyra, Mª.L.H. 2007. Disponibilidad léxica y escritura del español como lengua extranjera: propuesta de comparación de dos corpus. *Interlingüística* 17: 366–373.

García Domínguez, M.J., V. Marrero, V. Piñero, and J.A. Pérez Marín. 1994. Análisis de la variación según la ubicación y la clase de los centros escolares en las encuestas de disponibilidad léxica en Gran Canaria. *Revista de estudios de adquisición de la lengua española* 2: 65–72.

García Marcos, F., and M.V. Mateo. 1995. Disponibilidad léxica en Almería: primeros resultados. *III Seminario Internacional sobre 'Aportes de la lingüística a la enseñanza de la lengua materna*'. Carolina: Universidad de Puerto Rico, MS.
Germany, P., and N. Cartes. 2000. Léxico disponible en inglés como segunda lengua en instrucción personalizada. *Estudios Pedagógicos* 26: 39–50.
González Martínez, A. 1997. *Disponibilidad léxica de Cádiz.* Dissertation. Cádiz: Universidad de Cádiz.
Gougenheim, G. 1967. La statistique de vocabulaire et son application dans l'enseignement des langues. *Revue de l'Enseignement Supérieur* 3: 137–144.
Gougenheim, G., R. Michéa, P. Rivenc, and A. Sauvageot. 1956. *L'élaboration du français élémentaire*. Paris: Didier.
Gougenheim, G., R. Michéa, P. Rivenc, and A. Sauvageot. 1964. *L'élaboration du français fondamental 1er degré. Étude sur l'élaboration d'un vocabulaire et d'un grammaire de base*. Nouvelle édition refondue et augmentée. Paris: Didier.
Hernández Muñoz, N. 2010. Social aspects of oral and lexical written production in Spanish. *SKY Journal of Linguistics* 23: 101–123.
Hernández-Muñoz, N., C. Izura, and A.W. Ellis. 2006. Cognitive aspects of lexical availability. *European Journal of Cognitive Psychology* 18: 730–755.
Jiménez Catalán, R.M., and J. Ojeda Alba. 2009a. Disponibilidad léxica en inglés como lengua extranjera en dos tipos de instrucción. *Lenguaje y Textos* 30: 166–176.
Jiménez Catalán, R.M., and J. Ojeda Alba. 2009b. Girls' and boys' lexical availability in EFL. *ITL International Journal of Applied Linguistics* 158: 57–76.
Justo Hernández, H. 1987. *Disponibilidad léxica en colores*. Unpublished manuscript. México: Universidad Nacional Autónoma de México.
López Chávez, J. 1993. *El léxico disponible de escolares mexicanos*. Mexico: Alhambra Mexicana.
López Chávez, J., and L. Rodríguez Fonseca. 1992. Un modelo para la programación de la enseñanza del vocabulario. *I Congreso Internacional sobre la enseñanza del español*. Madrid, CEMIP. MS.
López Chávez, J., and C. Strassbuerger. 1991. *Un modelo para el cálculo de disponibilidad léxica individual*, ed. López Morales, 91–112.
López Morales, H. 1973. *Disponibilidad léxica de los escolares de San Juan*. MS.
López Morales, H. 1978. *Frecuencia léxica, disponibilidad y programación curricular*, ed. López Morales, 77–86.
López Morales, H. 1979. Disponibilidad léxica y estratificación socioeconómica. In *Dialectología y sociolingüística. Temas puertorriqueños*, ed. H. López Morales, 173–181. Madrid: Hispanova de Ediciones.
López Morales, H. 1983. *Lingüística estadística. Introducción a la lingüística actual*. Madrid: Playor.
López Morales, H. 1994. *Enseñanza de la lengua materna. Lingüística para maestros de español*. Madrid: Playor.
López-Morales, H. 2012. EL Proyecto Panhispánico. http://www.dispolex.com/. Accessed 20 Nov 2011.
López Morales, H., and F. García Marcos. 1995. Disponibilidad léxica en Andalucía. Proyecto de investigación. *Revista de Estudios de Adquisición de la Lengua Española* 3: 67–76.
López Rivero, E. 2008. *Estudio de disponibilidad léxica en 43 estudiantes de ELE. Memoria del Máster de Enseñanza del Español como Lengua Extranjera*. Universidad Antonio de Nebrija.
Lorán, R. 1983. *Un índice de disponibilidad léxica*. Technical report. Mayagüez: Departamento de Matemáticas, Universidad de Puerto Rico.
Lorán, R. 1987. *La tasa de sustitución e índice de disponibilidad léxica*. Technical report. Mayagüez: Departamento de Matemáticas, Universidad de Puerto Rico.
Lorán, R., and H. López Morales. 1983. *Nouveau calcul de l'indice de disponibilité*. MS.
Lucca, R. 1991. *Dominio activo y pasivo del lenguaje universitario: análisis sociolingüístico*. Unpublished Dissertation. Río Piedras: Universidad de Purto Rico.

Lucca, R. 1995. Léxico activo, léxico pasivo: análisis sociolingüístico. *III Seminario Internacional sobre 'Aportes de la lingüística a la enseñanza de la lengua materna*'. Carolina: Universidad de Puerto Rico, MS.

Mackey, W.C. 1971. *Le vocabulaire disponible du français*, 2 vols. Paris/Bruxelles/Montréal: Didier.

Mateo, M.V. 1994. Disponibilidad léxica: posibles aplicaciones. In *Actas de las Primeras Jornadas sobre estudio y enseñanza del léxico*, 141–147. Granada: Universidad de Granada.

Mateo, M.V. 1996a. *Estratificación social del léxico de la alimentación en el español de Motril*. Granada/Almería: Universidad de Almería.

Mateo, M.V. 1996b. Disponibilidad léxica en Andalucía. *Interlingüística* 5: 103–109.

Mateo, M.V. 1997. *Disponibilidad léxica en el C.O.U*. almeriense, 2 vols. Tesis doctoral. Granada: Universidad de Granada.

Michéa, R. 1953. Mots fréquents et mots disponibles. Un aspect nouveau de la statistique du langage. *Les Langues Modernes* 47: 338–344.

Müller, C. 1973. *Estadística lingüística*. Madrid: Gredos.

Murillo Rojas, M. 1993. Disponibilidad léxica en los preescolares: estudio de cinco campos semánticos. *Káñina [Universidad de Costa Rica]* 17: 117–127.

Murillo Rojas, M. 1994. Comidas y bebidas: estudio de disponibilidad léxica en preescolares. *Káñina [Universidad de Costa Rica]* 18: 117–133.

Njock, P.E. 1979. *L'universe familier de l'enfant africain*. Québec: CIRB.

Orama, C. 1990. *La planificación léxica en los textos de lectura de la serie básica y la adquisición de ese vocabulario en los alumnos de nivel elemental en el sistema público de Río Piedras*. Unpublished dissertation. Río Piedras: Universidad de Puerto Rico.

Pérez-Serrano, M. 2009. *Estudio de disponibilidad léxica en estudiantes de E/LE en los centros de interés "Medios de transporte" y "Profesiones y oficios"*. Final Master's report. Instituto Cervantes and Universidad Internacional Menéndez Pelayo.

Samper, J.A. 1995. Disponibilidad léxica: resultados de las encuestas grancanarias. *XXV Simposio de la Sociedad Española de Lingüística*, Zaragoza, MS.

Samper Hernández, M. 2002. *Disponibilidad léxica en alumnos de español como lengua extranjera.* ASELE, Colección Monografías nº 4. Málaga: ASELE.

Samper, J.A., and C.E. Hernández. 1995. Disponibilidad léxica en Las Palmas: análisis cuantitativo. *III Seminario Internacional sobre 'Aportes de la lingüística a la enseñanza de la lengua materna'*. Carolina: Universidad de Puerto Rico, MS.

Samper, J.A., and C.E. Hernández. 1997. El estudio de la disponibilidad léxica en Gran Canaria: datos iniciales y variación sociolingüística. *Contribuciones al estudio de la lingüística hispánica Homenaje al Profesor Ramón Trujillo*, Vol. II, 229–239. Santa Cruz de Tenerife: Montesinos.

Samper, J.A., and M. Samper Hernández. 2006. Aportaciones recientes de los estudios de disponibilidad léxica. *Lyn X* 5: 5–95.

Sanavitis, I. 1992. *Conocimiento del léxico básico de Puerto Rico por estudiantes de nuevo ingreso a la universidad*. Unpublished thesis. Río Piedras: Universidad de Puerto Rico.

Part I
Lexical Availability in English as L1 and L2

Chapter 2
Lexical Availability of Basic and Advanced Semantic Categories in English L1 and English L2

Roberto A. Ferreira Campos and Max S. Echeverría Weasson[†]

2.1 Introduction

Second language learning is a discipline that has become increasingly important around the world, thanks to the numerous opportunities for people to work and travel in multilingual environments. In this context, there is a growing interest in improving current teaching methods and materials in order to facilitate the process of second language acquisition. Overall, policies towards second language learning seem to be pointing in the right direction, since most countries promote this activity by offering a wide range of programs to learn not only the most popular foreign languages such as English, Spanish, French, or Mandarin; but also other less known languages.

When enrolled in any language programme, either within a university or a language institute, learners are usually assessed and then classified according to their initial proficiency in the second language. This is regularly done by using local language tests or standard international tests such as TOEIC[1] in English, DELE[2] in

[†]Author was deceased at the time of publication.

To the memory of Prof. Max Echeverría Weasson, who contributed significantly to the development of Linguistics, particularly through lexical availability studies in Spanish. I (Roberto A. Ferreira) will always be grateful to him for his teaching and guidance as a researcher, colleague, and friend.

[1] Test of English for International Communication (Educational Testing Service 2012).

[2] Diplomas de Español como Lengua Extranjera (Instituto Cervantes 2012).

R.A. Ferreira Campos (✉)
Departamento de Lenguas. Facultad de Educación, Universidad Católica de la Santísima Concepción, Alonso de Rivera 2850, Campus San Andrés, Concepción, Chile
e-mail: roberto.ferreira.c@gmail.com

R.M. Jiménez Catalán (ed.), *Lexical Availability in English and Spanish as a Second Language*, Educational Linguistics 17, DOI 10.1007/978-94-007-7158-1_2,

Spanish, and DELF-DALF[3] in French, among others. Once evaluated, second language (L2) learners are typically placed under categories such as beginner, pre-intermediate, intermediate, upper-intermediate, or advanced, which are common labels for the different stages of language acquisition. As L2 learners progress through these different phases, they are supposed to acquire and advance their knowledge about the different components of the language: vocabulary, grammar, syntax, phonology or orthography, in order to become fluent L2 speakers. However, the reality seems to suggest that this might not always be the case since even advanced L2 speakers sometimes fail to use appropriate grammar and accurate vocabulary. The current study focuses specifically on vocabulary and offers an assessment of the lexicon that advanced L2 learners are capable of eliciting during a lexical availability task, in comparison with native speakers.

If people are exposed to a new language in a foreign environment, probably the first words they will try to learn are those corresponding to greetings, places, food and drink, and so forth. Interestingly, some of the words and expressions they will encounter might not even appear in dictionaries or textbooks. This is because human language is very dynamic, constantly changing and incorporating new words to its network (Ferreira and Echeverría 2010). As a result, in order to teach a foreign language properly, it is important to know what vocabulary L2 learners should be exposed to at different stages of language learning and how this vocabulary should be presented. Generally, L2 instructors organize new vocabulary in different semantic categories which fit appropriate proficiency levels. For instance, beginners are likely to be exposed to new vocabulary from categories such as 'Body parts', 'Food and drink' or 'Parts of the house'; whereas advanced students are more likely to be taught new words within the categories of 'Health and medicine', 'Politics' or 'Economy and finance'. This way of dealing with vocabulary is believed to be very beneficial (e.g., Anwar Amer 1986; Channell 1990) and is widely used in different learning materials such as textbooks for basic and intermediate levels (e.g., McCarthy and O'Dell 2002; Pye 2002; Redman 2001), and for advanced L2 learners (e.g., Richards and Sandy 2008). While there is well-established agreement that organizing the vocabulary into semantic categories is advantageous, the efficiency of the methods and criteria for selecting words within each category can be questionable.

In the process of vocabulary selection for the second language class, most researchers seem to turn to frequency of use for answers. Frequency is a very powerful variable used quite extensively in the cognitive sciences and has been shown to affect reading aloud (Balota et al. 2004), lexical decision (Balota and Chumbley 1984), and object naming (Barry et al. 1997; Ellis and Morrison 1998; Cuetos et al. 1999), among other tasks. Thus, it is not surprising that it has traditionally been used as the main method for word selection in second language teaching. The early compilation of 10,000 words in the English language by Thorndike (1921), followed by Kucĕra and Francis (1967)'s work, the CELEX database by

[3] Diplôme d'études en langue française (DELF) and Diplôme approfondi de langue française (DALF) (Centre International d'études pédagogiques 2012).

Baayen et al. (1993), and more recently, Brysbaert and New (2009)'s improved word frequency measure are good examples of the long trajectory of frequency as a well-established reference for vocabulary selection. Despite its relevant role, there is growing concern that word frequency might fail to capture informal every-day vocabulary, and probably over represents formal vocabulary found in written texts and compilations of spoken language from which frequency is extracted (Hernández-Muñoz et al. 2006). This is potentially disadvantageous for L2 learners since they are probably not being exposed to the vocabulary native (L1) speakers use in everyday life.

In view of these facts and as explained by López-Morales in the introductory chapter, another less popular variable, *lexical availability*, has been regarded as an alternative approach for vocabulary selection and the study of learners' lexicons. Lexical availability measures are obtained by having participants elicit words from different semantic categories (e.g., 'Body parts', 'Food and drink'), which are similar to those found in second language learning textbooks. After the test is conducted, each generated word is then given a lexical availability value, which is calculated based on the number of participants who produce the word, its position on the list within a given category, and the lowest position the word occupies in any of the lists (see Sect. 2.2 for more detail). Since lexical availability is obtained directly from participants and not from edited written texts (unlike frequency), it might offer a very good representation of the functional everyday vocabulary people actually use in conversations. It is true that while a participant is performing a lexical availability test, he/she sometimes produces rare words. However, as these words are unlikely to be elicited by other participants, they never reach acceptable lexical availability values and end up at the bottom of the list.

In understanding the nature of lexical availability, researchers have investigated the contribution of different predictors that can drive the lexical availability effect. Hernández-Muñoz et al. (2006), in a multilevel multiple regression analysis, found that typicality, familiarity, and age of acquisition (AoA) were the only significant predictors of lexical availability. This means that what primarily drives individuals to produce words from a given category is how typical or familiar the items in each category are, and the age at which they learned those items. Unlike the above variables, frequency was not a significant predictor of lexical availability (Hernández-Muñoz et al. 2006). This strengthens the idea that lexical availability and frequency might target slightly different things and could generate deviant results when considered as reference to select words for inclusion in second language learning materials.

In the present study, we used a lexical availability task to compare the size and availability of the vocabulary that L1 and L2 speakers are able to retrieve from different semantic categories, within a time frame of 2 min.

First, we wanted to compare advanced university L2 learners and native speakers regarding number of words produced across basic and advanced semantic categories.[4] Second, we were also interested in looking at lexical availability values including

[4] Basic categories correspond to language units introduced at a beginners' level, whereas advanced categories represent units students learned at an advanced level.

speakers and type of semantic category (basic or advanced) as factors. Third, we also carried out a correlational analysis between lexical availability values of words generated by L1 and L2 speakers across basic and advanced semantic categories.

A number of hypotheses were tested. We first assessed the hypothesis that L1 speakers would outperform L2 speakers regarding the mean number of words produced in each semantic category. This might not be particularly surprising since L1 speakers in this study had lived in an English speaking environment since they were born, whereas L2 speakers learned English as a second language primarily in a school setting. We also expected that both L1 and L2 speakers would elicit more words for basic categories (e.g., 'Body parts') than for advanced categories (e.g., 'Terrorism and crime'). This is based on the assumption that words belonging to basic categories are likely to be more familiar, typical or learned earlier in life than words from advanced categories. Predictions regarding a direct comparison of lexical availability in L1 and L2 are not very straightforward. However, in line with the first hypothesis, lexical availability values should be higher in L1 than L2 speakers because the more words produced from a given category, the greater the chance of words to be repeated across participants, which would increase lexical availability values. Similarly, we expected that basic semantic categories would show higher lexical availability values in comparison with advanced categories. Basic categories seem to have higher familiarity and are generally acquired earlier in life, which would benefit word production and, consequently, lexical availability. As stated earlier, this is supported by Hernández-Muñoz et al. (2006)'s study, which found that familiarity and AoA were strong predictors of lexical availability. See also Chap. 3 by Jiménez Catalán, Agustín, Fernández, and Canga in this volume. Finally, we also expected to find a correlation between lexical availability values of words produced by native speakers and the same words elicited by L2 speakers.

2.2 Method

2.2.1 Participants

The data used in this Chapter is part of a larger data set collected by Ferreira (2006). The investigation included a total of 50 English native speakers (mean age 16.4, SD 0.6) and 50 advanced second language students (mean age 21.4, SD 0.4). All native speakers who qualified for the study were monolingual female students at The Royal School located in Haslemere, Surrey, United Kingdom. Prior to the lexical availability test, all prospective participants were asked orally whether they were able to speak a second language. Students who reported that they did so were excluded from the study before it took place. The fact that L1 speakers in this study were non-specialized secondary school students allowed us to obtain a sample of the average vocabulary English speakers can produce in a timeframe of 2 min. The L2 speakers were all undergraduate students from the University of Concepción,

Chile. Prior to enrolment at the university, they had studied English in a school setting for 8 years on average. All the students were in their fourth (last) undergraduate year, so they had completed at least 1,000 h of instruction in English. Their academic program included several general English language courses covering pre-intermediate, intermediate, upper-intermediate, and advanced levels. Other more advanced courses comprised phonetics, English literature, American and British history, applied linguistics, translation (English-Spanish); apart from optional courses such as academic writing, short-story writing, and drama.

2.2.2 Materials and Design

The full data set by Ferreira (2006) included ten semantic categories, which were selected on the basis of previously established categories as part of the Panhispanic Project[5] (see also López-Morales 2012, for details) and English as a Second Language (ESL) textbooks such as the Interchange series (Richards et al. 2005). The current investigation only used four categories in order to examine relevant factors and interactions more carefully. In Ferreira (2006) all semantic categories were classified into basic or advanced, depending on the degree of specialization of the lexicon they contained. Here, we revalidated this classification by asking 20 currently employed English teachers to classify all ten categories into basic or advanced. Participants were told to choose five categories or units that they would normally use to teach beginner students and five categories they were more likely to use with advanced students. See Appendix 2.1 for Instructions sheet. All participants agreed that 'Body parts', 'Food and drink', 'Holidays', 'Clothes', and 'Entertainment' were more likely to be taught at a beginners' level; whereas 'Economy and finance', 'Terrorism and crime', 'Politics', 'Pollution and the environment', and 'Health and medicine' were more suited for an advanced audience. For the current publication, we randomly selected two basic ('Body parts' and 'Food and drink') and two advanced ('Terrorism and crime', and 'Health and medicine') categories.

2.2.3 Procedure

Both L1 and L2 speakers were given a paper-based lexical availability test in a classroom setting. They were presented with all ten semantic categories in a pseudorandom order, in order to ensure that categories corresponding to the same classification (basic or advanced) never appeared together. Each category was displayed on a different page and participants were told to read the name of the category (appearing on top) and then write as many words as possible from the given category within a time period of 2 min. A table with 50 spaces was provided

[5] See Chap. 1 by Humberto López Morales in this volume.

$$D(P_j) = \sum_{i=1}^{n} e^{\left(-2.3\times\left(\frac{i-1}{n-1}\right)\right)} \times \frac{f_{ji}}{I_1}$$

Fig. 2.1 Formula to calculate lexical availability ('D(Pj)' represents the lexical availability value of the word j within a semantic category; 'I' denotes the total number of participants who performed the test; 'i' represents the position of the word j in a given list; 'f' is the number of participants who elicited the word j at that position in their list; 'n' denotes the lowest position obtained by word j in any list produced for the category; and 'e' is the natural number 2.718181818459045 (see Hernández-Muñoz et al. 2006))

for the purpose (see Appendix 2.2). Participants were not allowed to move on to the following page until the end of the 2-min period, and were asked to immediately hand in the test after all semantic categories were presented. The complete test lasted around 20 min.

In order to edit and process the lists of words obtained from the participants, a set of criteria was adopted. Windows XP Note block (Microsoft Corporation 2007) was used to type in the responses produced by both groups of participants. First, two different number codes were used to differentiate the different types of speakers: L1 speakers were coded as 11111, while L2 speakers were identified with the number 11112. Each participant in each group was identified with a number ranging from 001 to 050 and a similar procedure was adopted to code semantic categories, which ranged from 01 to 10. Responses from each participant were entered in lower-case including the group code first, followed by the participant's code and finally the category code. Each code was separated from other codes and the words by a single spacebar press, whereas words were separated from each other by a comma followed by a spacebar press (e.g., 11111 001 01 leg, arm, hand). Each list of words corresponding to the same category and the same participant was separated from subsequent lists by an *Enter* press so that each list would be placed on a different line. Regular nouns and adjectives were typed in singular form, but irregular nouns were kept in their original form. Except for gerunds and participles, all other verb forms were transformed to infinitive. Finally, compound nouns, short phrases, or expressions (e.g., *orange squash*, *september eleventh*, etc.) were hyphenated (e.g., *orange-squash*, *september-eleventh*) and turned into a single entry.

After all words were entered in Block note, they were saved in a single txt. file in order to be processed. Data processing was carried out using *Dispogen II* (Echeverría et al. 2005), which allowed us to obtain lexical availability values for each word elicited in each semantic category. This software is an application created in MATLAB version 7 (The Math Works Inc. 2005) and uses a formula developed by López-Chávez and Strassburguer-Frías (1991) (see Fig. 2.1), which computes lexical availability values according to the position that a word takes in a list, the number of participants who elicit the word at those positions, and the lowest position the word is observed in any of the lists (see Hernández-Muñoz et al. 2006). Based on this formula, words produced by a large number of participants and which appear early on the lists will obtain a high lexical availability value, whereas words elicited

by few participants and appearing at the bottom of the lists will rank low in lexical availability.

As stated earlier, only four categories out of ten were used in the current analysis. These corresponded to 'Body parts', 'Food and drink' (basic), and 'Terrorism and crime', 'Health and medicine' (advanced).

2.3 Results

Results were obtained for three main analyses. The first analysis aimed to examine differences between L1 and L2 speakers and between semantic categories regarding mean number of words produced (see Table 2.1). The second analysis included a direct comparison of lexical availability values across speakers and categories for the 100 most available words in each category and each group of speakers (see Table 2.2). Mixed-factorial analysis of variance (ANOVA) was used in the first and second analyses. When further analyses were required, one-way within subjects ANOVAs and Bonferroni-corrected t-tests were used. Effect sizes were reported using partial Eta squared (η_p^2) and when sphericity was violated, Greenhouse-Geisser correction was used to adjust *p* values. The third analysis included a correlation (Spearman's rho) between the lexical availability values of the 100 most available words produced by native speakers in each category and the values of the same words generated by L2 speakers. When a word elicited by native speakers was not generated by L2 speakers, it received a lexical availability value of '0'.

Table 2.1 Total number of different words per category and mean number of words per participant in each group of speakers

	Speakers	BP	F&D	T&C	H&M
Total number of different words	L1	206	450	413	468
	L2	109	253	316	276
Mean number of words	L1	26.7	29.6	21.0	23.6
	L2	21.2	23.6	14.9	15.1

Note: *BP* 'Body parts', *F&D* 'Food and drink', *T&C* 'Terrorism and crime', *H&M* 'Health and medicine'

Table 2.2 Mean lexical availability values for the first 100 words in L1 and L2 speakers

	BPL1	BPL2	F&DL1	F&DL2	T&CL1	T&CL2	H&ML1	H&ML2
Mean lexical availability	0.14	0.11	0.11	0.11	0.08	0.06	0.09	0.06
SD	0.18	0.17	0.08	0.09	0.09	0.06	0.11	0.09
Range	0.01–0.79	0.00–0.76	0.03–0.49	0.02–0.38	0.03–0.48	0.02–0.31	0.03–0.75	0.02–0.58

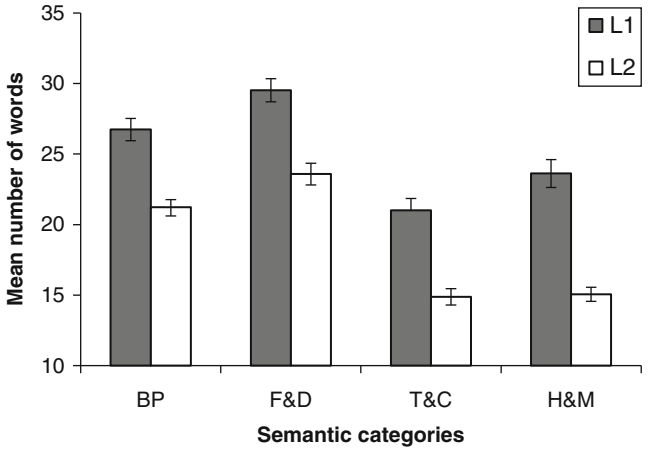


Fig. 2.2 Mean number of words produced by participants in each group (L1 and L2)

2.3.1 Analysis 1: Mean Number of Words Produced

As explained earlier, this analysis assessed the number of words retrieved by L1 and L2 speakers in each semantic category, and the possible interactions between group and semantic category. See Table 2.1 and Fig. 2.2.

The first step in the analysis was carried out using a 2 × 4 mixed-factorial ANOVA with speaker (L1, L2) and category ('Body parts', 'Food and drink', 'Terrorism and crime', 'Health and medicine') as the main factors. The mixed-factorial ANOVA showed a significant main effect of group, with native speakers outperforming non-native speakers, $F_1(1, 98) = 2808.65$, MSE = 68.65, $p < .001$, $\eta_p^2 = .39$. There was also a significant effect of semantic category, $F_1(1, 98) = 112.86$, MSE = 17.27, $p < .001$, $\eta_p^2 = .53$, and a significant group x category interaction, $F_1(1, 98) = 3.31$, MSE = 17.27, $p = .03$, $\eta_p^2 = .03$.

In order to explore the group x category interaction, separate one-way ANOVAs and post hoc tests were conducted for the data from each group of speakers. Results of the ANOVA conducted on the L1 data showed a significant main effect of category, $F_1(1, 98) = 40.91$, MSE = 20.08, $p < .001$, $\eta_p^2 = .46$. Bonferroni-corrected paired-samples t-tests were used for all post hoc analyses. Results showed that native speakers produced a significantly higher number of words for 'Food and drink' than for all other three categories ($p < .001$). The second most productive semantic category was 'Body parts', where participants elicited significantly more words than for 'Terrorism and crime', and 'Health and medicine' ($p < .001$). The least productive category was 'Terrorism and crime' that showed a significantly lower number of words produced than 'Health and medicine' ($p < .01$).

Results for the one-way ANOVA run on the data corresponding to L2 speakers also showed a significant main effect of semantic category, $F_1(1, 98) = 82.58$,

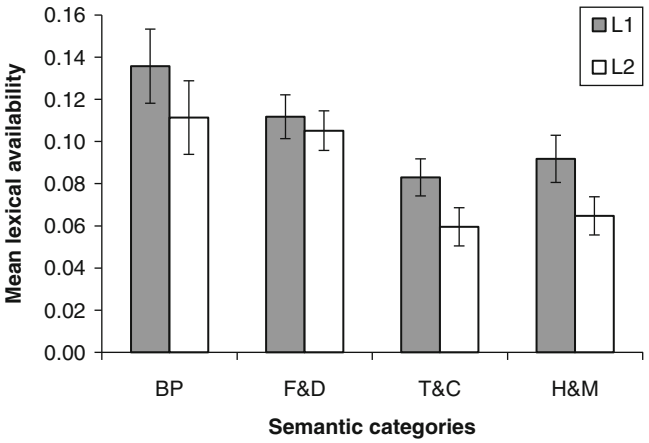


Fig. 2.3 Mean lexical availability for the 100 words with the highest values in L1 and L2 speakers

MSE = 15.49, $p < .001$, $\eta_p^2 = .63$. Post hoc tests (Bonferroni-corrected) revealed that advanced L2 speakers (similar to L1 speakers) produced significantly more words for 'Food and drink' than for 'Body parts' ($p < .01$), 'Terrorism and crime' ($p < .001$), and 'Health and medicine' ($p < .001$). The second most productive category was also 'Body parts', which showed significantly more words than 'Terrorism and crime', and 'Health and medicine'. Unlike the results in the L1 group, 'Terrorism and crime' and 'Health and medicine' did not differ in the L2 group ($p = 1.0$).

In summary, native speakers outperformed L2 speakers within all semantic categories. Overall, both native and non-native speakers produced more words for basic semantic categories ('Body parts', and 'Food and drink') than for advanced categories ('Terrorism and crime', and 'Health and medicine'). The group x category interaction can be explained by the fact that L1 speakers elicited more words for 'Health and medicine' than for 'Terrorism and crime', whereas these two categories were not significantly different from each other in the group of L2 speakers.

2.3.2 Analysis 2: Lexical Availability

In order to perform Analysis 2, the 100 words with the highest lexical availability values from each group of speakers in each semantic category were selected. See Appendix 2.3 for a sample of ten words in each category and each group of speakers Fig. 2.3.

As in Analysis 1, a 2 × 4 mixed-factorial ANOVA was first conducted on the data and included the same main factors. The factorial ANOVA found no effect of group, $F_1(1, 98) = 1.79$, MSE = 0.05, $p = .18$, $\eta_p^2 = .01$. However, there was a highly significant

Table 2.3 Correlations of lexical availability in L1 and L2 speakers for the first 100 words in each category

	Semantic categories	Spearman's rho
Basic	'Body parts'	.79***
	'Food and drink'	.51***
Advanced	'Terrorism and crime'	.45***
	'Health and medicine'	.47***

Note: *** = <.001

effect of semantic category, $F_1(1, 98) = 45.22$, MSE = 0.01, $p < .001$, $\eta_p^2 = .19$. The group x category interaction did not reach significance, $F_1(1, 98) = 82.58$, MSE = 15.49, $p < .001$, $\eta_p^2 = .63$. Since there was no effect of group or interaction, but a significant effect of category, post hoc tests (Bonferroni-corrected) were run on the L1 and L2 data combined. No difference was found between the two basic categories, 'Body parts' and 'Food and drink', (p = .13). However, both these categories showed significantly higher lexical availability values than any of the advanced categories ('Terrorism and crime', and 'Health and medicine' (p < .001)). At the same time, 'Health and medicine' showed higher lexical availability than 'Terrorism and crime' (p = .04).

2.3.3 Analysis 3: Correlation of Lexical Availability Between Native and Non-native Speakers

Despite the fact that in the previous analysis lexical availability values were compared across speakers and categories, this did not clarify whether lexical availability values of words in L1 speakers correlate with the values of the same words in L2 speakers. In order to investigate this, the 100 most available words produced by native speakers in each category were used once again. However, this time the analysis compared the lexical availability values of these words with the same words produced by L2 speakers. Bivariate nonparametric correlations (Spearman's rho) were performed on the data from each semantic category. As observed in Table 2.3 and Fig. 2.4, there was a high correlation between lexical availability values in L1 and L2 speakers across the different categories. Both basic ('Body parts', and 'Food and drink') and both advanced semantic categories ('Economy and finance', and 'Terrorism and crime') showed a highly significant correlation (p < .001). It is also important to notice that basic categories, especially 'Body parts' showed higher correlations than advanced categories.

Despite the fact that there were very significant correlations between L1 and L2 speakers, a substantial number of words with high lexical availability values (among the 100 most available) produced by L1 speakers were not elicited by L2 speakers. Some examples include *kidney, fingernail*, and *torso* in 'Body parts'; *crisp, carbohydrate*, and *protein* in 'Food and drink'; *september-eleventh, saddam-hussein*, and *burglary* in 'Terrorism and crime'; *paracetamol, nhs*, and *penicillin* in 'Health and medicine'. See Appendix 2.4 for full list.

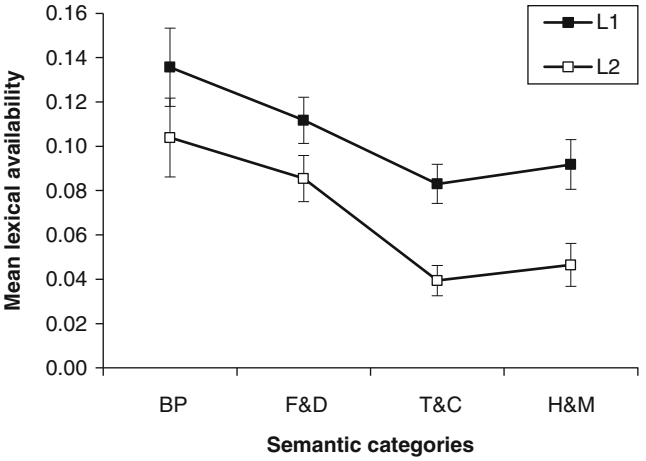


Fig. 2.4 Correlation of lexical availability for the 100 words with the highest values in L1 speakers and their translation equivalent in L2 speakers

2.4 Discussion

The ultimate aim of this investigation was to compare the lexicon of native speakers and advanced students of English as a second language across different semantic categories, using number of words produced and lexical availability as dependent variables. Another important aim was to examine whether basic (e.g., 'Food and drink') and advanced (e.g., 'Health and medicine') semantic categories would show divergent results across L1 and L2 speakers.

The first part of the investigation focused on the average number of words produced by each participant in each semantic category: 'Body parts', 'Food and drink', 'Terrorism and crime', and 'Health and medicine'. The first hypothesis stated that L1 speakers would outperform L2 speakers regarding number of words produced in each semantic category. The results confirmed this hypothesis since L1 speakers clearly produced a higher number of words than L2 speakers across all semantic categories. This was not surprising considering that native speakers are exposed to their mother tongue at all times, so they clearly get more exposure to the language than L2 speakers. However, it is important to notice that this might not be the only reason why L1 performed better than advanced L2 language users. Another important factor could be the fact that native speakers in this study were all monolinguals, so they were able to elicit words in their mother tongue without facing competition from words in another language. There is widespread evidence suggesting that bilingual language processing is nonselective (e.g., Ferreira 2011; Dijkstra 2005; Costa et al. 1999), which means that words from both languages become activated and compete for selection during word production. In this particular case, it is possible that when the L2 participants were asked to produce words from a given semantic category, they encountered more difficulties than

monolinguals to select appropriate words in their L2, due to possible interferences from the L1. This implies that L2 speakers might have lost time suppressing words in their L1 in order to only elicit words that are part of the L2 lexicon. This almost unnoticeable phenomenon is also likely to increase the demand on memory resources, which can certainly delay word production or make the process more error-prone (Ferreira 2011).

The second hypothesis in this study predicted that basic semantic categories would show a greater number of words than advanced categories across both groups of participants. This hypothesis was also confirmed since the two basic categories ('Body parts' and 'Food and drink') showed significantly more words than the advanced categories ('Terrorism and crime', and 'Health and medicine'). This might be because words from basic categories tend to be more familiar than words from advanced categories, thus fostering production and increasing lexical availability. This explanation is in line with the results of a multilevel regression analysis performed by Hernández-Muñoz et al. (2006), which showed that familiarity was one of the strongest predictors of lexical availability. Familiarity has been defined as a measure of how often people think of concepts or things and is obtained by having participants rate different concepts (Cycowicz et al. 1997). Based on these ratings, familiarity has been found to influence word recognition (e.g., Cuetos et al. 2002) and word production (e.g., Ellis and Morrison 1998). In a lexical availability test, where participants are required to produce words from different semantic categories, words with higher familiarity are more likely to get activated and, consequently, elicited than words with lower familiarity. Another factor that can also influence the number of words produced in a lexical availability task is age of acquisition (AoA) or order of acquisition. This variable is a strong predictor of accuracy and speed in different language tasks such as reading (Monaghan and Ellis 2002; Morrison and Ellis 2000) and object naming (Carroll and White 1973; Bates et al. 2001; Ellis and Morrison 1998; Snodgrass and Yuditsky 1996). Since words belonging to basic categories are likely to be learned early in life (e.g., *head, water*), they might be easy to activate and elicit as opposed to words from advanced categories, which are more likely to be learned at a much later stage in life (e.g., *murder, drug*). Given the above, AoA seems to be another important factor contributing to the difference between basic and advanced semantic categories.

The fact that both L1 and L2 speakers produced more words for basic than advanced semantic categories reflects, to some extent, that both groups of language users behaved similarly regarding category type. The only difference found was outlined by the significant group x semantic category interaction, which reflects the advantage of 'Terrorism and crime' over 'Health and medicine' (advanced categories) only present in native speakers. This suggests that despite the underlying differences in vocabulary production between the two groups, the words acquired by L2 speakers throughout the learning stages take similar pathways to those of native speakers. Perhaps this represents similarities in exposure, acquisition, and organization of the words in the mental lexicon.

The second set of predictions involved lexical availability values. In line with the number of words produced, it was expected that lexical availability values would be much higher for native speakers than for L2 speakers. However, this hypothesis was

not confirmed since no difference between the two groups of speakers was found. This result might suggest that both native and non-native speakers show similar patterns of organization for the most available vocabulary, independently of whether they know the same words. This result is supported by the fact that no interaction between group and semantic category was found, which implies that both groups of language users show a similar pattern of behavior when producing words from different semantic categories.

It was also predicted that basic categories would show an advantage in comparison with advanced categories. This hypothesis was successfully confirmed since the two basic categories ('Body parts' and 'Food and drink') showed significantly higher lexical availability values than the advanced categories ('Terrorism and crime', and 'Health and medicine'). This means that words generated from advanced categories varied more across participants than words produced from basic categories. This high variability is explained by the fact that each individual word in the advanced categories was produced by fewer participants. This difference in lexical availability had previously been demonstrated for abstract ('Intelligence') versus concrete categories (e.g., 'Animals'), where 'Intelligence' showed lower lexical availability values than four different concrete categories (Hernández-Muñoz et al. 2006), but had never been assessed for proficiency levels such as basic and advanced. An important factor that could help understand underlying differences in lexical availability between semantic categories is familiarity. As stated earlier, it seems that basic categories tend to gather words with higher familiarity than advanced categories. This is supported by the fact that familiarity has been found to correlate strongly with lexical availability (Hernández-Muñoz et al. 2006) and has been reported to benefit performance during different tasks such as object naming (e.g., Ellis and Morrison 1998; Cuetos et al. 1999), and semantic categorization (e.g., Larochelle and Pineau 1994; Malt and Smith 1982). Age of acquisition (AoA) is another variable that has shown a high (negative) correlation with lexical availability (Hernández-Muñoz et al. 2006) and provides more insights into the nature of the available lexicon. In this line, we can argue that basic categories are more likely to contain words acquired early in life than advanced categories since words belonging to basic categories showed higher lexical availability than those in advanced categories.

The third analysis of this study looked at correlations between lexical availability values in L1 and L2 across basic and advanced semantic categories. All categories showed very high correlations between L1 and L2 speakers, but it is important to notice that basic semantic categories, especially 'Body parts', showed a much higher correlation than advanced categories. This might suggest that basic categories experience less lexical variability across speakers, and that L2 learners acquire vocabulary from these categories more accurately. Even though all correlations between L1 and L2 speakers were highly significant, a sizeable number of words with high lexical availability values in L1 speakers was not produced by L2 speakers. This shows that despite their high proficiency level, advanced L2 speakers still struggle to produce relatively common words (used by native speakers) when a semantic category is presented as stimulus. The failure to produce these apparently common words might reflect difficulties during the retrieval of the words' lexical representations. This is likely to be caused by incomplete word knowledge or

interference from the nontarget language as has been shown in studies of word production (e.g., Ferreira 2011; Costa 2005; De Bot 1992; Poulisse 1999). It is important to notice that available words, which were not elicited by L2 speakers, might not show differences in performance between L1 and L2 speakers in other tasks such as naming or recognition memory. In a word learning study conducted by Ferreira (2011), L1 speakers clearly outperformed advanced L2 speakers in a production task, where participants were asked to elicit novel words based on orthographic cues and a definition. However, participants did not differ in the two recognition tasks: reading aloud and recognition memory. This implies that advanced L2 speakers are comparable to native speakers in recognition tasks, but they struggle to match native speakers in production tasks. This might be due to the fact that production tasks require participants to activate different components of the words in order to elicit their lexical representation, which is probably harder for L2 speakers, who face competition from the L1 language.

The process of word production seems to be hierarchical (Caramazza 1997; Levelt 1989), so activation flows from conceptual representations to phonological and orthographic representations. This also seems to be the case for bilingual word production (Costa 2005), except that lexical representations from both languages can be activated. In a lexical availability task, participants are asked to produce words from a semantic category (e.g., 'Body parts'). If we assume that word production is hierarchical, then semantic representations related to 'Body parts' would be first activated in the speaker's mental lexicon. Then activation would spread to the lexical level, where lexical representations start competing for selection. At this stage, hierarchical monolingual models propose the activation of several candidates within the target language (e.g., *head*, *leg*, *arm*, etc., for 'Body parts'). Since bilingual language processing appears to be nonselective (e.g., Costa et al. 1999; De Bot 1992; Poulisse 1999), hierarchical bilingual models of word production propose that lexical competition also includes words from the nontarget language, in this case, Spanish. Thus, L2 speakers in the current investigation probably also activated words such as *cabeza*, *pierna*, and *brazo* when attempting to produce *head*, *leg*, and *arm*. The fact that words from the nontarget language become activated might make retrieval in the target language harder for L2 speakers than for L1 speakers. This together with the fact that L2 words are also less integrated in the mental lexicon could produce a decline in performance, with L2 speakers failing to elicit words such as *kidney* in 'Body parts' or *plaster* in 'Health and medicine', which are relatively common words in the English language.

In summary, this investigation has provided new insights into the nature of the lexicon of advanced L2 speakers in comparison with monolingual native speakers. It has been demonstrated that the latter outperform non-native speakers regarding number of words produced in each semantic category. However, groups do not significantly differ when comparing lexical availability values for the 100 most available words, which suggests that despite the difference in number of words produced, both L1 and L2 speakers share similar patterns of integration and organization of the words in the mental lexicon. This is also supported by the fact that lexical availability values correlate highly between native and non-native speakers in each semantic

category. It is important to notice, however, that despite the similarities across groups in relation to lexical availability values, advanced L2 speakers are still unable to activate and elicit a number of words which are highly available among native speakers. This fact is important to consider when teaching English as a second language. Perhaps the current methodologies used to teach new vocabulary are not completely appropriate, or the vocabulary itself might not entirely correspond to the one used by native speakers. Due to the fact that most textbooks used at school or even in higher education have been elaborated based on word frequency and do not take into account other variables such as lexical availability, they might be failing to capture the average vocabulary that native speakers use in everyday life. It is worth noticing that L1 speakers in this investigation were not specialized in any particular area, so their 'available' lexicon is a glimpse inside the average vocabulary used by native speakers of English. Given the results of our investigation, we propose lexical availability as a complement of word frequency in the selection of words for inclusion in different types of ESL materials. By combining both methods we can ensure that ESL students do get access to some words, which even though are highly common in the language, they are not captured by word frequency alone.

Regarding basic and advanced semantic categories, it was discovered that both native and non-native speakers produced more words from basic than advanced categories. Lexical availability values were also higher for basic than advanced semantic categories. These two findings are particularly important since they reveal that L2 speakers follow the same pattern of vocabulary growth and organization as native speakers. This implies that traditional teaching methods, which present vocabulary organized in a progression of units or lessons, are fairly accurate in simulating the way native speakers deal with vocabulary, but probably fail to introduce all relevant lexicon commonly used by native speakers.

In conclusion, our findings suggest that even though L2 speakers resemble native speakers in different aspects of vocabulary development, they might still need to incorporate relevant words to their available lexicon. Noteworthy, we need to be cautious about our claims since the samples used (50 participants in each group) are certainly not representative of the entire population of English native speakers or that of advanced L2 English users. Our research is only a first attempt to directly compare monolingual native speakers and advanced L2 speakers regarding their performance in a lexical availability task, and to provide relevant cognitive explanations about the processes underlying word production. Future studies will require much bigger samples, perhaps from different geographical regions, in order to cover the full spectrum of target populations. It would also be advisable to have tighter control over sociocultural and economic variables that might have an effect on the results.

Acknowledgments This study was partially supported by Proyecto Fondecyt 1050598 (2005). We would like to thank Linda Salter and Paul Salter for their collaboration in collecting the data from the English speakers. We would also like to thank Prof. Lilian Gómez who allowed us to take part of her teaching time in order to conduct the lexical availability test at the University of Concepción. Many thanks to the students from the Royal School Haslemere, and the students from the University of Concepcion, who participated in the study. Finally, we would also like to thank Mariya Bistrina for proofreading this chapter.

Appendices

Appendix 2.1. Instructions for Classifying Semantic Categories into Basic or Advanced

If you were to teach English as a second language to two different groups of students, one 'beginners' and one 'advanced'; which of the following units would you use with beginners and which do you think you would rather use with advanced students? Choose five different units for each group. If you think that one unit can fit both groups, please tick (√) both indicating the order of preference (first or second). If you need to explain your choices, please write below in the space provided next to 'comments'.

Units	Beginners	Advanced
'Pollution and the environment'	___	___
'Entertainment'	___	___
'Body parts'	___	___
'Health and medicine'	___	___
'Politics'	___	___
'Food and drink'	___	___
'Terrorism and crime'	___	___
'Holiday'	___	___
'Clothes'	___	___
'Economy and finance'	___	___

Comments: _____
_____
_____
_____
_____
_____
_____

Appendix 2.2. Lexical Availability Test

Instructions

Read the instructions carefully before you begin.

1. On each of the following pages you will be presented with the name of a category (topic). Write down as many words from that category as you can within a given time period of 2 min.

2. Do not look at the categories until the experimenter allows you to.
3. Use clear/readable handwriting.
4. When the experimenter says you can start, turn to next page. There you will find the name of the first category, read it carefully and write down any possible words that come to your mind.
5. Stop writing when the experimenter says "Stop". Then turn to next page and repeat the same procedure until all the categories are covered.
6. When you reach the end of the test, you will not be allowed to add new words to previous categories, so just hand in your test.
7. The complete test will take around 25 min as there are only 12 categories.

Now you are ready to start!

'Body parts'	
1	26
2	27
3	28
4	29
5	30
6	31
7	32
8	33
9	34
10	35
11	36
12	37
13	38
14	39
15	40
16	41
17	42
18	43
19	44
20	45
21	46
22	47
23	48
24	49
25	50

Appendix 2.3. Lexical Availability Values for the Ten Most Available Words in Each Semantic Category and Each Group of Speakers

L1 speakers		L2 speakers	
'Body parts'	Lexical availability	'Body parts'	Lexical availability
leg	0.79	*head*	0.76
arm	0.77	*arm*	0.68
head	0.67	*leg*	0.62
eye	0.63	*hand*	0.61
finger	0.56	*eye*	0.57
hand	0.54	*nose*	0.55
nose	0.52	*mouth*	0.53
toe	0.52	*finger*	0.50
ear	0.43	*ear*	0.48
mouth	0.40	*hair*	0.36
'Food and drink'		'Food and drink'	
water	0.49	*meat*	0.38
apple	0.38	*tomato*	0.37
orange	0.29	*apple*	0.36
coke	0.29	*potato*	0.35
chip	0.27	*bread*	0.35
chocolate	0.26	*milk*	0.33
pasta	0.26	*chicken*	0.28
banana	0.23	*lettuce*	0.28
carrot	0.22	*juice*	0.26
wine	0.22	*orange*	0.26
'Terrorism and crime'		'Terrorism and crime'	
bomb	0.48	*gun*	0.31
death	0.43	*bomb*	0.30
police	0.37	*murder*	0.30
gun	0.34	*death*	0.29
murder	0.33	*robbery*	0.19
prison	0.32	*war*	0.16
twin-towers	0.24	*kill*	0.16
iraq	0.22	*police*	0.15
war	0.19	*blood*	0.13
jail	0.18	*eta*	0.13
'Health and medicine'		'Health and medicine'	
doctor	0.75	*doctor*	0.58
hospital	0.58	*hospital*	0.47
nurse	0.53	*nurse*	0.38
cancer	0.35	*illness*	0.35
paracetamol	0.25	*pill*	0.29
injection	0.25	*cancer*	0.19
nhs	0.24	*disease*	0.17
death	0.23	*medicine*	0.15
surgery	0.21	*sick*	0.15
disease	0.21	*aids*	0.12

Appendix 2.4. Words That Were Not Elicited by L2 Speakers Despite Being Among the 100 Most Available Words Produced by L1 Speakers

'Body parts'	'Food and drink'	'Terrorism and crime'	'Health and medicine'
kidney	*crisp*	*september-eleventh*	*paracetamol*
fingernail	*carbohydrate*	*saddam-hussein*	*nhs*
torso	*protein*	*burglary*	*penicillin*
bladder	*healthy*	*bus*	*nurofen*
tummy	*plum*	*train*	*calpol*
intestine	*curry*	*london*	*gp*
artery	*ribena*	*abuse*	*help*
calf	*orange-squash*	*evil*	*exercise*
organ	*duck*	*weapon*	*pharmacy*
tendon	*wheat*	*injury*	*clean*
oesophagus	*unhealthy*	*racism*	*smoking*
forehead	*raisin*	*london-bombing*	*private*
skull	*Taste*	*torture*	*happy*
bum	*raspberry*	*arrest*	*unhealthy*
rib	*hunger*	*punishment*	*well*
collar-bone	*baked-beans*	*hostage*	*plaster*
pancreas	*sparkling*	*shoplifting*	*care*
joint	*tango*	*palestine*	*ibuprofen*
biceps	*sweetcorn*	*suicide-bomber*	*unwell*
triceps	–	*wrong*	*anaesthetic*
biology	–	*new-york*	*hospital-bed*
pelvis	–	*loss*	*expensive*
anatomy	–	*fraud*	*sti*
femur	–	*stupid*	*chemotherapy*
boob	–	*jury*	*poorly*
ugly	–	*trial*	*malaria*
gall-bladder	–	*prisoner*	*drip*
human	–	*crying*	*running*
ear-lobe	–	*egypt*	*bandage*
animal	–	*security*	*treatment*
internal-organ	–	*chaos*	*waiting-room*
small-intestine	–	*bad*	*bupa*
different	–	*underground*	*hay-fever*
ribcage	–	*anger*	*healthy-diet*
large-intestine	–	*cruel*	*waiting-list*
spot	–	*son-of-bin-laden*	*anorexia*
trachea	–	*pikey*	*phd*
body	–	*life-sentence*	*cream*
spleen	–	*legal*	*a + e*
coccyx	–	*penalty*	*std*
urethra	–	–	*drink*
Feature	–	–	*bad*
Alveoli	–	–	*liquid*
Digit	–	–	*morphine*
Thin	–	–	–

References

Anwar Amer, A.A. 1986. Semantic field theory and vocabulary teaching. *English Teaching Forum* 24: 30–31.

Baayen, R.H., R. Piepenbrock, and H. van Rijn. 1993. *The CELEX lexical database (CD-ROM)*. Philadelphia: Linguistic Data Consortium.

Balota, D.A., and J.I. Chumbley. 1984. Are lexical decisions a good measure of lexical access? The role of word frequency in the neglected decision state. *Journal of Experimental Psychology: General* 133: 283–316.

Balota, D.A., M.J. Cortese, S.D. Sergent-Marshall, D.H. Spieler, and M.J. Yap. 2004. Visual word recognition of single-syllable words. *Journal of Experimental Psychology: General* 133: 283–316.

Barry, C., C.M. Morrison, and A.W. Ellis. 1997. Naming the Snodgrass and Vanderwart pictures: Effects of age of acquisition, frequency and name agreement. *Quarterly Journal of Experimental Psychology* 50A: 560–585.

Bates, E., C. Burani, S. D'Amico, and L. Barca. 2001. Word reading and picture naming in Italian. *Memory and Cognition* 29: 986–999.

Brysbaert, M., and B. New. 2009. Moving beyond Kučera and Francis: A critical evaluation of current word frequency norms and the introduction of a new and improved word frequency measure for American English. *Behavior Research Methods* 41: 977–990.

Caramazza, A. 1997. How many levels of processing are there in lexical access? *Cognitive Neuropsychology* 14: 177–208.

Carroll, J.B., and M.N. White. 1973. Word frequency and age-of-acquisition as determiners of picture-naming latency. *Quarterly Journal of Experimental Psychology* 25: 85–95.

Centre International d'études pédagogiques. 2012. Diplôme d'études en langue française DELF and Diplôme approfondi de langue française DALF. http://www.ciep.fr/delfdalf/index.php. Accessed 5 Jan 2012.

Channell, J. 1990. Vocabulary acquisition and the mental lexicon. In *Meaning and lexicography*, ed. J. Tomasczyk and B. Lewandowska-Tomasczyk, 21–30. Amsterdam: John Benjamins.

Costa, A. 2005. Lexical access in bilingual production. In *Handbook of bilingualism: Psycholinguistic approaches*, ed. J.F. Kroll and A.M.B. De Groot, 308–325. Cary: Oxford University Press.

Costa, A., M. Miozzo, and A. Caramazza. 1999. Lexical selection in bilinguals: Do words in the bilingual's two lexicons compete for selection? *Journal of Memory and Language* 41: 365–397.

Cuetos, F., A.W. Ellis, and B. Alvarez. 1999. Naming times for the Snodgrass and Vanderwart pictures in Spanish. *Behavior Research Methods, Instruments, and Computers* 31: 650–658.

Cuetos, F., G. Aguado, C. Izura, and A.W. Ellis. 2002. Aphasic naming in Spanish: Predictors and errors. *Brain and Language* 82: 344–365.

Cycowicz, Y.M., D. Friedman, M. Rothstein, and J.G. Snodgrass. 1997. Picture naming by young children: Norms for name agreement, familiarity, and visual complexity. *Journal of Experimental Child Psychology* 65: 171–237.

De Bot, K. 1992. A bilingual production model: Levelt's speaking model adapted. *Applied Linguistics* 13: 1–24.

Dijkstra, T. 2005. Bilingual visual word recognition and lexical access. In *Handbook of bilingualism: Psycholinguistic approaches*, ed. J.F. Kroll and A.M.B. De Groot, 179–201. Cary: Oxford University Press.

Echeverría, M.S., P. Urzúa, and I. Figueroa. 2005. *Dispogen II. Programa computacional para el análisis de la disponibilidad léxica*. Concepción: Universidad de Concepción.

Educational Testing Service. 2012. Test of English for international communication TOEIC. http://www.ets.org/toeic. Accessed 5 Jan 2012.

Ellis, A.W., and C.M. Morrison. 1998. Real age of acquisition effects in lexical retrieval. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 24: 515–523.

Ferreira, R.A. 2006. Disponibilidad léxica en inglés como lengua materna e inglés como lengua extranjera: Estudio del léxico disponible desde un enfoque psicolingüístico. Unpublished dissertation. Concepción: University of Concepción.
Ferreira, R.A. 2011. Learning the meaning of new words: Behavioural and neuroimaging evidence. Unpublished doctoral dissertation. York: University of York.
Ferreira, R.A., and M.S. Echeverría. 2010. Redes semánticas en el léxico disponible de inglés L1 e inglés LE. *Onomazein* 21: 133–153.
Hernández-Muñoz, N., C. Izura, and A.W. Ellis. 2006. Cognitive aspects of lexical availability. *European Journal of Cognitive Psychology* 18: 730–755.
Instituto Cervantes. 2012. Diplomas de español como lengua extranjera. http://diplomas.cervantes.es/es/informacion/inscripcion_fechas_examen_dele.html. Accessed 5 Jan 2012.
Kucĕra, H., and W.N. Francis. 1967. *Computational analysis of present-day American English*. Providence: Brown University Press.
Larochelle, S., and H. Pineau. 1994. Determinants of response times in the category verification task. *Journal of Memory and Language* 33: 796–823.
Levelt, W.J.M. 1989. *Speaking: From intention to articulation*. Cambridge, MA: MIT Press.
López-Chávez, J., and C. Strassburguer-Frías. 1991. Un modelo para el cálculo del índice de disponibilidad léxica individual. In *La enseñanza del español como lengua materna*, ed. H. López-Morales, 91–112. Río Piedras: Universidad de Puerto Rico.
López-Morales, H. 2012. EL Proyecto Panhispánico. http://www.dispolex.com/. Accessed 12 June 2013.
Malt, B.C., and E.E. Smith. 1982. The role of familiarity in determining typicality. *Memory and Cognition* 10: 69–75.
McCarthy, M., and F. O'Dell. 2002. *English vocabulary in use: Advanced*. Cambridge: Cambridge University Press.
Microsoft Corporation. 2007. Windows XP note block. http://windows.microsoft.com/es-XL/windows/products/windows-xp. Accessed 3 Apr 2007.
Monaghan, J., and A.W. Ellis. 2002. What, exactly, interacts with spelling-sound consistency in word naming? *Journal of Experimental Psychology: Learning, Memory, and Cognition* 28: 183–206.
Morrison, C.M., and A.W. Ellis. 2000. Real age of acquisition effects in word naming and lexical decision. *British Journal of Psychology* 91: 167–180.
Poulisse, N. 1999. *Slips of the tongue: Speech errors in first and second language production*. Amsterdam: Benjamins.
Pye, G. 2002. *Vocabulary in practice 1*. Cambridge: Cambridge University Press.
Redman, S. 2001. *English vocabulary in use: Pre-intermediate and intermediate*. Cambridge: Cambridge University Press.
Richards, C.J., and C. Sandy. 2008. *Passages student's book*, 2nd ed. New York: Cambridge University Press.
Richards, C.J., J. Hull, and S. Proctor. 2005. *Interchange*, 3rd ed. New York: Cambridge University Press.
Snodgrass, J.G., and T. Yuditsky. 1996. Naming times for the Snodgrass and Vanderwart pictures. *Behavior Research Methods, Instruments, and Computers* 28: 516–536.
The Math Works Inc. 2005. MATLAB version 7.1. Natick, Microsoft Corporation.
Thorndike, E.L. 1921. *A teacher's word book of 20,000 words*. New York: Teacher's College.

Chapter 3
The Effect of Age on EFL Learners' Lexical Availability: Word Responses to the Cue Words 'Town' and 'Countryside'

Rosa María Jiménez Catalán, María del Pilar Agustín Llach, Almudena Fernández Fontecha, and Andrés Canga Alonso

3.1 Introduction

Age is a short word that conveys multiple meanings. According to its definition in the *Longman Dictionary of English Language and Culture*, age is 'the period of time someone has been alive or something has existed (1992: 19)'. It is usually represented in plain figures as in 'he is 12', or 'she is 90' but figures in themselves do not convey much information. We know that 25 is older than 12 and that being older is related to more experience and knowledge in certain facets of life as well as more participation in society. As Eckert (1998) notes:

> Aging is central to human experience. It is the achievement of physical and social capacities and skills, a continual unfolding of the individual's participation in the world, construction of personal history, and movement through the history of the community and of society (p. 151).

Relating Eckert's ideas to foreign language education, it could be said that age is among those factors that have a bearing on language learning. It has to do with the accomplishment of capacities and skills in the target language, and the manifestation of learners' participation in the community of English speakers and users.

There are different approaches to the study of learners' performance in a target language; in this chapter, we make use of a lexical availability task to explore the relationship between age and learners' lexical production on two specific semantic domains: 'Town' and 'Countryside'. Our aim is to ascertain whether there are differences or similarities in the number and in the characteristics of the words that young and adult EFL learners retrieve in relation to those domains. Either the 'Town'

R.M. Jiménez Catalán (✉) • M.P. Agustín Llach • A. Fernández Fontecha • A. Canga Alonso
Departamento de Filologías Modernas. Facultad de Letras y Educación, Universidad de La Rioja, C/San José de Calasanz s/n, 26004 Logroño, La Rioja, Spain
e-mail: rosa.jimenez@unirioja.es; maria-del-pilar.agustin@unirioja.es; almudena.fernandez@unirioja.es; andres.canga@unirioja.es

R.M. Jiménez Catalán (ed.), *Lexical Availability in English and Spanish as a Second Language*, Educational Linguistics 17, DOI 10.1007/978-94-007-7158-1_3,

or the 'Countryside' is the scenario in which schools are located. The study of the words that come to EFL learners' minds in response to cue words representing those domains may provide us with some insights into the bonds learners establish with those scenarios. Likewise, it may offer useful clues to identify variation in learners' word knowledge. As we will see, children and adult EFL learners share a similar mental representation in their available lexicons related to 'Town' and 'Countryside', as shown by their word responses to these prompts; however, differences also emerge that may have to do with age in the sense of experience as defined in the dictionary, but also with age in the sense of cognitive capacities, skills and construction of identity suggested in Eckert's quotation above.

3.2 Background

In this section we briefly review language variation according to age in three fields: sociolinguistics, second language acquisition (SLA), and lexical availability. In our review, we pay particular attention to research on young and adult foreign language learners.

3.2.1 The Age Factor in Sociolinguistics

In sociolinguistics, age is a factor of language variation. Together with ethnicity, class, status, and gender, age is among the variables that can introduce variation in the way people use language. In this field, age has been mostly studied from a contrastive perspective, in which broad categories such as adolescents versus adults are compared. Wardhaugh (1996) points to age-differences in English language. In his view, "younger speakers can be observed to use the language differently from older speakers" (196); age variation occurs at all levels of language such as the pronunciation of certain vowels and diphthongs but particularly at the creation of words and expressions. According to Wardhaugh, the reasons for age difference lie in how society is organised:

> There is a variety of possible relationships between language and society. One is that social structure may either influence or determine linguistic structure and/or behaviour. Certain evidence may be adduced to support this view: the *age-grading* phenomenon, whereby young children speak differently from other children and, in turn, children speak differently from mature adults. (10) (emphasis in the original)

The *age-grading* phenomenon is usually linked to slang in teenagers, which serves to identify them as members of the peer group. As Rickford (1996) notes, this phenomenon can be considered an in-group marker. These specific features tend either to disappear as the group becomes older or establish innovative changes in language.

3.2.2 The Age Factor in SLA

Age is a recurrent variable in SLA research both in natural contexts, and in formal or school contexts. Natural and formal contexts present different characteristics which often stand in literature in opposition, the former referring to natural contexts, the latter to school contexts. Most research on natural contexts has focused on brain lateralization and critical period hypothesis (CPH) as well as on memory capacity of young and older learners. In contrast, in school contexts, research on age-related differences has paid more attention to the relationship between starting age and language performance. In this respect, Yamada et al. (1980) claim the supremacy of young learners over older learners as far as rote-memory and motor ability are concerned: they exhibit a greater ability to pronounce and imitate sounds. However, other studies have shown that older learners are faster in the acquisition of most linguistic aspects, and also more efficient in the fulfilment of school tasks (Cenoz 2002; Lightbown 2008; Muñoz 2008).

Three main factors have been attributed to age-related differences in L2 acquisition: neurological factors e.g., the (CPH), affective factors related to motivation, with younger learners displaying higher levels of intrinsic motivation (Lightbown 2008; Muñoz 2006; MacIntyre et al. 2002; García Mayo and García Lecumberri 2003; Cenoz 2003), and cognitive factors (Singleton 2003; Cenoz 2002; Lightbown 2008; Muñoz 2000, 2008). As for this last issue, research has shown that older learners are more mature cognitively and linguistically. Muñoz (2001: 262–263) observes that academic mastery of the L1 displayed by older learners favours L2 learning see also Lightbown (2008: 14). Likewise, the superior cognitive development of older learners has proved to benefit this group as cognitive development goes hand in hand with the kind of tasks learners are capable of performing: understanding the nature of learning tasks as well as choosing the appropriate strategies to accomplish them is of paramount importance for language learning in formal contexts (cf. Muñoz 2001, 2008).

3.2.3 The Age Factor in Lexical Availability Studies

Most studies on lexical availability –both in L1 and in L2– have looked at the effect of social class (Dimitrijević 1969), geographical location monolingual versus bilingual areas (Bailey 1971), rural versus 'Town' speakers' origin (Dimitrijević 1969), type of school instruction (private versus state, bilingual versus regular instruction (Germany and Cartes 2000; Jiménez Catalán and Ojeda Alba 2009a)), or sex (Hernández Muñoz 2010; Jiménez Catalán and Ojeda Alba 2009b). Other studies have looked at the effect of intelligence or family background (Dimitrijević 1969). Few studies have dealt with the effect of age from the perspectives undertaken in SLA studies. However, this does not mean that age has not been addressed in lexical availability studies: it has, but most research has been carried out under the label of 'school grade' rather than 'critical period' or 'starting age' (See for instance Carcedo 1997–1998; Samper Hernández (2002); Echeverría 1991). Whereas in SLA, the age

factor has tended to be investigated from a cognitive approach, in lexical availability research this variable has tended to be addressed from a sociolinguistics stance. In other words, regarding the age factor, lexical availability research is closer to sociolinguistics than to psycholinguistics and most have looked at Spanish as L1. An example of this trend is found in the study by Carcedo (1997–1998), in which he looks at the lexical availability of Finnish and Swedish learners of Spanish. His findings are particularly relevant for the present study as he compares primary school and university learners of Spanish. Carcedo reports an increase in the number of word types according to grade/age; however, this scholar also reports similarities in two respects: on the one hand, concerning the productivity of the cue words, as some are more productive than others both for secondary school students and for university students; on the other, concerning the first five responses to the prompts provided by all groups. These findings mirror the ones obtained by Samper Hernández (2002) with L1 Spanish primary and secondary school students in Canary Islands. According to Carcedo, the striking similarities observed in learners of different levels/ages are due to the universal processes that underline second language acquisition and the organization of L2 mental lexicon.

In our view, there are two arguments in favour of Carcedo's interpretation: the first argument is that identical results are attained in Spanish as L1 and L2 studies; the second is that the same task was used in data collection and learners were allowed an equal amount of response time for each prompt included in the task. However, at this point, a question not answered yet is whether similar or different results would come up in learners of other target languages. In this regard, a question not posed so far in lexical availability studies is whether holding the variable language level constant, children and adult foreign language learners will behave the same or differently in a lexical availability task. Our study aims to contribute to this line of research by combining quantitative and qualitative analyses in the study of the lexical availability of English foreign language learners (EFL) of the same level but different group age. The questions we address are as follows:

1. Will adult EFL learners retrieve a higher number of words than young learners in response to the prompts 'Town' and 'Countryside'?
2. Do adult and young learners activate the same or different words when asked to write responses to the prompts 'Town' and 'Countryside'?
3. Given the same level but different age, will there be variation in the characteristics of the word responses provided by the two groups?

3.3 Method

3.3.1 Participants

In this exploratory study we look at English word responses to a lexical availability task accomplished by 26 Spanish students, learners of English as a foreign language (EFL). The sample is distributed into two groups of different age but same

vocabulary level: thirteen 11–12 year-olds from 6th primary education in a medium size city in La Rioja, northern Spain (children), and thirteen 18–19 year-olds first year students enrolled in the degree of Primary Education at the University of La Rioja (adults). We selected each group out of a larger pool of students on the basis of their scores on a vocabulary test described in the next section. Thus, despite being different in age, the two groups are comparable on the basis of their language level.

3.3.2 Data Collection, Procedures and Analysis

We administered two tests to our informants: a vocabulary level test and lexical availability test. The former was the 2,000 frequency band of the Vocabulary Levels Test (VLT), version 2 designed by Schmitt et al. (2001). This test has been widely used in vocabulary research to place learners' vocabulary level and to draw EFL learners' vocabulary size. The latter consisted of a questionnaire made up of 15 prompts traditionally used in L1 Spanish lexical availability studies. These prompts were: (1) 'Parts of the body', (2) 'Clothes', (3) 'Parts of the house', (4) 'Furniture' Home, (5) 'Food and drink', (6) 'Objects found on the table at meals', (7) 'Kitchen utensils', (8) 'School', (9) 'Town', (10) 'The 'Countryside', (11) 'Means of transport', (12) 'Animals', (13) 'Hobbies', (14) 'Professions', and (15) 'Colours'.

Our sample of EFL learners were given 10 min to complete the VLT and 30 min to respond to the lexical availability questionnaire (2 min per prompt).

The data collection was run by the authors themselves in students' regular classes in two gatherings. We took special care to maintain the same conditions in the two groups such as: (i) giving oral instructions in Spanish before administering the tests, (ii) keeping an identical order in test administration: VLT first, then the lexical availability questionnaire. Regarding the latter, we maintained the same prompts and test format, and encouraged students to write down as many words as came to their minds for each prompt.

Once the data elicitation process was completed, we moved on to encoding students' scores on VLT into an Excel file. This was followed by the editing of word responses, adopting the same criteria as in Jiménez Catalán and Ojeda Alba (2009a, b), who in turn, followed the protocols established in L1 Spanish lexical availability studies. Basically these criteria were: (i) correcting spelling mistakes, (ii) counting repeated words only once per prompt, (iii) discarding unintelligible words and Spanish words, (iv) inserting a hyphen in lexical units containing more than one word (e.g., orange-squash), (v) deleting proper names that have the same spelling in English and Spanish as for instance, Paris, Portugal, but keeping those that are written in a different way in these languages (e.g., New York, London).

We created an electronic plain-text file for each questionnaire, in which we typed the words provided by the students but in their edited form. A lemmatised corpus was built upon with the responses to 'Town' and 'Countryside', and by means of the text analyser WordSmith Tools (Version 6), we counted the words per prompt and grouped them into alphabetical and frequency lists. From our point of view, using this tool is an innovative way of looking at learners' lexical availability; as it allows

combining the use of a lexical availability task with instruments used in corpus studies. The electronic arrangement of learners' words into alphabetical and frequency lists provides researchers with an objective point of reference to compare lexical availability in two age-group learners.

3.4 Results

The first research question was concerned with whether adult EFL learners would retrieve more words than young EFL learners in response to 'Town' and 'Countryside'. In order to answer this question we looked at the responses in both groups. Regarding children, the means of word types were (11.30 sd 5.03) for 'Town' and (9.61 sd 3.42) for 'Countryside'. As for adults, the means were (13.46 sd 2.75) for 'Town' and (10.15 sd 5.71) for 'Countryside'. The descriptive statistics confirmed the differential hypothesis in favour of the adult group: on average, 1st Grade Education students wrote a higher number of words for each prompt than 6th primary school students and they did it for both prompts. However, inferential statistics, specifically a t-test for a sample meeting the normality assumption, applied to the data indicated that differences were non-significant at the $p < .001$ level as proved by the values obtained: for 'Town' ($p = 0.193$) and for 'Countryside' ($p = 0.774$).

In research question two, we aimed to identify the words that adults and children activate in their English lexicon regarding 'Town' and 'Countryside' as to determine whether the words would be the same or different in the two groups. Here we will focus on the comparison of the words accessed by each group, and we will do it on the basis of the number of students who retrieved each word. (Table 3.1 for 'Town') and (Table 3.2 for 'Countryside') display the lists of words produced by each group together with the number of students who generated each word.

In order to gain deeper understanding of the words produced by children and adults in 'Town' and 'Countryside', in the following paragraphs we will group responses into the categories of shared and non-shared vocabulary. In turn, for each category we will classify the words into semantic domains, when they share meaning features. It is important to remark here that the classification of word responses into semantic domains categories obeys practical reasons. As we will see below, there is an overlap among categories: responses such as *sheep*, *pig*, or *cow* could perfectly be categorized under several different but related fields as for instance: 'Fauna', 'Animals', 'Rural life', or 'Domestic animals'.

3.4.1 'Town'

Table 3.1 displays the words retrieved by our informants ranked in descending order by the number of students who retrieved each word. As can be observed, there are shared and non-shared responses between children and adults: 31 words are common to both groups, whereas 47 are exclusive to children and 56 to adults.

Table 3.1 'Town': List of the words retrieved by children and adults distributed according to the number of informants who retrieved each word

Children	**school** (10), **house** (8), **park** (7), **shop** (7), **car (**5), **people** (5), Town hall (5), **tree** (5), **library** (4), **street** (4), **university** (4), **bank (**3), **cinema** (3), post office (3), bookshop (2), **bus (**2), **church** (2), **flat (**2), **person** (2), **restaurant** (2), river (2), theatre (2), traffic lights (2), zoo (2), airport (1), baker's (1), **bar** (1), bench (1), box (1), bread shop (1), **bridge** (1), **building** (1), bus stop (1), butcher's (1), butcher's shop (1), **centre** (1), **chemist** (1), **children** (1), chop shop (1), **factory** (1), **flower** (1), garage (1), grass (1), hill (1), hospital (1), institute (1), lamp street (1), light street (1), light traffic (1), market (1), meat (1), office (1), palace (1), pavement (1), pilot (1), **police** (1), police office (1), port (1) **pub** (1), **road** (1), sea (1), seesaw (1), shop assistant (1), **skyscraper** (1), slide (1), sports centre (1), street bin (1), street lights (1), **supermarket (**1), teacher (1), telephone (1), tent (1), tourist (1), traffic signs (1), vendor (1), window (1), **work** (1), zebra crossing (1)
Adults	**car (**10), **park** (10), **house** (9), **people** (9), **school** (8), **street** (8), **shop** (7), disco (5), **tree** (5), **university** (5), London (3), **restaurant** (3), **road** (3), **bank** (2), bike (2), **building** (2), **children** (2), **cinema (**2), drug (2), friend (2), football team (2),monument (2), museum (2), New York (2), parking (2), party (2), policeman (2), rubbish (2), shopping centre (2), stadium (2), **supermarket** (2), animal (1), avenue (1), **bar** (1), **bridge** (1), bus station (1), **bus** (1), café (1), camp (1), castle (1), cat (1), **centre** (1), charity (1), **chemist** (1), **church** (1), city hall (1), city (1), contamination (1), dog (1), enterprise (1), environment (1), **factory** (1), **flat** (1), **flower** (1), food (1), football (1), garden (1), government (1), gym (1), hairdresser (1), job (1), lake (1), **library** (1), little (1), lorry (1), map (1), market (1), mayor (1), motorbike (19), noise (1), paper (1), **person** (1), **police** (1), polluted (1), priest (1), **pub** (1), **skyscraper** (1), smoking (1), square (1), swimming pool (1), tennis club (1), tourist office (1), train station (1), wall (1), water (1), **work** (1), worker (1)

Table 3.2 'Countryside' list of the words retrieved by children and adults distributed according to the number of informants who retrieved each word

Children	**tree** (10), **flower** (9), **river** (9), **lake** (7), animal (6), **grass** (6), **mountain** (6), **bird** (4), volcano (4), **cow** (3), **dog** (2), eagle (2), **farm** (2), hill (2), **house** (3), hunter (2), **insect** (2), **plant** (2), **pig** (2), sheep (2), bear (1), bridge (1), bull (1), bus (1), **cat** (1), **chicken** (1), elm (1), factory (1), falcon (1), **farmer** (1), fox (1), **fruit** (1), **horse** (1), hut (1), land (1), monkey (1), **mouse** (1), oak (1), path (1), rabbit (1), rain (1), road (1), **rock** (1), snake (1), **snow** (1), street (1), tent (1), tractor (1), train (1), **village** (1), vulture (1), water (1), wolves(1)
Adults	**cow** (7), **river** (6), **bird** (5), **cat** (5), **dog** (5), **flower** (4), **horse** (4), sun (4), **tree** (4), **farm** (3), **farmer** (1), **grass** (3), green (3), **mountain** (3), **plant** (3), sky (3), **chicken** (3), fire (2), fish (2), **pig** (2), **village** (2), air (1), apple (1), beach (1), beautiful (1),bee (1), block (1), bored (1), bread (1), butterfly (1), climbing (1), colour (1), elephant (1), family (1), fly (1), food (1), fresh (1), **fruit** (1), grandma (1), grandpa (1), happiness (1), healthy (1), holiday (1), **house** (1), **insect** (1), **lake** (1), liberty (1), **mouse** (1), nature (1), not (1), old (1), onion (1), parrot (1), people (1), pine (1), pink (1), pork (1), **rock** (1), sea (1), sheep (1), **snow** (1), Spain (1), sweet (1), tomato (1), uncle (1), vegetables (1)

Shared Vocabulary

Regarding the shared responses, the most remarkable feature is that, independently of age, four words were retrieved by most students: *school*, *house*, *park*, and *shop*, and another two, *car* and *tree*, by half of them. However, except for these words, which show a high degree of availability, a considerable number of words were retrieved only once.

The words shared by children and adults can be subsumed in the following semantic fields: facilities (bank, cinema, chemist, church, library, office, restaurant, school, shop, supermarket, and university), means of transport (car, bus, and road), work (factory, police), urban elements (centre, bridge, building, flat, house, skyscraper, street,) or individuals/agents people, (person, children, and tourist).

Non-shared Vocabulary

As to the non-shared words, the exclusive words for children were related to road safety (*traffic lights, zebra crossing, traffic signs, bus stop*), means of transport (*airport, pilot, port*), or education (*teacher, institute*). In contrast, adults' exclusive vocabulary is related to hobbies or sports (*disco, football, museum, swimming, tennis, stadium, gym, hairdresser*), or city names (*London, New York*).

3.4.2 'Countryside'

As shown in the frequency lists (Table 3.2), only 23 words were shared by children and adults. The number of exclusive words in children was 30, compared to 42 in adults.

Concerning the shared words, we note that *tree*, *flower*, *river*, *lake*, *grass*, *mountain*, were common responses in the two groups, although they were retrieved by children more frequently than by adults. In contrast, the number of students who retrieved the words *cow*, *cat*, *dog*, or *horse* was considerably higher in the adult group. As with what happened with 'Town', a great number of words were retrieved only once.

Regarding children's exclusive words, the percentages in the frequency list point to the existence of a wide dispersion in children's responses. Except for *animal* and *volcano*, retrieved, respectively by six and four children, the majority of words that make up children's exclusive vocabulary were yielded by a rather reduced number of informants: a closer look at Table 3.2 reveals that 32 words in children's exclusive vocabulary were retrieved only on one occasion.

The same tendency observed in children's exclusive words is found when we examine adults' exclusive vocabulary. Out of a total of 67 word types, 45 words occur only once. High cohesion in adults' responses is found in the word *cow*, moderate cohesion in *river*, *bird*, *cat*, and *dog*. As to the remaining words in the list, cohesion is rather low or non- existent.

Taking the above observations into account, we can go further to classify children's and adults' shared vocabulary into the following semantic fields:

Fauna: *bird*, *cow*, *dog*, *insect*, *pig*, *cat*, *chicken*, *horse*, *mouse*.
Flora: *tree*, *flower*, *grass*, *plant*, *fruit*.
Nature: *river*, *lake*, *mountain*, *rock*, *snow*.
Rural life: *village*, *farm*, *house*, *farmer*.

As far as children's and adults' exclusive vocabularies are concerned, some patterns come up. Firstly, a high percentage of children produced the superordinate 'animal' but no adult retrieved this word in response to 'Countryside' although, they did so in response to 'Town'. Secondly, both children and adults produced a high number of hyponyms within the category fauna, but the two groups differ considerably as far as the domain these words belong to: children tended to generate hyponyms for wild animals, not found in their real world (although found in cartoons, tales, and movies); in comparison, adults' responses tended to be more related to either domestic animals or to animals closely related to their daily life experience such as *fish, bee, butterfly, chicken,* or *fly*. Thirdly, adults produced a higher number of words related to food than children as for instance, *food,* and *vegetables* as superordinates, *apple, bread, onion, pork,* or *tomato*, as hyponyms.

Finally, in children's production, we detect a series of words which apparently are not so closely related to the typical associations most people make up for 'Countryside': e.g., words connected to means of transport (*bus, train*), or architectural elements (*bridge, factory, street*). On the other hand, in adults' production, we identify words that mostly express aspects or descriptions connected to positive emotions in respect to country life (*happiness, healthy, holiday, liberty, green, beautiful, fresh, not contaminated*) – in this group there is only a word expressing some negative emotion (*bored*) – actions or hobbies (*climbing*), and family relationships (*family, grandma, grandpa*). None of these words were found in children's lexical availability production for this prompt.

3.4.3 Word Classes

In our third research question we aimed to ascertain whether there would be variation concerning the characteristics of the word responses provided by the adult and younger group. In order to answer this question we will pay attention to the following aspects: word class, and word morphology.

As far as word class is concerned, figures in Tables 3.1 and 3.2 uncover a similar pattern for 'Town' and 'Countryside' in the two groups: the words retrieved by young and adult learners are content words rather than grammatical words, and within this category, there is a predominance of nouns in both groups. However, in adults' lexical production, two adjectives are identified in their responses for 'Town' and eight for 'Countryside'. In contrast, no adjectives are found in children's responses to 'Town' or 'Countryside'.

As to the morphology of the words retrieved by our sample of adult and young EFL learners, two features are particularly notorious. Regarding responses to 'Town' the young group produced a considerable number of lexical units rather than single units as for instance *Town hall*, *traffic lights*, or *shop assistant*. Some of them could be regarded as lexical creations for example, *bread shop* and *chop shop*. Adults also retrieved lexical units but they did it in a lower percentage than children.

3.5 Discussion

As to lexical availability production research question (1), descriptive statistics do show an advantage in the number of responses in favour of adult learners; however this advantage does not reach statistical significance. Apparently this result proves that age does not affect lexical availability production, at least as far as responses to 'Town' and 'Countryside' are concerned. Our results contradict the hypothesis held in studies on age-related differences in SLA (e.g., Lightbown 2008; Muñoz 2008), in which it is claimed that older learners are better and more efficient, especially in school settings. The small size of our sample may in part explain the lack of significant differences.

On the other hand, our findings corroborate lexical availability studies and sociolinguistic studies, in which language variation according to age-groups has been reported. A close look at our data uncovers subtle differences between adult and young EFL learners. Among the words retrieved by the adult group, we find quite a number of cognates, most of them, abstract nouns (e.g., *contamination, nature, tomato, liberty, monument*). Younger learners hardly use this strategy and resort instead to other more creative strategies such as word invention (e.g., *chop shop, street bin, light street*) (see below).

However, our results also show the existence of shared vocabulary together with non shared vocabulary in both groups (research questions 2 and 3). Regarding shared vocabulary, it is not surprising that both groups agree in their word responses, since they come from the same place and similar social and cultural backgrounds. However, in our view, there is more than sharing place and backgrounds as similar results are reported in research on lexical availability on Spanish L2 (Carcedo Gonzále 2000; Samper Hernández 2002). In the analysis of the lexical availability of learners of Spanish as L2, the field 'Town' proved to be far more productive than 'Countryside'. In agreement with lexical availability in Spanish as L1 and L2, 'Town' generates more responses than 'Countryside', and the two fields show low cohesion indexes, with 'Town' having a higher index, in general. Compared to 'Countryside', the prompt 'Town' generates a higher number of words and concentrates a higher number of respondents who retrieved identical responses. However, both fields allow for many associative possibilities, they may recall different experiences, emotions and views in adult and young learners, and as a result, they may explain why there are more non-shared than shared words between the two groups.

According to Samper Hernández (2002), the higher productivity of 'Town' compared to 'Countryside' may be related to the urban environment in which her informants were acquiring Spanish as L2: Salamanca, a middle-size Spanish city. However, this explanation does not apply to our informants, as they are learners of English in a foreign language context rather than in second language context. In other words, our informants learn English in the context of the classroom, where English is part of the curricula and not a language spoken in the community for communication in daily life. We are more inclined to believe that the higher productivity of 'Town' compared to 'Countryside' observed in our informants may be due to instruction such as the vocabulary input learners are exposed to in foreign classrooms: 'Town' and 'Countryside' form part of the topics and activities contained in most course books for primary, secondary and tertiary EFL learners.

The most revealing result regarding questions 2 and 3 was that pertaining to the existence of exclusive vocabularies. Our results seem to suggest that adult and young EFL learners differ in their available vocabulary, even if they hold the same language level. We may postulate three types of explanation to account for this: (1) cognitive differences concerning creative processes and learning styles, (2) experiential differences of adult and young learners and (3) instructional differences. Let us discuss each of these aspects in turn.

Regarding cognitive-creative differences, although we do not find many studies in which adult and young EFL learners are compared, we find studies on the language of adolescents in reference to adults' (Eckert 1988; Aitchison and Koppel 1990; Kerswill 1996; Rickford 1996). Research has shown that adolescents make use of creative language strategies such as the invention of words, and language changes as a result. Our sample of young learners was closer to adolescence than to adulthood. The creativeness reported in the studies on the language of adolescents might explain the larger number of invented compounds in children's responses in comparison with adults' responses.

Another explanation for the different words elicited by the lexical availability task may have to do with the learning styles related to each group. Research in L2 has shown that children rely more on rote memory (Yamada et al. 1980); our data seem to corroborate this tendency in the higher number of lexical units retrieved by primary school learners, such as compounds: *post office, bread-shop, traffic lights, zebra crossing*, just to highlight a few, in comparison to adults, where lexical units are not so frequent. This tendency may suggest different patterns of linguistic categorization in child and adult learners. However, care should be taken in this interpretation as Samper Hernández (2002: 37) found invented words in her data of young adults in the field of 'Town' for the following: *basura sitio* (rubbish-place) for *vertedero* (dumping site) for *tiracarta* (letter-throw) or *buzón* (mailbox). Although the author does not report such mechanisms of lexical creation for the field 'Countryside' just as with what happens in our data, the results in both studies might be due to the specific characteristics of the samples as well as to the languages involved.

The use of "chunking" or fixed lexical phrases in the production of young learners is related to unanalyzed language chunks learnt by heart (Boers et al. 2010).

This points to distinctive learning styles but above all, to different ways of conceptualizing words. Our sample of young learners might conceive lexical phrases as single units rather than as the sum of separate words. On the other hand, adult learners might perceive them as separate words, subject to analysis rather than as chunks. Support for this interpretation is found in (Wray 2002), who claims that L2 learning is more difficult for adults, because they have already developed an "analytical" learning style and this hinders their learning of multiword sequences as wholes. As a result, L2 chunks are less likely to be stored (Boers et al. 2010: 244).

Differences in cognitive processing and semantic categorization may also explain the appearance of adjectives in the adult's lexical availability production but not in children's. Adjectives are abstract words; they qualify or describe the quality of things, states, and emotions. Their use implies a higher cognitive development and a different categorization of the world, which has to do with the different experiences adults and children are exposed to.

However, adjectives do not abound in adults' lexical availability production: for each age-group, we obtained mostly nouns in response to the cue-words under examination in the present study. Our finding is in line with most lexical availability studies in Spanish and English as L2, where nouns rather than verbs, adjectives or adverbs have been frequently reported (Carcedo Gonzále 2000; Samper Hernández 2002; Jiménez Catalán and Ojeda Alba 2009a, b). This may be due to the following factors: (a) the stimulus word is a noun and thus nouns are elicited, if it were an adjective or a verb, we could expect different results; (b) nouns are the most readily available vocabulary, the easiest to acquire, and the first to be mastered (López Morales 1992; cf. Ellis and Beaton 1993). As learners grow and develop cognitively, they are capable of categorizing the world by means of different word classes. Our data back this interpretation as in the present study adjectives appear only in the adult production. Maturing processes go hand in hand with the development of categorization, which result in children and adults having undergone different categorization processes (see Mervis 1987). The latter have more experience and linguistic information to complete or restructure their categorization.

Still, a third explanation can account for the different words retrieved by each group of learners: the effect of instruction. We believe that the vocabulary input coming from course books and instructional approaches results in different words available to the learner. In an analysis of the vocabulary input in English course books aimed at 6th primary education and 10th (4 year of secondary education in Spain), (Jiménez Catalán and Mancebo Francisco 2008) found differences in the vocabulary included in course books of different grade (and also within course books of the same grade). Our findings provide evidence of the great disparity observed in children and adults when responding to an identical cue-word. A revealing example is the word *volcano*, produced four times in the Primary school group and none in the University group. Other words that may be motivated by learners' course books may be: *elm, hunter, oak, tent,* or *vulture*. These words are exclusively found in children's production, although they do not show a high degree of availability as they were retrieved by a reduced number of informants. We also find further support in the production of wild animal names by the younger group, which most probably do not belong to

learners' direct experience or world, but might possibly come from instruction. In this regard, our results mirror the ones obtained in English as L1: (Dimitrijević 1969) found a higher number of words of exotic and wild animals than domestic ones in Scottish primary school students' responses to the prompt 'Animals'. Our findings may be due to the high presence of exotic and wild animals in tales, movies, and TV programmes for children.

3.6 Conclusion

This chapter looked at the effect of age on the lexical availability of young and adult EFL learners of the same language level. Age effect has been studied from two different standpoints: number of words, and type of words retrieved by the two groups in response to the prompts 'Town' and 'Countryside' in a lexical availability task.

Regarding the number of words generated by each group, we have seen how overall results and means were higher in the adults' group than in the younger group, but also, how these differences did not prove to be significant. Therefore, we cannot conclude that adult EFL learners have a higher lexical availability than younger learners, at least as far as the prompts analysed are concerned. Our findings in this respect provide counter evidence against the hypothesis held in age-related studies that adult EFL learners will surpass younger learners (e.g., Lightbown 2008; Muñoz 2008). However, care should be taken in the extrapolation of this interpretation to other groups as these findings were based on a rather reduced sample of informants. The small size of our sample may, in part, explain the lack of significant differences.

As for the words generated by adult and young EFL learners, the quantitative and qualitative analyses applied to the data provided us with evidence of the existence of shared and non-shared words in the two groups. Regarding the shared words, *school*, *house*, *park*, and *shop* for 'Town', and *tree*, *flower*, *river*, *lake*, *grass*, *mountain* for 'Countryside' were available to most learners in the two groups. But a high degree of variation in the responses to each prompt was practically the norm rather than the exception. On the one hand, variation appeared in the exclusive vocabulary elicited from each group both for 'Town' and for 'Countryside'. On the other, high idiosyncratic word responses were detected as many words were retrieved only once. This pattern of behaviour was similar in both groups.

The non-shared vocabulary suggests different conceptualization of realities in the minds of the two age-groups examined: whereas young learners' exclusive vocabulary has to do mainly with school issues, adults' exclusive vocabulary refers to hobbies and sports; likewise, we see how the young group retrieved words related to wild animals, usually familiar to them by means of cartoons, tales and movies, whereas the adult group retrieved words referring to domestic animals. In this group we also found abstract words that seemed to be associated to emotions towards 'Countryside', not found in the younger group.

Finally, regarding our third question, we saw that most words retrieved by the two groups, and for the two prompts, were nouns. Although adjectives appeared in adults' production, they were not as frequent as nouns. This finding was not surprising, as practically all lexical availability studies in L1 and L2 have systematically reported the predominance of nouns over other word classes. Overall, these tendencies might be pointing to a different phase in cognitive and affective development in the two groups.

Acknowledgments We would like to acknowledge the financial support of Ministerio de Ciencia e Innovación through grant nº FFI2010-19334.

References

Aitchison, J., and A. Koppel. 1990. *The lexicon in 11–14 year olds. Overview of LSE-Longman project on children's vocabulary*. Manuscript, Language Studies Centre, London School of Economics.

Bailey, V.J. 1971. A study of lexical availability among monolingual-bilingual speakers of Spanish and English. Unpublished thesis, Rice University.

Boers, F., J. Denconick, and S. Lindstromberg. 2010. Choosing motivated chunks for teaching. In *Fostering language teaching efficiency through cognitive linguistics*, ed. Sabine De Knop, Frank Boers, and Antoon De Rycker, 239–257. Berlin/New York: De Gruyter.

Carcedo, González A. 1997–1998. Desarrollo de la competencia léxica en el español LE. *Pragmalingüística* 5–6: 75–94.

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

Cenoz, J. 2002. Age differences in foreign language learning. *ITL Review of Applied Linguistics* 135–136: 125–142.

Cenoz, J. 2003. The influence of age on the acquisition of English: General proficiency, attitudes and code mixing. In *Age and the acquisition of English as a foreign language*, ed. Maria Pilar García Mayo and Maria Luisa García Lecumberri, 77–93. Clevedon: Multilingual Matters.

Dimitrijević, N. 1969. *Lexical availability*. Heidelberg: Julius Groos Verlag.

Echeverría, M.S. 1991. Crecimiento de la disponibilidad léxica en estudiantes chilenos de nivel básico y medio. In *La enseñanza del español como lengua materna*, ed. Humberto López Morales, 61–78. Río Piedras: Universidad de Puerto Rico.

Eckert, P. 1988. Adolescent social structure and the spread of linguistic change. *Language in Society* 17: 183–207.

Eckert, P. 1998. Age as a sociolinguistic variable. In *The handbook of sociolinguistics*, ed. Florian Coulmas, 151–168. London: Blackwell.

Ellis, Nick C., and A. Beaton. 1993. Psycholinguistic determinants of foreign language vocabulary learning. *Language Learning* 43(4): 559–617.

García Mayo, M.P., and M.L. García Lecumberri (eds.). 2003. *Age and the acquisition of English as a foreign language*. Clevedon: Multilingual Matters.

Germany, P., and N. Cartes. 2000. Léxico disponible en inglés como segunda lengua en instrucción personalizada. *Estudios Pedagógicos* 26: 39–50.

Hernández Muñoz, N. 2010. Social aspects of oral and lexical written production in Spanish. *SKY Journal of Linguistics* 23: 101–123.

Jiménez Catalán, R.M., and R. Mancebo Francisco. 2008. Vocabulary input in ELT textbooks. *RESLA* 21: 147–166.

Jiménez Catalán, R.M., and J. Ojeda Alba. 2009a. Disponibilidad léxica en inglés como lengua extranjera en dos tipos de instrucción. *Lenguaje y Textos* 30: 166–176.

Jiménez Catalán, R.M., and J. Ojeda Alba. 2009b. Girls' and boys' lexical availability in EFL. *ITL International Journal of Applied Linguistics* 158: 57–76.

Kerswill, P. 1996. Children, adolescents, and language change. *Language Variation and Change* 8: 177–202.

Lightbown, P.M. 2008. Easy as Pie? Children learning languages. *COPAL Concordia Working Papers in Applied Linguistics* 1: 5–29.

López Morales, H. 1992. En torno al aprendizaje del léxico. Bases psicolingüísticas de la planificación curricular. *RLA: Revista de lingüística teórica y aplicada* 30: 39–50.

MacIntyre, P.D., S.C. Baker, R. Clément, and L.A. Donovan. 2002. Sex and age effects on willingness to communicate, anxiety, perceived competence, and L2 motivation among junior High School French immersion students. *Language Learning* 52(3): 537–564.

Mervis, Carolyn B. 1987. Child-basic object categories and early lexical development. In *Concepts and conceptual development*, ed. Ulrich Neisser, 201–234. Cambridge: Cambridge University Press.

Muñoz, C. 2000. Bilingualism and trilingualism in school students in Catalonia. In *English in Europe. The acquisition of a third language*, ed. Jasone Cenoz and Ulrike Jessner, 157–178. Clevedon: Multilingual Matters.

Muñoz, C. 2001. Factores escolares e individuales en el aprendizaje formal de un idioma extranjero. In *Estudios de Lingüística. Anexo 1. Tendencias y Líneas de Investigación en Adquisición de Segundas Lenguas*, ed. Susana Pastor Cesteros, and Ventura Salazar García. Departamento de Filología Española, Lingüística General y Teoría de la Literatura, Universidad de Alicante.

Muñoz, C. (ed.). 2006. *Age and the rate of foreign language learning*. Clevedon: Multilingual Matters.

Muñoz, C. 2008. Symmetries and asymmetries of age effects in naturalistic and instructed L2 learning. *Applied Linguistics* 29(4): 578–596.

Rickford, J.R. 1996. Regional and social variation. In *Sociolinguistics and language teaching*, ed. Sandra Lee McKay and Nancy H. Hornberger, 151–194. Cambridge/New York: Cambridge University Press.

Samper Hernández, M. 2002. *Disponibilidad léxica en alumnos de español como lengua extranjera*. Málaga: ASELE.

Schmitt, N., D. Schmitt, and C. Clapham. 2001. Developing and exploring the behaviour of two new versions of the Vocabulary Levels Test. *Language Testing* 18(1): 55–89.

Singleton, D. 2003. Critical period or general age factors. In *Age and the acquisition of English as a foreign language*, ed. Maria Pilar García Mayo and M. Luisa García Lecumberri, 3–22. Clevedon: Multilingual Matters.

Wardhaugh, R. 1996. *An introduction to sociolinguistics*. Oxford: Blackwell.

Wray, A. 2002. *Formulaic language and the lexicon*. Cambridge: Cambridge University Press.

Yamada, J., S. Takatsuka, N. Kotake, and J. Kurusu. 1980. On the optimum age for teaching foreign vocabulary to children. *IRAL* 18(3): 245–247.

Chapter 4
The Incidence of Previous Foreign Language Contact in a Lexical Availability Task: A Study of Senior Learners

Francisco Gallardo del Puerto and María Martínez Adrián

4.1 Introduction

This chapter responds to the need for research on three areas of interest in Second Language Acquisition (SLA). More precisely, we address (i) the study of individual differences among learners, such as the incidence of previous Foreign Language (FL) contact; (ii) receptive and productive vocabulary acquisition and (iii) the study of senior learners. Our aim will be to investigate the incidence of FL contact on senior learners' productive vocabulary. A lexical availability task will enable us to discover whether there are any differences between True and False beginner learners in the total number of words produced, in the most productive semantic fields and with respect to the most available words in each semantic field.

4.2 Background

4.2.1 Previous Foreign Language Contact

The study of individual differences in SLA deals with several variables that can affect the acquisition process. With respect to vocabulary, studies dealing with the relationship between lexical range and individual differences are still quite scarce.

F. Gallardo del Puerto
Departamento de Filología, Universidad de Cantabria,
Avda. De los Castros, s/n 39005 Santander, Spain
e-mail: francisco.gallardo@unican.es

M. Martínez Adrián (✉)
Departamento de Filología Inglesa y Alemana. Facultad de Letras,
University of the Basque Country, 01006 Vitoria-Gasteiz, Spain
e-mail: maria.martineza@ehu.es

R.M. Jiménez Catalán (ed.), *Lexical Availability in English and Spanish as a Second Language*, Educational Linguistics 17, DOI 10.1007/978-94-007-7158-1_4,

Most of them have focused on individual differences related to: (i) Second Language (L2) proficiency (Fan 2000; Hanania and Shikhani 1986; Jochems and Montens 1988; Lapkin and Swain 1977), (ii) gender (Agustín Llach et al. 2005; Edelenbos and Vinjé 2000; Jiménez Catalán 2003; Jiménez Catalán and Ojeda Alba 2009; Jiménez Catalán and Terrazas Gallego 2005–2008) and (iii) age, (Cenoz 2002; Miralpeix 2008; Turner 1983). Other variables such as 'previous FL contact' have received scant attention in vocabulary acquisition, except for the study carried out by Martínez Adrián and Gallardo del Puerto (2010), which proved that this variable had a positive effect on the acquisition of both receptive and productive FL vocabulary.

When talking about 'previous FL contact', a distinction between True Beginners (TBs) and False Beginners (FBs) is usually made. TBs could be defined as those students with no previous knowledge of a language when they start learning. FBs, on the contrary, can be defined as students who start to study a language from the beginning again, although they may have a slight knowledge of it (Macmillan Dictionary 2009–2011). FBs have been reported to be more successful at the beginning of university level courses (Halff and Frisbie 1977; Lange et al. 1992; Watt 1997). Nevertheless, these differences between TBs and FBs do not usually persist by the end of courses (Lange et al. 1992) and TBs may even outperform FBs, as the latter do not seem to use the learning strategy of self-correction and therefore, they can make more errors than true beginners (Nakamura 1997). According to Fukai (2000), prior FL learning experience provides background information that makes FBs feel more secure, and thus, FBs are usually less anxious. In this study, Fukai (2000) collected data from two American college students of Japanese, a TB and a FB, by means of the Foreign Language Classroom Anxiety Scale (Horwitz et al. 1986) and semi-structured interviews. A FB in Japanese, who had been taught Japanese at high school for 4 years, reported feeling lucky to have had this prior experience and consequently, was less nervous or anxious in class than the TB. In a more recent study, Frantzen and Sieloff Magnan (2005) compared TBs and FBs learning either L2 French or L2 Spanish. In line with previous research, TBs were reported to express more anxiety than FBs, especially during both processing and output stages. Additionally, TBs received lower grades than FBs. In our study, the impact of the variable 'previous FL contact' on L2 English productive vocabulary acquisition is studied with respect to senior learners, as there is a dearth of research regarding the impact of this variable on vocabulary acquisition. Moreover, most studies addressing the impact of 'previous FL contact' have dealt with secondary students. However, our research is pioneer as it examines a different age group, older adults who have been away from FL study for at least 20 years.

4.2.2 Lexical Availability

According to Carcedo González (1998a), the use of lexical availability tasks allows us (i) to ascertain the available vocabulary of a community of speakers, (ii) to examine the different stages of the acquisition process in the L1 or in a FL, (iii) to make

comparisons between different communities of speakers in the same language or between native speakers and non-native speakers of a given language, and (iv) to have a solid statistical base from which to select those lexical units that have to be taught in the classroom.

Even though lexical availability studies may have these stated advantages, research on lexical availability in a FL is still in its infancy. Studies dealing with L2 Spanish are quite scarce (Carcedo González 1998b; López Rivero 2008; Samper Hernández 2002; Šifrar Kalan 2012, among others) and those dealing with L2 English are even more limited (Jiménez Catalán and Ojeda Alba 2009, 2010; Germany and Cartes 2000). Carcedo González (1998b) analyses the relationship between the variable 'sex' and the lexical availability of 78 Finnish secondary students in L2 Spanish. This author concludes that there are neither quantitative nor qualitative differences between men and women with respect to the total number of words produced and the most available words in a lexical availability task. López Rivero (2008) examines the incidence of the variables 'sex', 'age', 'sociocultural level', 'knowledge of other languages', 'L1', 'length of exposure to L2 Spanish' and 'teaching method followed' on the lexical availability of 43 L2 Spanish students in Madrid. She finds a positive correlation existing among the variables 'sociocultural level', 'knowledge of other languages', 'length of exposure to L2 Spanish', 'sex' and 'age'. Samper Hernández (2002) examined the lexical availability of L2 Spanish learners with a variety of L1s. More recently, Šifrar Kalan (2012) has compared the lexical availability of Slovene students of Spanish as a FL with the lexical availability of native Spanish students so as to detect differences and similarities between foreign and native Spanish speakers in ten semantic categories.

As for L2 English, Jiménez Catalán and Ojeda Alba (2009) test the variable 'sex' in the lexical availability of 210 primary school children. A lexical availability task was designed following the type of research conducted for L1 Spanish Carcedo González (1998a). The results obtained indicate that at the end of primary education, students are capable of generating 109 English words. Additionally, some cue words are more productive than others in eliciting words from students: 'Food and drink' followed by 'Animals' and 'School', all of which supports previous studies on L1 Spanish (García Marcos and Mateo García 1997; Etxebarría Arostegui 1996), as well as on L1 English (Hernández Muñoz et al. 2006) and Spanish and English as a FL (Carcedo González 1998b; Germany and Cartes 2000). With respect to the incidence of the variable 'sex', statistically significant differences were found between girls and boys regarding the average number of words produced by the two sexes, this result being consistent for all cue words. In a similar study, Jiménez Catalán and Ojeda Alba (2010) examined 12 year-old children in Primary Education Grade 6. These learners were divided into two groups taking into account whether they had participated in an immersion programme during the previous two academic years. The immersion group had received 709 h of English, and the other group following traditional instruction had attended English classes for 629 h. Contrary to what was expected, results revealed that the non-immersion group scored higher than the immersion group (119.59 vs. 108.52), even though the difference was not statistically significant. The same result is repeated for practically all

the cue words. The authors also found that the most productive cue words for both groups were 'Food and drink', 'Animals', 'Town' and 'Parts of the body', while the least productive cue words were 'Kitchen and 'Table'. Germany and Cartes (2000) also tested the variable 'type of instruction' in the lexical availability of 100 Chilean secondary students focusing only on the analysis of three cue words: 'Food', 'Human body' and 'House'. 'Human body' turned out to be the most productive cue word, followed by 'Food' and 'House'. Three different types of schools were considered in this study: a bilingual school, a private school and a state school. The bilingual school used English as a vehicular language, whereas the private school employed a communicative method with an integration of the four skills. The state school's approach focused entirely on grammatical aspects of the language, disregarding the communicative perspective. The Bilingual school outperformed both the private school and the state school in the lexical availability task.

4.2.3 The Senior Learner and Second Language Acquisition

As stated in the previous section, the two studies devoted to lexical availability in English in second or foreign languages comprised primary and secondary school learners. To our knowledge, no studies concerning senior learners' lexical availability in English have been conducted so far. Our study will hopefully help initiate research in this area.

Adult learners participate in the domain of Post-initial Education. One of its modalities is Recurrent Education, which refers to older-aged participants at the University of the Third Age (Rogers 1996). In Recurrent Education programmes, older people are provided with opportunities to develop knowledge and skills for survival, discover new role options and enrich their lives. Older learners participate in these educational programmes not for credit or formal recognition, but primarily for immediate application, personal satisfaction and socialization (Blacklock 1985).

Regarding language acquisition, few studies exist on older second language learners (Brandt 1983; Brown 1983; Burling 1981; Zdenek 1986). Older learners seem to succeed in learning another language (Grognet 1997) as their learning ability does not decline with age, even in the face of apparent obstacles such as their heightened fear of failure (Roumani 1978), their doubt that adults can learn a language (Grognet 1997), and their return to a dramatically new educational environment (Singleton and Ryan 2004). If older people remain healthy, their intellectual abilities and skills do not decline (Ostwald and Williams 1981). Even more, as noted by Schleppegrell (1987), especially in the areas of vocabulary and language structure, adults have been found to be better language learners than children (Walsh and Diller 1978). The reason lies in older learners' more highly developed cognitive systems, in their ability to make higher order associations and generalizations, as well as in their capacity to integrate new language input with their substantial learning experience (See also Chap. 3 in this volume). They also rely on long-term memory rather than on the short-term memory function used by children and

younger learners for rote learning (Schleppegrell 1987). More recently, Martínez Adrián and Gallardo del Puerto's (2010) study provided positive support for the ability of the older learner to acquire an FL.

4.3 Research Questions

Taking into account previous findings regarding lexical availability in L2 learners (Carcedo González 1998a, b for Spanish), (Germany and Cartes 2000; Jiménez Catalán and Ojeda Alba 2009, 2010 for English) and the incidence of the variable 'previous FL contact' in L2 acquisition (Fukai 2000; Lange et al. 1992; Nakamura 1997), we addressed the following research questions:

RQ1 Do FBs outperform TBs in the total number of words produced?
RQ2 Are there any differences between TBs and FBs regarding the most productive semantic fields?
RQ3 Are there any differences between TBs and FBs regarding the most available words in each semantic field?

4.4 The Study

4.4.1 Sample

To answer these questions, we gathered data from 18 seniors who were learning L2 English in an institutional context. Due to the low number of participants, this study is exploratory in nature. At the time of data collection, subjects were attending English lessons at the *Aulas de la Experiencia* (University of the Third Age) from the University of the Basque Country for 3 h a week. English was an elective subject from a 3-year degree in Human Sciences. The methodology employed on this course was a communicative-focus-on-form approach (Doughty and Williams 1998), combining the use of a textbook (*Burlington English for Adults 1*) and additional materials prepared by the teacher. As for the language of instruction, all the lessons were conducted in English, except for those occasions when grammar and vocabulary points were not clear enough for the students.

The 18 participants were divided into two research groups depending on their previous FL contact: a group of TBs and a group of FBs. In order to make up those two groups, we took into account the responses provided by the students in a general background questionnaire, as well as the data provided by these learners' instructor, who is also the co-author of this chapter. The TB group was composed of those learners who answered that they had never learned English before, and the FB group was made up of those learners who responded that they had had previous contact with English. Table 4.1 displays the main characteristics of both research

Table 4.1 The sample

	TBs (n = 7)	FBs (n = 11)
Gender	3 male/4 female	3 male/8 female
Mean Age (range)	62.57 (56–68)	61.82 (59–69)
Mean Exposure (range)	115.71 (90–180 h.)	114.55 (90–180 h.)
Previous FL Contact	none	3 years

groups, namely their gender distribution, their mean age and age range, their mean exposure and exposure range, and their previous FL contact.

As can be observed in Table 4.1, the TB group, i.e. those with no previous knowledge of the English language, was composed of three males (42.86 %) and four females (57.14 %), whereas three males (27.27 %) and eight females (72.73 %) made up the FB group. The TB sample was aged between 56 and 68, with a mean of 62.57. The FBs were very much alike in this respect, as their mean age was 61.82 and their ages ranged from 59 to 69. Regarding time spent learning English at the University of the Third Age, both research groups included some learners who had completed one semester, a total of 90 h of classes (71.43 % in the TB group and 72.73 % in the FB group), and other learners who had been studying for two semesters, 180 h (28.57 % in the TB group and 27.27 % in the FB group). The average number of class hours of each group was very similar (115.71 vs. 114.55 h). Additionally, the FB group had had an average of 3 years of previous contact with the FL. It is worth noting that all the false beginners had not studied English for at least 20 years.

4.4.2 Instruments and Procedure

Participants filled in a general background questionnaire where they had to include personal data and to indicate whether they had learned or used English prior to their lessons at the University of the Third Age.

They also performed a lexical availability task based on Carcedo González (1998b) for L2 Spanish. This instrument has been proved to be a reliable measure in L1 Spanish (Alba 1998; Etxebarría Arostegui 1996; López Morales 1999, among many others) and L2 Spanish (Carcedo González 1998b; López Rivero 2008; Samper Hernández 2000) lexical availability. To our knowledge, only three studies have used this type of task with L2 English learners –Germany and Cartes (2000) for Secondary Education learners, and Jiménez Catalán and Ojeda Alba (2009, 2010) for Primary Education learners. Even though the use of this task in L2 English may be scarce and exploratory in nature, it is important to administer this task to a different population, i.e. senior learners, and thus, swell the list of references regarding L2 English.

This task was made up of 15 prompts, each representative of an area of interest (center of interest) related to everyday life: (1) ‘Parts of the body’, (2) ‘Clothes’,

(3) 'House', (4) 'Furniture', (5) 'Food and drink', (6) 'Table', (7) 'Kitchen', (8) 'School', (9) 'Town', (10) 'Countryside', (11) 'Means of transport', (12) 'Animals', (13) 'Hobbies', (14) 'Professions', and (15) 'Colours'. Participants were given a written questionnaire containing the 15 prompts. Each prompt was written on a separate sheet of paper, where participants had 30 lines, on each of which they could write one word. On every sheet of paper with a prompt, participants were asked to write as many words as possible related to the given cue word, in the order that the words came into their heads. They were allowed 2 min for each cue word, and the task lasted 30 min as there were 15 cues.

In order to analyse the data gathered from the lexical availability task, we edited the responses as follows: (i) words with spelling errors were counted, (ii) repeated words were counted only once for each cue word, (iii) unintelligible words or words in other languages were discarded, (iv) words which did not refer to the semantic field where they were included were also rejected, (v) each word received a specific number (from 1 to 30) depending on its position on the sheet of paper provided for each cue prompt, as this would allow us to measure word availability in learners' mental lexicon for each semantic field, words in the first positions being more available than those in the last ones. Afterwards, we typed the responses for each prompt into the computer for processing by the *SPSS* program. Finally, we counted the total number of words produced by each participant for all the prompts in the lexical availability task.

Non-parametric Mann–Whitney tests analyses were conducted in order to compare TBs and FBs with regard to their performance in the lexical availability task.

4.5 Results

In this section, results are shown according to the three research questions of our study. As mentioned in Sect. 4.1, we wanted to discover whether there were differences between true and false beginner learners with regard to three different criteria: (RQ1) the total number of words produced; (RQ2) the semantic fields which were most productive; and (RQ3) the words which were most available to students in each semantic field.

As regards RQ1 (*Do FBs outperform TBs in the total number of words produced?*), Table 4.2 displays the mean number of words, as well as the standard deviations and the minimum and maximum number of words, in each prompt for both research groups. The two columns in bold display the mean number of words produced by TBs on the left and FBs on the right. FBs outperformed TBs in all cases. Overall results indicate that FBs' production was much larger than that of TBs (93.56 vs. 65.86). Similar results were obtained when we looked at the different lexical cues.

When analyses for statistical significance were run, we observed that in two cases (15. 'Colours'; overall results) probability was below .01, which means that differences between the groups were highly significant. Some other comparisons turned out to be significant at below .05 level. This is the case of nine of the

Table 4.2 Mean number of words in each semantic field

	TBs N = 7				FBs N = 11				
	Mean	SD	Min	Max	Mean	SD	Min	Max	Mann–Whitney
1. 'Parts of the body'	3.57	4.04	0	12	5.67	4.07	0	15	−2.048*
2. 'Clothes'	2.29	2.06	0	5	3.94	2.84	0	12	−2.070*
3. 'House'	2.86	2.12	1	6	4.61	2.75	1	10	−2.013*
4. 'Furniture '	4.43	3.31	1	9	5.33	3.09	1	12	−.825
5. 'Food/drink'	6.29	3.64	0	11	9.28	5.05	0	24	−2.138*
6. 'Table'	1.71	2.36	0	6	3.33	2.81	0	9	−2.148*
7. 'Kitchen'	2.71	2.75	0	6	4.72	3.21	0	11	−2.116*
8. 'School'	5.86	2.73	4	11	7.39	2.95	4	13	−1.797#
9. 'Town'	5.57	3.41	0	9	9.33	5.35	0	19	−2.454*
10. 'Countryside'	3.71	3.95	0	12	5.56	4.34	0	17	−1.773#
11. 'Means of Transport'	4.00	1.73	1	6	5.28	1.84	1	8	−2.204*
12. 'Animals'	4.43	2.30	2	8	6.89	3.69	2	17	−2.433*
13. 'Hobbies'	4.14	2.73	2	8	5.50	3.94	0	13	−.961
14. 'Professions'	6.71	1.25	5	9	7.56	2.59	5	13	−.506
15. 'Colours'	7.43	1.72	4	9	8.67	1.68	4	11	−2.655**
OVERALL RESULTS	65.86	26.30	28	108	93.56	39.40	28	189	−2.628**

Note: Statistical significance is indicated at (<.01**) and (<0.5*) levels. Statistical tendency is also marked at (<.09 #) level

prompts (1. 'Parts of the body'; 2. 'Clothes'; 3. 'House'; 5. 'Food and drink'; 6. 'Table'; 7. 'Kitchen'; 9. 'Town'; 11. 'Means of transport'; 12. 'Animals'). Still, in two other cases (8. 'School'; 10. 'Countryside') the comparison yielded a statistical tendency at below .09 level. Therefore, there were just three cases (4. 'Furniture'; 13. 'Hobbies'; 14. 'Professions') in which FBs' superiority was not statistically significant.

In order to be able to find an answer to RQ2 (*Are there any differences between TBs and FBs regarding the most productive semantic fields?*), research groups were compared according to the total number of words, as well as the pertaining percentage shared, in each semantic field. Table 4.3 displays these results, organized in descending order, for each research group. The prompts are ordered from highest to lowest according to the number and percentage of words produced within each semantic field. For clarity we have shaded the highest percentages in light grey and the lowest percentages in dark grey. Intermediate percentages are not shaded.

We observed that the different lexical prompts are equally distributed for TBs and FBs into these three different production categories. Both research groups behaved similarly as regards the prompts which triggered the largest, the shortest and the intermediate number of lexical productions. It is worth noting that the same six semantic fields ('Food and drink', 'School', 'Town', 'Animals', 'Professions', and 'Colours') figured in the top six of both groups. There were also striking similarities regarding the least productive prompts. TBs' four poorest fields ('Clothes', 'House',

Table 4.3 Percentage of words in each semantic field

	TBs N=7			FBs N=11		
	Prompt	% of words	N of words total: 461	Prompt	% of words	N of words total: 1,684
1	'Colours'	11.28	52	'Town'	9.98	168
2	'Professions'	10.20	47	'Food/drink'	9.92	167
3	'Food/drink'	9.54	44	'Colours'	9.26	156
4	'School'	8.89	41	'Professions'	8.08	136
5	'Town'	8.46	39	'School'	7.90	133
6	'Animals'	6.72	31	'Animals'	7.36	124
7	'Furniture'	6.72	31	'Parts of the body'	6.06	102
8	'Hobbies'	6.29	29	'Countryside'	5.94	100
9	'Transport'	6.07	28	'Hobbies'	5.88	99
10	'Countryside'	5.64	26	'Furniture'	5.70	96
11	'Parts of the body'	5.42	25	'Transport'	5.64	95
12	'House'	4.34	20	'Kitchen'	5.05	85
13	'Kitchen'	4.12	19	'House'	4.93	83
14	'Clothes'	3.47	16	'Clothes'	4.22	71
15	'Table'	2.60	12	'Table'	3.56	60

'Table', 'Kitchen') correspond to FBs' lowest positions, and both groups had the prompts 'Clothes' and 'Table', in that order, as the least productive ones.

As far as RQ3 (*Are there any differences between TBs and FBs regarding the most available words in each semantic field?*) is concerned, we compared the five most often cited words in each semantic field for the two groups. Table 4.4 organizes these five words from highest to lowest availability from left to right. Below each word, its mean frequency score in the lexical availability task is displayed. This score ranges from 1 to 30, 1 indicating the highest frequency and 30 the lowest.

The analysis of these data reveals that both research groups showed much agreement in word availability. All semantic fields presented words common to both TBs and FBs. On two occasions ('Clothes' and 'Colours'), the two groups exhibited a hundred per cent agreement, their five most available words being exactly the same. Besides, four matches (out of five) were found in five other semantic fields ('Table', 'Kitchen', 'School', 'Town', 'Means of transport'). Seven other lexical prompts ('House', 'Furniture', 'Food and drink', 'Countryside', 'Animals', 'Hobbies', 'Professions') yielded three shared words. Finally, in only one case, ('Clothes') were there two common items.

Table 4.4 Five most available words in each semantic field

		Word 1	Word 2	Word 3	Word 4	Word 5
1. 'Parts of the body'	**TBs**	*eye*	*hair*	*foot*	*finger*	*leg*
		6.00	15.00	19.14	23.14	23.29
	FBs	*eye*	*foot*	*leg*	*finger*	*hair*
		5.82	9.73	13.64	16.55	19.09
2. 'Clothes'	**TBs**	*jacket*	*shirt*	*slip*	*hat*	*pullover*
		22.57	22.71	23.00	26.71	26.71
	FBs	*shoe*	*shirt*	*trousers*	*skirt*	*jacket*
		11.18	15.45	18.78	20.33	22.09
3. 'House'	**TBs**	*room*	*living room*	*bathroom*	*kitchen*	*window*
		10.71	18.57	23.14	23.71	26.71
	FBs	*window*	*door*	*room*	*kitchen*	*bedroom*
		13.00	14.45	16.18	16.91	18.64
4. 'Furniture'	**TBs**	*table*	*chair*	*bed*	*radio*	*curtain*
		1.86	18.43	22.71	23.29	23.29
	FBs	*table*	*chair*	*bed*	*armchair*	*television*
		2.55	7.27	9.00	13.27	22.09
5. 'Food &drink'	**TBs**	*water*	*orange*	*banana*	*sandwich*	*apple*
		10.71	19.43	19.43	19.71	23.00
	FBs	*wine*	*orange*	*water*	*apple*	*beer*
		8.64	11.73	11.82	16.27	18.73
6. 'Table'	**TBs**	*plate*	*bowl*	*knife*	*glass*	*spoon*
		18.71	18.71	23.00	23.14	26.71
	FBs	*knife*	*plate*	*glass*	*spoon*	*bottle*
		7.82	12.36	13.09	15.82	20.40
7. 'Kitchen'	**TBs**	*table*	*television*	*chair*	*spoon*	*knife*
		14.43	23.00	23.29	26.71	26.86
	FBs	*table*	*chair*	*knife*	*washing-machine*	*spoon*
		2.09	10.18	19.00	20.82	21.55
8. 'School'	**TBs**	*pen*	*book*	*table*	*chair*	*paper*
		6.43	7.71	15.00	15.71	20.29
	FBs	*pen*	*book*	*pencil*	*table*	*paper*
		3.91	6.09	6.64	10.00	17.09
9. 'Town'	**TBs**	*street*	*car*	*bus*	*house*	*cinema*
		10.14	10.29	19.86	23.29	23.29
	FBs	*house*	*street*	*car*	*bus*	*people*
		5.27	9.64	11.55	20.18	22.09
10. 'Countryside'	**TBs**	*tree*	*bird*	*flower*	*green*	*camp*
		14.14	15.00	18.14	18.71	26.86
	FBs	*tree*	*flower*	*green*	*mountain*	*dog*
		8.45	11.28	21.00	21.18	23.55
11. 'Means of Transport'	**TB**	*car*	*bus*	*bicycle*	*train*	*taxi*
		9.86	10.57	15.71	19.29	19.29
	FB	*bicycle*	*bus*	*car*	*train*	*plane*
		4.00	4.18	5.00	13.09	19.09

(continued)

Table 4.4 (continued)

		Word 1	Word 2	Word 3	Word 4	Word 5
12. 'Animals'	**TB**	*dog*	*cat*	*horse*	*bird*	*mouse*
		2.14	10.14	14.71	15.14	19.86
	FB	*dog*	*cat*	*horse*	*snake*	*elephant*
		1.91	4.45	16.36	19.45	19.55
13. 'Hobbies'	**TB**	*football*	*dancing*	*tennis*	*skiing*	*reading*
		10.00	18.43	18.57	22.71	23.00
	FB	*reading*	*football*	*cinema*	*tennis*	*music*
		15.60	20.18	20.82	20.91	21.81
14. 'Professions'	**TB**	*teacher*	*mechanic*	*waiter*	*doctor*	*hairdresser*
		7.14	10.00	12.00	15.86	19.14
	FB	*mechanic*	*doctor*	*hairdresser*	*shop-assistant*	*engineer*
		3.18	18.36	14.50	15.36	16.55
15. 'Colours'	**TB**	*blue*	*green*	*black*	*red*	*white/yellow*
		4.14	6.86	7.29	12.00	12.86
	FB	*black*	*yellow*	*blue*	*green*	*white*
		3.09	4.64	5.55	5.82	6.45

4.6 Discussion

Our study sought to explore the influence of previous foreign language contact on senior learners' acquisition of L2 English at the initial state. More specifically, we analysed three different aspects of these learners' productions in a lexical availability task: the amount of words (RQ1), the productivity of different semantic fields (RQ2), and the available words in each semantic field (RQ3).

Concerning the first aim of the study, we found differences between our two groups of participants, as FBs outperformed TBs significantly in the number of words produced in the lexical availability task, both when analysing the overall task results and when examining most of the individual lexical fields. This finding supports previous studies indicating a better performance by FBs in language learning when compared to TBs (Fukai 2000; Halff and Frisbie 1977; Lange et al. 1992; Frantzen and Sieloff Magnan 2005; Watt 1997). This can be explained by the fact that first year learners who had studied English before had acquired some vocabulary when they first started learning English, and although this had happened over 20 years before recommencing English at the University of the Third Age, they could still recall some English words. Additionally, it could be the case that this initial advantage is caused by lower levels of anxiety shown by FBs owing to their familiarity with the target language, as has been suggested by Fukai (2000) and Frantzen and Sieloff Magnan (2005) for L2 Japanese and L2 Spanish learners, respectively. Although this explanation still remains hypothetical and goes beyond the scope of our study, preliminary analyses of our participants' language anxiety indicate that

students without previous foreign language contact exhibit a significantly higher level of anxiety. Therefore, the potential relationship between participants' lexical availability and anxiety decidedly deserves further thinking and examination in future research on the incidence of the variable previous foreign language contact.

Besides, a comparison of the present study on Adult Education with research on L2 English vocabulary acquisition by primary school children (Jiménez Catalán and Ojeda Alba 2009, 2010), suggests that the rate of vocabulary acquisition seems to be faster in seniors than in children. Jiménez Catalán and Ojeda Alba examined the lexical availability of several groups of 12 year-old children in Grade 6, that is, at the end of their Primary Education after 5 years of English (629–709 h). Results revealed that the mean number of words produced by the different primary school groups ranged between 109 and 120 words. It is worth noting that these rates are very close to the ones obtained by the FB older learners in our study (93.56) after just 1 year of English instruction (180 h). Additionally, even though our TB older learners had received six times less exposure (90 h) than Jiménez Catalán and Ojeda Alba's (2009, 2010) children, they produced more than half the number of words that these children (65.86 vs. 109/120). In the light of these comparisons we cannot but suggest that senior learners seem to acquire English vocabulary faster than schoolchildren, which is in agreement with late starters' faster rate of acquisition at the initial stage (Gallardo del Puerto 2007; García Mayo and García Lecumberri 2003; Muñoz 2006). This finding is also in line with previous research on the older learner (Ostwald and Williams 1981; Schleppegrell 1987) which indicates that adults are usually better language learners than children in the areas of vocabulary and language. It is worth mentioning that age is not a monolithic factor and encompasses a myriad of variables such as type of instruction, motivation, cognitive capacity and world experience, among others, and future research should tackle the influence of these factors independently.

With regard to the second aim of our study, we found that both types of learners (TBs and FBs) showed remarkable similarities regarding productivity in each of the different semantic fields contained in the lexical availability task. Both research groups behaved similarly as far as the cues which triggered the largest ('Food and drink', 'School', 'Town', 'Animals', 'Professions', and 'Colours'), the shortest ('Clothes', 'House', 'Table', and 'Kitchen') and the intermediate ('Parts of the body', 'Furniture', 'Countryside', 'Means of Transport' and 'Hobbies') lexical productions. Thus, like Jiménez Catalán and Ojeda Alba (2009, 2010) and Germany and Cartes (2000) for L2 English, as well as Carcedo González (1998b) for L2 Spanish, we also found that 'Food and drink', 'Animals' and 'Town' were among the most productive cue words, whereas the least productive lexical prompts were 'Kitchen' and 'Table'. This could be accounted for by the fact that 'Kitchen' and 'Table' are semantic fields of a more reduced spectrum than the other fields, as well as having received less attention in the classroom. Hence, the comparison of our senior study with research on children and secondary school students suggests that similarities between children, adolescents and adults are greater than differences as regards their lexical availability in the cue prompts examined.

These findings allow us to suggest that, irrespective of age or previous foreign language contact, beginners go through similar developmental stages in the

acquisition of the lexicon. A possible reason could be that the most productive cue words may be the ones first learnt by students in the process of learning English and that some words are more difficult to learn than others (Jiménez Catalán and Ojeda Alba 2009). This result may also be accounted for by the fact that the same topics are usually dealt with in books designed for adults and children at beginner level. As regards our sample in particular, the fact that both research groups were instructed jointly may have further contributed to homogenizing outcomes in this regard. Nevertheless, this explanation may be disproved in the light of other studies where subjects belong to different instruction environments and they still show much agreement in word availability (Carcedo González 1998a, b; Jiménez Catalán and Ojeda Alba 2009, 2010).

Regarding the third aim of our study, that is, the analysis of words of highest availability in each semantic field, we ascertained that TBs and FBs also displayed striking similarities. All prompts happened to elicit items shared by the two research groups among their five most available words, these coincidences reaching a 60–100 % agreement on 14 of the 15 semantic fields. This finding is also in line with the results obtained in the study by Carcedo González (1998b), which proved there were no qualitative differences, in this case ascribed to sex, with respect to the most available words in the lexical availability task.

Our results may be explained by the fact that these words are among the most basic, and they are the ones that appeared in the textbook used in class at the University of the Third Age (*Burlington English for Adults 1*). Again, it is worth mentioning that the two experimental groups (TBs and FBs) were in the same class, which surely resulted in higher TB-FB agreement in word activation when performing the task.

We believe that this study is a worthwhile addition to the currently scarce amount of literature available on older second language learners (Brandt 1983; Brown 1983; Burling 1981; Zdenek 1986). It also provides supportive evidence that demonstrates the ability of the older learner to learn a language and helps confirm the claim that ageing is not necessarily linked to language learning impairments (Singleton and Ryan 2004). From our teaching experience we know that the profile of the University of the Third Age student is that of a highly motivated language learner (Blacklock 1985; Losada Friend et al. 2007), which may counterbalance both ageing effects and older learners' doubts about their own learning capacity (Grognet 1997). Whatever the case, future research should further disentangle the effects of older people's specificities (e.g., hearing or eyesight loss, cognitive decline) on L2 vocabulary acquisition.

4.7 Conclusion

The main goal of our study was to analyse the effect of the variable 'previous FL contact', operationalized as the comparison between true and false beginner learners, on older adults' acquisition of L2 English productive vocabulary at the initial

state. More specifically, we have investigated whether (i) FBs outperform TBs in the total number of words produced, (ii) whether there are any differences between TBs and FBs regarding the most productive semantic fields and (iii) whether there are any differences between TBs and FBs regarding the most available words in each semantic field.

The results obtained in our study confirm that FBs outperform TBs significantly in the total number of words produced in the lexical availability task, as well as in most of the individual lexical fields contained in the task, which agrees with previous research findings on better FB performance (Fukai 2000; Frantzen and Sieloff Magnan 2005, among others). In line with other studies (Carcedo González 1998b; Germany and Cartes 2000; Jiménez Catalán and Ojeda Alba 2009, 2010), we also found that the productivity of the different semantic fields is very similar for both TBs and FBs. 'Food and drink', 'Animals' and 'Town' are among the most productive cue words, while 'Kitchen' and 'Table' are the least productive prompts. These similarities observed in senior learners and in other populations allow us to suggest that beginners seem to experience similar developmental stages in acquiring lexicon. Regarding the analysis of words of highest availability in each semantic field, there were many matches between the TBs and FBs, which is fully consistent with earlier work on lexical availability (Carcedo González 1998b).

This study also helps swell the list of language learning references concerning the senior learner. The results of the present study indicate that the ability of the older adult to learn a language is not impaired (Singleton and Ryan 2004). Thanks to senior learners' high motivation for language learning (Blacklock 1985; Losada Friend et al. 2007), ageing effects and older learners' doubts about their own learning capacity (Grognet 1997) might not exert as big an impact as thought on the learning process.

For future research, it would be interesting to analyse the relationship between participants' lexical availability and other variables such as the level of anxiety when learning a language, as TBs have been reported to present higher levels of anxiety than FBs (Fukai 2000). In addition, as the two experimental groups were from the same class, it would be worth investigating whether the results would be similar when comparing participants enrolled in different classes. Finally, it would also be useful to explore the impact of specificities in the elderly (cognitive decline, hearing loss, eyesight alterations, among others) on L2 vocabulary acquisition, as well as to analyse which of these features are of greater effect on vocabulary learning.

References

Agustín Llach, M.P., A. Fernández Fontecha, and S. Moreno Espinosa. 2005. Responding to different composition topics: A quantitative analysis of lexical error production. *Glosas Didácticas. Revista Electrónica Internacional de Didáctica de las Lenguas y sus Culturas* 13: 129–141.

Alba, O. 1998. Variable léxica y dialectología hispánica. *La Torre. Revista de la Universidad de Puerto Rico* 7–8: 299–316.

Blacklock, K.L. 1985. Lifelong learning for the older adult. *Journal of Extension* 233: 1–4.
Brandt, H. 1983. The acquisition of German as a second language by two adult learners: A case study. Unpublished doctoral dissertation. Los Angeles: University of Southern California.
Brown, C. 1983. The distinguishing characteristics of the older adult second language learner. Unpublished doctoral dissertation. Los Angeles: University of Southern California.
Burling, R. 1981. Social constraints on adult language learning in native language and foreign language acquisition. *Annals of the New York Academy of Sciences* 379: 279–290.
Carcedo González, A. 1998a. Sobre las pruebas de disponibilidad léxica para estudiantes de español/LE. *RILCE* 14(2): 205–224.
Carcedo González, A. 1998b. Tradición y novedad en las aportaciones hispánicas a los estudios de disponibilidad léxica. *Lingüística* 10: 5–68.
Cenoz, J. 2002. Age differences in foreign language learning. *ITL Review of Applied Linguistics* 135–6: 125–142.
Doughty, C., and J. Williams (eds.). 1998. *Focus on form in classroom second language acquisition*. Cambridge: Cambridge University Press.
Edelenbos, P., and M.P. Vinjé. 2000. The assessment of a foreign language at the end of primary elementary education. *Language Testing* 17(2): 144–162.
Etxebarría Arostegui, M. 1996. Disponibilidad léxica en escolares del País Vasco: variación sociolingüística y modelos de enseñanza bilingüe. *Revista Española de Lingüística* 262: 301–325.
Fan, M. 2000. How big is the gap and how to narrow it? An investigation into the active and passive vocabulary knowledge of L2 learners. *RELC Journal* 312: 105–119.
Frantzen, D., and S. Sieloff Magnan. 2005. Anxiety and the true beginner-false beginner dynamic in beginning French and Spanish classes. *Foreign Language Annals* 382: 171–186.
Fukai, M. 2000. College Japanese classroom anxiety: True vs. false beginners. *Paper presented at the annual meeting of the Midwest conference of Asian affairs*, Bloomington, October 2000.
Gallardo del Puerto, F. 2007. On the effectiveness of early foreign language instruction in school contexts. In *Fremdsprachenkompetenzen für ein wachsendes Europa. Das Leitziel "Multiliteralität"*, ed. D. Elsner, L. Küster, and B. Viebrock, 215–229. Frankfurt/Main: Peter Lang.
García Marcos, F.J., and M.V. Mateo García. 1997. Resultados de las encuestas sobre disponibilidad léxica realizadas en Almería. *REALE: Revista de Estudios de Adquisición de la Lengua Española* 7: 57–68.
García Mayo, M.P., and M.L. García Lecumberri (eds.). 2003. *Age and the acquisition of English as a foreign language*. Clevedon: Multilingual Matters.
Germany, P., and N. Cartes. 2000. Léxico disponible en inglés como segunda lengua en instrucción personalizada. *Estudios Pedagógicos* 26: 39–50.
Grognet, A. G. 1997. *Elderly refugees and language learning*. Washington, DC: Center for Applied Linguistics. (ERIC No. ED416721).
Halff, N.E., and D.A. Frisbie. 1977. College foreign language study related to high school foreign language study. *Modern Language Journal* 61: 401–406.
Hanania, E., and M. Shikhani. 1986. Interrelationships among three tests of language proficiency: Standardized ESL, cloze and writing. *TESOL Quaterly* 201: 97–100.
Hernández Muñoz, N., C. Izura, and A. Ellis. 2006. Cognitive aspects of lexical availability. *European Journal of Cognitive Psychology* 185: 730–755.
Horwitz, E.K., M.B. Horwitz, and J. Cope. 1986. Foreign language classroom anxiety. *Modern Language Journal* 70: 125–132.
Jiménez Catalán, R.M. 2003. Sex differences in L2 vocabulary learning strategies. *International Journal of Applied Linguistics* 131: 54–78.
Jiménez Catalán, R.M., and J. Ojeda Alba. 2009. Girls' and boys' lexical availability in EFL. *ITL, International Journal of Applied Linguistics* 158: 57–76.
Jiménez Catalán, R.M., and J. Ojeda Alba. 2010. Disponibilidad léxica en inglés como lengua extranjera. *Lenguaje y Textos* 30: 167–176.
Jiménez Catalán, R.M., and M. Terrazas Gallego. 2005–2008. The receptive vocabulary of English foreign language young learners. *Journal of English Studies* 5–6: 173–191.

Jochems, W., and F. Montens. 1988. The multiple choice cloze test as a general language proficiency test. *ITL Review of Applied Linguistics* 81: 139–159.
Lange, D., D. Prior, and W. Sims. 1992. Prior instruction, equivalency formulas and functional proficiency. Examining the problem of secondary school-college articulation. *Modern Language Journal* 76: 284–294.
Lapkin, S., and M. Swain. 1977. The use of English and French cloze tests in a bilingual education program evaluation: Validity and error analysis. *Language Learning* 272: 279–314.
López Morales, H. 1999. *Léxico disponible de Puerto Rico*. Madrid: Arco Libros.
López Rivero, E. 2008. *Estudio de disponibilidad léxica en 43 estudiantes de ELE*. Memoria del Máster de Enseñanza del Español como Lengua Extranjera. Universidad Antonio de Nebrija. http://www.educacion.gob.es/dctm/redele/Material-RedEle/Biblioteca/2008_BV_09/2008_BV_09_2_semestre/2008_BV_09_16Lopez_Rivero.pdf?documentId=0901e72b80e2ad5f
Losada Friend, M., E. Garrido Anes, and J. Bando Domínguez. 2007. Perspectivas y estrategias de innovación en la enseñanza del inglés en el aula de mayores. In *Proceedings of the 30th international AEDEAN conference*, ed. M. Losada Friend, P. Ron Vaz, S. Hernández Santano, and J. Casanova, 1–6. Huelva: Universidad de Huelva.
Martínez Adrián, M., and F. Gallardo del Puerto. 2010. L2 English receptive and productive vocabulary in senior learners. *ITL, International Journal of Applied Linguistics* 160: 112–131.
McMillan Dictionary online. 2009–2011. http://www.macmillandictionary.com/dictionary/british/. Accessed 12 Dec 2011.
Miralpeix, I. 2008. The influence of age on vocabulary acquisition in English as a foreign language. Doctoral dissertation. Barcelona: Universitat de Barcelona. http://tdx.cat/bitstream/handle/10803/1670/01.IMP_CHAPTER_1
Muñoz, C. (ed.). 2006. *Age and the rate of foreign language learning*. Clevedon: Multilingual Matters.
Nakamura, T. 1997. *What makes language learners false beginners?* Washington, DC: ERIC Digest. ED420194.
Ostwald, S.K., and H.Y. Williams. 1981. Optimizing learning in the elderly: A model. *Lifelong Learning* 9(1): 10–13.
Rogers, A. 1996. *Teaching adults*. Buckingham: Open University Press.
Roumani, J. 1978. Foreign language learning for older learners: Problems and approaches. Peace corps, Washington, DC. ERIC Document. http://www.eric.ed.gov/ERICWebPortal/search/detailmini.jsp?_nfpb=true&_&ERICExtSearch_SearchValue_0=ED192627&ERICExtSearch_SearchType_0=no&accno=ED192627. Accessed 10 Jan 2012.
Samper Hernández, M. 2000. *Disponibilidad léxica en alumnos de ELE*. Salamanca: Universidad de Salamanca.
Samper Hernández, M. 2002. *Disponibilidad léxica en alumnos de español como lengua extranjera*. Málaga: ASELE.
Schleppegrell, M. 1987. *The older language learner*. Washington, DC. ERIC Clearinghouse on Language & Linguistics FL No. 016 969.
Šifrar Kalan, M. 2012. Análisis comparative de la disponibilidad léxica en español como lengua extranjera (ELE) y lengua materna (ELM). *Macroele: Revista de Didáctica ELE* 15: 1–19.
Singleton, D., and L. Ryan. 2004. *Language acquisition: The age factor*. Clevedon: Multilingual Matters.
Turner, G. 1983. Teaching French vocabulary: A training study. *Educational Review* 351: 81–88.
Walsh, T.M., and K.C. Diller. 1978. Neurolinguistic foundations to methods of teaching a second language. *International Review of Applied Linguistics* 16: 1–14.
Watt, Y.I. 1997. Attrition and performance in Japanese language courses: A study of articulation between high school and university levels. *Foreign Language Annals* 30: 327–334.
Zdenek, J.W. 1986. A survival course for the older learner. *Hispania* 69: 723–7275.

Chapter 5
Lexical Variation in Learners' Responses to Cue Words: The Effect of Gender

María del Pilar Agustín Llach and Almudena Fernández Fontecha

5.1 Introduction

The field of lexical availability in second or foreign languages is very recent, but rapidly growing. General studies on lexical availability of Spanish and English native speakers have examined the issue from different perspectives and addressed different variables. In the present study, we are interested in exploring the effect of gender on learners' lexical availability. With this purpose in mind, in the following sections, we review the studies and give an account of the main results obtained so far by investigations on gender and second language acquisition.

5.1.1 Gender and Language Acquisition

Gender has been occasionally addressed in L1 and L2 acquisition studies. General tendencies can be identified concerning the connection between this variable and language proficiency. Basically, the literature distinguishes among studies that found no differences for male and female learners, those that found female advantage, and those that present findings where male learners do better than their female peers. A mixture of social, cultural and biological aspects has been brandished as potential explanations for gender differences. Research conducted in the last 20 years identified a male advantage in mathematics and visual-spatial abilities (Hyde and Linn 1988; Linn and Petersen 1985), and a female advantage in different verbal skills (Andreou et al. 2005; Kiss and Nikolov 2005; Stumpf 1995). Regarding

M.P. Agustín Llach (✉) • A. Fernández Fontecha
Departamento de Filologías Modernas. Facultad de Letras y Educación,
Universidad de La Rioja, C/San José de Calasanz s/n, 26004 Logroño, La Rioja, Spain
e-mail: maria-del-pilar.agustin@unirioja.es; almudena.fernandez@unirioja.es

R.M. Jiménez Catalán (ed.), *Lexical Availability in English and Spanish as a Second Language*, Educational Linguistics 17, DOI 10.1007/978-94-007-7158-1_5,

L2 learning, inconclusive results or lack of gender differences are reported in general language achievement (Lin and Wu 2003; Spellerberg 2011), listening comprehension (Bacon 1992) or reading comprehension (Brantmeier 2003; Young and Oxford 1997). Other studies have identified a female advantage in motivation towards foreign language (FL) learning (Fernández Fontecha 2010; Kissau 2006), general language achievement (Wen and Johnson 1997; Schaer and Bader 2003), or L2 listening comprehension (Lin and Wu 2003). Still, some other research studies found male learners outperforming female peers in test performance in L2 (Al-Ohtman 2004), or L2 listening comprehension (Boyle 1987).

5.1.2 Gender and Foreign Language Vocabulary Acquisition

Systematic research on the relationship between gender and FL vocabulary acquisition is still embryonic. Jiménez and Ojeda (2009) classify these studies into those that look at the relationship of gender and vocabulary in different language skills (Agustín Llach 2009; Brantmeier 2003), those that examine gender-related differences in learners' receptive or productive vocabulary knowledge (Jiménez and Terrazas 2005–2008), studies that include other variables such as the different use of vocabulary strategies by sex (Jiménez 2003), or research that looks at the effect of topic, task or test type on males' and females' performance (Takala 2000; Yang 2001; Young and Oxford 1997).

L2 research framed in the same context as the present study yields inconclusive findings. In general, girls outperform boys in the range of types of words and semantic fields (Jiménez and Ojeda 2008; Ojeda and Jiménez 2007) as well as in receptive and productive vocabulary size (Jiménez and Ojeda 2008; Ojeda and Jiménez 2007). Non-statistically significant differences in favour of females are also reported concerning lexical errors (Agustín Llach 2009), receptive vocabulary size (Jiménez and Terrazas 2005–2008), or lexical creations (Agustín Llach 2010). In other studies, however, males outperform females in vocabulary retention and access to translations (Grace 2000) or in a test of academic vocabulary recognition, understanding and use (Scarcella and Zimmerman 1998).

5.1.3 Gender Differences in Lexical Availability

Gender is together with socio-cultural status one of the most frequently examined factors in studies of lexical availability. Mother tongue, L2 proficiency, or language learning context (monolingual vs. bilingual context) are also relevant factors in L2 lexical availability studies. Broadly speaking, within lexical availability studies, be it in the L1 (García Marcos and Mateo García 1997 in Carcedo González 2000, and also in Samper Padilla and Hernández Cabrera 1997) or more recently in the L2 (Carcedo González 2000; Samper Hernández 2002), there is agreement in showing

lack of gender differences in quantitative terms. In this sense, and despite the appearance of specific or incidental differences in particular semantic fields, most authors agree on the little relevance of the gender factor in lexical availability studies. However, Samper Hernández (2002) found female advantage in the number of tokens produced for all semantic fields. Unfortunately, she does not offer inferential statistics to compare male and female productions, and thus generalize results. Additionally, since she observes neither qualitative differences nor differences in the number of types produced, she disregards the gender issue as irrelevant. In opposition to this, Hernández Muñoz (2010) concludes that gender is a differentiating factor in qualitative analysis of lexical availability. That is, when the specific words produced are taken into account.

The findings of the above reviewed studies clearly throw non-conclusive results regarding the gender variable. Accordingly, more research into gender differences and similarities is called for.

Furthermore, previous studies on lexical availability have concentrated on investigating lexical availability of adolescent or adult L2 learners, both at specific stages and longitudinally. We have no knowledge of studies dealing either with younger learners or exploring the development of gender differences. Nevertheless, age may be a relevant factor in identifying gender differences.

With all these considerations in mind, we assume that exploring the evolution of gender differences in the lexical availability of young L2 learners would help improve our understanding of gender and lexical availability. We set out to investigate the following research questions:

1. Are there any gender differences in lexical availability of English as FL (EFL) learners at 6th (end of Primary education) and 9th grade (3rd grade of Spanish Secondary education)?
2. How do gender differences evolve across grades and L2 proficiency?
3. What are the most productive semantic fields in males and females?

5.2 Method

5.2.1 Participants

We analyzed the lexical availability of 190 Spanish learners at two different moments of their EFL learning process. At the first moment of data collection, learners were 12 years old on average and attended 6th grade (end of Primary education) and had received approximately 629 h of instruction in the foreign language; at the second moment, learners were 15 years old on average and attended 9th grade (3rd grade of Spanish Secondary education), and had received around 944 h of instruction in the foreign language. The sample is divided into 106 male learners (55.8 %) and 84 female learners (44.2 %). Learners come from a similar social, cultural and economic background: urban, middle-class.

Table 5.1 Mean values for cue words in 6th grade

Semantic fields	Males–Mean values	Semantic fields	Females–Mean values
'Animals'	10.82	'Food and drink'	12.35
'Food and drink'	10.3	'School'	12.03
'School'	10.17	'Animals'	11.88
'Parts of the body'	9.16	'Parts of the body'	11.18
'Town'	8.18	'Town'	10.3
'Hobbies'	7.36	'Hobbies'	8.47
'Transport'	6.13	'Countryside'	7.26
'Countryside'	6	'Transport'	7
'Professions'	4.81	'Professions'	6.03
Total	72.94	Total	86.5

5.2.2 Instruments and Procedures

To gather the data for this study, we used a lexical availability task out of which we selected the following nine prompts out of the original 15: 'Parts of the body', 'Food and drink', 'School' 'Town', 'Countryside', 'Transport', 'Animals', 'Sports' and 'Professions'. The reason for this selection was that we only had information at the two data collection times from these nine prompts. Learners had to generate as many related words or expressions as possible for each prompt or cue word. Thirty minutes were allotted to complete the task, 2 min per cue word. The number of words produced for each semantic field was looked at.

5.3 Results

Table 5.1 lists the mean values of the responses given by male and female learners for the selected prompts in the 6th grade. Results show that females write significantly more words corresponding to each cue word than their male peers with the exception of 'Hobbies', 'Transport' and 'Animals', for which results are not significant.[1] With minor differences in the scale of the results, males and females coincide in the most and least productive cue words. The most productive for males, from highest to lowest mean values, are 'Animals', 'Food and drink' and 'School'. The four most productive prompts for females – from highest to lowest mean values – are 'Food and drink', 'School' and 'Animals'. The least productive fields for males – from lowest to highest mean values – are 'Professions', 'Countryside' and 'Transport'. The cue word 'Professions' is also the least productive among the females' answers, followed by 'Transport' and 'Countryside'. In the mid positions, we identify the

[1] The Wilcoxon-signed rank test for non-parametric comparison of two related samples was conducted. Significance values for all pairs are $p = .000$.

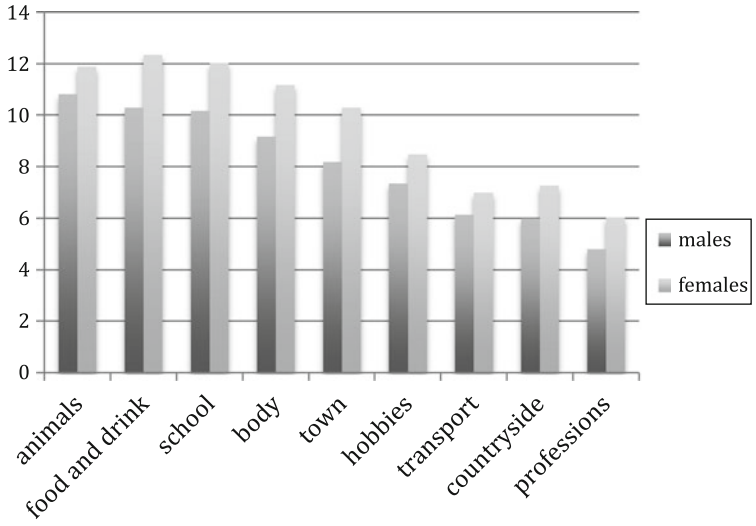


Fig. 5.1 Mean number of responses in 6th grade

Table 5.2 Mean values for cue words in 9th grade

Cue word	Males–Mean values	Cue word	Females–Mean values
'School'	15.83	'Food and drink'	18.72
'Food and drink'	14.63	'School'	17.46
'Animals'	14.16	'Parts of the body'	16.58
'Parts of the body'	14	'Animals'	16.4
'Town'	13.03	'Town'	14.66
'Hobbies'	12.05	'Hobbies'	13.52
'Transport'	9.07	'Transport'	10.7
'Countryside'	8	'Professions'	10.65
'Professions'	4.81	'Countryside'	10.27
Total	110.62	Total	129

same prompts and in the same order for both genders: 'Parts of the body', 'Town' and 'Hobbies'. Figure 5.1 depicts the parallelism in the results of both groups.

Table 5.2 displays the results for the 9th graders. As in 6th grade, the 9th grade girls wrote significantly more words for most prompts than their male peers. Significant results are not observed in the case of 'Transport' and 'Town'.[2] An increase in the total mean value of both groups is perceived.

As in 6th grade, the tendency in the 9th grade is that both groups of students coincide largely in the highest, lowest and mid positions of the mean values ranking.

[2] The Mann–Whitney test for non-parametric comparison of two independent samples was conducted. The p values are as follows: 'Transport': .117, 'Town': .110, 'School': .05, 'Professions': .046, 'Countryside': .019, 'Hobbies': .014, 'Parts of the body': 002, 'Animals': .005, 'Food and drink': .000. Significant level is set at p = .05.

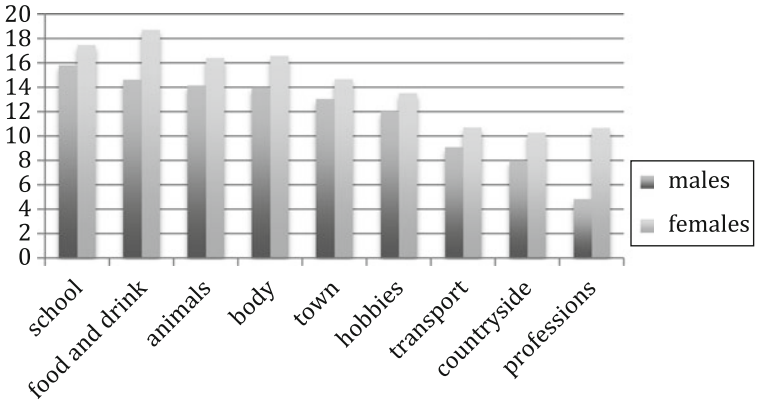


Fig. 5.2 Mean number of responses in 9th grade

From highest to lowest mean values, the prompts that elicit more responses from males are 'School' and 'Food and drink'. The rank of the most productive cue words for girls is 'Food and drink' and 'School'. At the bottom end of the ranking by boys, from highest to lowest values, the prompts 'Transport', 'Countryside' and 'Professions' are found. In the case of females, from the least to the most productive prompts at the bottom of the list we detect 'Countryside' and 'Professions'. In the middle part of the rank, both groups exhibit largely the same results: 'Animals', 'Parts of the body', 'Town' and 'Hobbies'. Figure 5.2 illustrates the large coincidences in the mean value ranking of males and females.

To compare the results obtained by male and female learners in the 6th and 9th grades, we established a three-level ranking of productivity for the prompts under examination: (A.) very productive prompt, (B.) in-between prompt, and (C.) little productive prompt. Results show that 'Animals', 'Food and drink', 'School' and 'Parts of the body' are the most productive prompts for all learners at the two grades; 'Transport', 'Countryside' and 'Professions' are the least productive for both groups at the two collection moments, and finally 'Town' and 'Hobbies' are in the middle see Table 5.3.

Both male and female learners retrieve significantly[3] more words for all prompts in 9th grade than in the 6th. Very few outliers are identified, i.e. learners who perform notably well in the 6th grade but not in the 9th learners 23 and 262 (see Fig. 5.3), or learners who show a remarkable poor performance in the lexical availability test in 6th grade but produce many words in the 9th (learner 168). Figure 5.3 shows a boxplot where learners are depicted and outliers identified. Each circle represents a learner and each learner is located at the intersection between the mean number of words produced for all semantic fields in the 6th grade and that in the 9th grade.

[3] The Wilcoxon-signed rank test for non-parametric comparison of two related samples was conducted. Significance values for all pairs are $p = .000$.

Table 5.3 Comparison of rankings of productivity for prompts in 6th and 9th grades

Availability Ranking		6th Grade		9th Grade	
		Male	Female	Male	Female
(A.) Very productive	1	'Animals'	'Food and drink'	'School'	'Food and drink'
	2	'Food and drink'	'School'	'Food and drink'	'School'
	3	'School'	'Animals'	'Animals'	'Parts of the body'
	4	'Parts of the body'	'Parts of the body'	'Parts of the body'	'Animals'
(B.) In-betweeners	5	'Town'	'Town'	'Town'	'Town'
	6	'Hobbies'	'Hobbies'	'Hobbies'	'Hobbies'
(C.) Little productive	7	'Transport'	'Countryside'	'Professions'	'Transport'
	8	'Countryside'	'Transport'	'Transport'	'Professions'
	9	'Professions'	'Professions'	'Countryside'	'Countryside'

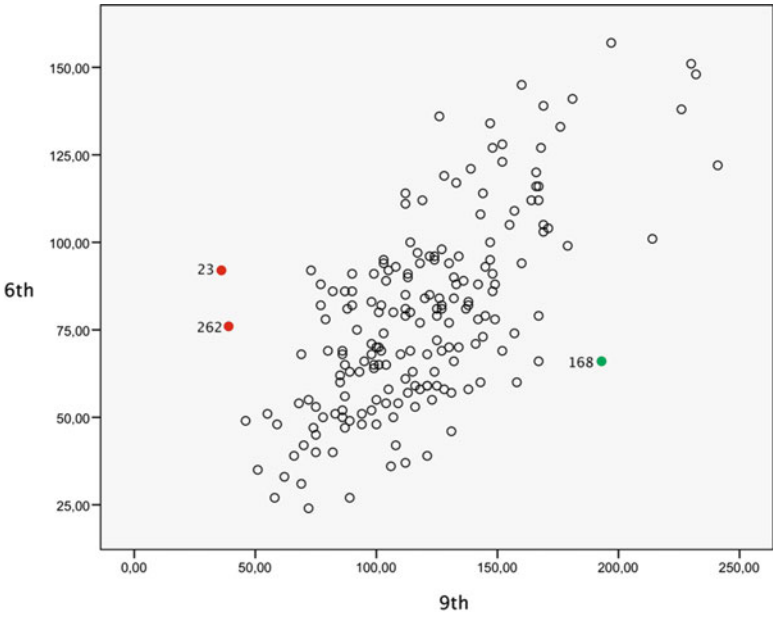


Fig. 5.3 6th and 9th lexical production

5.4 Discussion

In the present study, we set out to find out whether there are gender differences in lexical availability at two different moments and how these differences evolve across grades, as learners get older and more proficient. These two variables co-occur in our design. Analysis of the number of responses provided for each prompt shows that girls produce significantly more responses than their male peers at both testing moments, with the only exception of 'Hobbies', 'Transport'

and 'Animals' for 6th grade and 'Town' and 'Transport' for grade 9. Gender differences in lexical availability were thus found, and these differences are sustained as learners move up school grades and thus grow older and more proficient.

Our results are consistent with the female advantage found in other studies on lexical availability in L1 and L2 (Carcedo González 1998; Fernández Fontecha 2010; Jiménez and Ojeda 2009; Reyes Diaz 1999; Samper Hernández 2002), and productive vocabulary knowledge in L2 (Jiménez 2010). However, our results run against those studies that found no gender-related differences in L2 receptive vocabulary knowledge of learners of similar characteristics to ours (Jiménez and Terrazas 2005–2008). This difference in results leads us to think that gender-related differences are task-dependent (Chavez 2000: 8; Jiménez 2010). In this sense, free-response tasks seem to favour females (cf. Hellekant 1994). Additionally, gender differences might also be skill dependent, since differences were found for productive vocabulary knowledge, but not for receptive vocabulary knowledge as just mentioned above. This is a very similar finding to the one obtained by Jiménez Catalán (2010) in her comparison of learners' performance across vocabulary tests and free production tasks.

Some possible explanations for female advantage may be that girls are more motivated toward FL learning (Jiménez and Ojeda 2009; Mori and Gobel 2006). In addition, Hernández Muñoz (2010) believes that females undergo different cognitive processes as concerns lexical categorization; she also alludes to attitudinal differences, with girls sticking strictly to the task, taking it more seriously and responding to it more formally. In her opinion, cognitive and attitudinal differences help explain gender-related differences in lexical availability studies.

If we look at the cue words' responses where no significant gender differences were found in the 6th grade, that is, 'Animals' and 'Hobbies', we notice that males are especially productive in these fields. They include many different sport types in 'Hobbies', and a wide range of word responses for 'Animals'.

A careful look at our results points to quantitative rather than qualitative gender differences. There is no difference in the particular prompts that elicit the most and least number of responses. In other words, the most and least productive fields are the same for male and female learners at the two data collection moments.

In general, most studies on lexical availability coincide in pointing out differences in the mean values of tokens produced by male and female learners, but find similar ranking in the responses to prompts as concerns their productivity, i.e. the most and least productive prompts (Carcedo González 2000; Samper Hernández 2002). This finding indicates that for both male and female learners the most and least productive vocabulary domains coincide. It might be revealing a similar organization of the available lexicons for male and female learners in the same line that it is suggested by Ferreira Campos and Echeverria this volume regarding Spanish L1 and L2 users.

This result concurs with the findings consistently reported in previous lexical availability studies (Hernández Cabrera and Samper Hernández 2006; Jiménez and Ojeda 2009) that cue words such as 'Food and drink', 'School' or 'Animals' generate a large number of responses and others such as 'Professions', 'Transport'

or 'Countryside' hardly elicit any responses. Jiménez and Ojeda (2009) allude to three possible reasons for this. First, the more exposure to words of some of the semantic fields makes them more productive. Learners, especially young learners, learning English in formal instruction receive more exposure to words within the 'School' or 'Animals' semantic fields than in for instance, 'Town' or 'Countryside'. Second, the interlexical characteristics of some words in particular semantic fields might make them more difficult to learn. Finally, some cue words are more productive, simply because the words belonging to the semantic field they represent may be acquired earlier naturally. For example, the prompts 'Food and drink', 'Animals' and 'School' are easily accessible or available fields, probably because they are part of the young learner's world. Prompts of other semantic fields such as 'Professions' may be less accessible or available to students at this age, probably because this field is not in their immediate surrounding or universe. The most productive fields ('Food and drink', 'Animals', 'School', 'Parts of the body') may include everyday vocabulary, which is frequently used in the learner's world and daily life, words belonging or describing the world around them. In the same line, Ferreira Campos and Echeverria and also Hernández, Izura and Tomé (Chaps. 2 and 10 in this volume) identify exposure, familiarity with the topic, and age of acquisition of the words at stake as key factors in determining the availability values of words and of vocabulary domains in general. The familiarity-with-the-topic factor turns especially relevant if we compare our results with those of Gallardo del Puerto and Martínez Adrián with senior learners (Chap. 5). Older learners clearly show a very different order of productivity for semantic fields, which suggests different lexical availability values for older adults and young adolescent learners (cf. also Chap. 3 by Jiménez Catalán, Agustín Llach, Fernández Fontecha and Canga Alonso). In our view, the tendency observed can most probably be due to the different realities, marked by age, of the two learners' groups.

We may dare speculate still another two reasons to account for the productivity of prompts of certain semantic fields. These results may be a consequence of, firstly, the instruction students have received, or, secondly, it may just be the result of their psychological, cognitive, and linguistic development at the stage of acquisition which resists instruction. Still, it may also be the result of a combination of these two factors, i.e. teaching or pedagogy adapts in topics and linguistic structures to the psychological, cognitive, development of learners.

From a pedagogic point of view, our results suggest that teaching should pay attention to those areas or semantic fields, where learners are least productive. In this sense, explicit vocabulary instruction of words from these fields can contribute to enriching learners' lexical repertoire.

Finally, results also point to significant differences in learners' lexical availability at the two grades under examination, with significant gender differences being maintained. Since vocabulary acquisition is an incremental process both at the receptive (e.g., Agustín Llach and Terrazas Gallego 2013) and at the productive level (Jiménez 2010), the increase in the words available to learners as they grow older is not surprising as shown also by Samper Hernández in Chap. 7. It rather reflects normal lexical development. A more intriguing result concerns the

evolution of gender differences across grades. In relation to this issue, Brantmeier (2004:2) noticed that although at the early stages of language instruction gender may be a "critical factor", it is not at the upper levels. Casey (1996) had already called attention to gender differences decreasing with increasing age as Brantmeier (2004) notes. San Mateo Valdehita (2003–2004) obtained similar results regarding vocabulary learning in Spanish L2. In our data, time span might be too short to appreciate a big reduction of gender differences. Even so, we can observe significance values which are smaller for 9th grade gender differences than for 6th grade. We can thus only speculate that this might be a first sign of the gender factor loosing relevance as learners grow older.

This study serves mainly descriptive purposes as concerns the vocabulary available to Spanish EFL students of the 6th and 9th grades, gender-related differences, and development as age and proficiency increase, rather than having explanatory or predictive power. Therefore, we can only speak of general tendencies observed by our informants. As learners grow older, they have more words available in their lexicon, but while the number of words in each field increases, the order of productivity of the prompts keeps stable from the 6th to the 9th grade.

5.5 Conclusion

It is rare to find lexical availability studies with target languages other than Spanish as L1. However, we think the notion of available vocabulary offers insights into lexical acquisition, and it is worth taking up such studies.

Our study was intended as a first step into covering that gap in research addressing gender differences in lexical availability of young EFL learners. Three main findings stand out in this research. First, females write significantly more responses than their male peers both in 6th grade and 3 years later in 9th grade. Gender differences in the number of responses stay across grades. The tendency though is for significance values to get smaller as learners grow older. Second, male and female learners coincide in the semantic fields for which they produce most and least cued responses and again this result stays over time. Third, learners in grade 9 produce significantly more responses than in grade 6. This shows a bigger lexical availability for older learners in accordance with studies of general vocabulary knowledge, which shows that vocabulary size increases over time and exposure to the FL (cf. Agustín Llach and Terrazas Gallego 2013).

The present results confirm previous findings in lexical availability studies, as reported above. This systematicity in results leads to believe that the lexical availability task is a reliable task that throws consistent results, making it thus a very appropriate measure of lexical development.

Future studies should examine the role of education context on lexical availability. Germany and Cartes (2000) found out that the teaching approach is decisive in lexical availability results of students. Thus, they conclude that learners in bilingual programmes have higher degrees of lexical availability because they use the FL as a

means of communication. It would be very enlightening to follow this line of research (see López in Chap. 9 in this volume).

Further studies with students of an older age which might throw more light on productive semantic fields and lexical availability are also called for from our results. Likewise, comparison of these results with lexical availability studies in L1 may show whether differences in productive semantic fields are due to language issues or rather to psychological issues. Specifically, a qualitative study of the particular words in the semantic fields may be very helpful to see whether gender or age-related differences appear in the specific words used in each semantic field.

Our research does not intend to be exhaustive, but a small contribution to the area of study of gender. In our exploration of gender differences and similarities from the perspective of lexical availability, we have found the lexical availability task to be a very useful tool, since it allows for both quantitative and qualitative data as concerns learners' productive vocabulary knowledge. However, our research is preliminary and clearly needs a qualitative follow-up that would allow for a deeper exploration of the relationship between the gender factor and learners' lexical availability. Certainly, it would also help us in understanding the reasons that lead to the differences found in the present study, and thus being able to explain them better, predict them better, and prevent them, if this were necessary or desirable.

Acknowledgments We would like to acknowledge the financial support of FEDER and Ministerio de Ciencia e Innovación through grant nº FFI2010-19334.

References

Agustín Llach, M.P. 2009. *Gender differences in vocabulary acquisition in the foreign language in primary education: Evidence from lexical errors*, Biblioteca de Investigación 54. Logroño: La Rioja University Press.

Agustín Llach, M.P. 2010. Exploring the role of gender in lexical creations. In *Gender perspectives on vocabulary in foreign languages*, ed. R.M. Jiménez Catalán, 74–92. Basingstoke: Palgrave Macmillan.

Agustín Llach, M.P., and M. Terrazas Gallego. 2013. Vocabulary knowledge development and gender differences in a second language. *ELIA* 22: 45–75.

Al-Ohtman, N.M.A. 2004. The relationship between gender and learning styles in Internet-based teaching – A study from Kuwait. *The Reading Matrix* 41: 38–54.

Andreou, G., F. Vlachos, and E. Andreou. 2005. Affecting factors in second language learning. *Journal of Psycholinguistic Research* 345: 429–438.

Bacon, S. 1992. The relationship between gender, comprehension, processing strategies, and cognitive and affective response in foreign language listening. *The Modern Language Journal* 76: 160–178.

Boyle, J.P. 1987. Sex differences in listening vocabulary. *Language Learning* 372: 273–284.

Brantmeier, C. 2003. Does gender make a difference? Passage content and comprehension in second language reading. *Reading in a Foreign Language* 151: 1–27.

Brantmeier, C. 2004. Gender, violence-oriented passage content and second language reading comprehension. *The Reading Matrix* 42: 1–19.

Carcedo González, A. 1998. Sobre las pruebas de disponibilidad léxica para estudiantes de español LE. *RILCE* 142: 205–224.

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

Casey, M.B. 1996. Gender, sex, and cognition: Considering the interrelationship between biological and environmental factors. *Learning and Individual Differences* 81: 39–53.

Chavez, M. 2000. *Gender in the language classroom*. New York: McGraw-Hill.

Fernández Fontecha, A. 2010. Gender and motivation in EFL vocabulary production. In *Gender perspectives on vocabulary in foreign languages*, ed. R.M. Jiménez Catalán, 93–116. Basingstoke: Palgrave Macmillan.

García Marcos, F., and M.V. Mateo García. 1997. Disponibilidad Léxica y Enseñanza de Lenguas Extranjeras. *Docencia e Investigación*. Cuenca: Universidad de Castilla La Mancha.

Germany, P., and N. Cartes. 2000. Léxico disponible en inglés como segunda lengua en instrucción formalizada. *Revista Estudios Pedagógicos* 26: 39–50.

Grace, C. 2000. Gender differences: Vocabulary retention and access to translations for beginning language learners in CALL. *The Modern Language Journal* 84(2): 214–224.

Hellekant, J. 1994. Are multiple-choice tests unfair to girls? *System* 22(3): 349–352.

Hernández Cabrera, C.E., and M. Samper Hernández. 2006. *Léxico disponible, norma culta y norma popular. In memoriam Manuel Alvar, II.* Rosa M.ª Castañar y José M.ª Enguita, 1341–1358. Zaragoza: Institución "Fernando el Católico".

Hernández Muñoz, N. 2010. Social aspects of oral and written lexical production in Spanish. *SKY Journal of Linguistics* 23: 101–123.

Hyde, J.S., and M.C. Linn. 1988. Gender differences in verbal abilities. A meta-analysis. *Psychological Bulletin* 1041: 53–69.

Jiménez, R.M. 2003. Sex differences in L2 vocabulary learning strategies. *International Journal of Applied Linguistics* 131: 54–77.

Jiménez, R.M. 2010. Gender tendencies in EFL across vocabulary tests. In *Gender perspectives on vocabulary in foreign languages*, ed. R.M. Jiménez Catalán, 117–139. Houndmills/Basingstoke/Hampshire: Palgrave Macmillan.

Jiménez, R.M., and J. Ojeda. 2008. The English vocabulary of boys and girls: Similarities or differences? Evidence from a quantitative study. In *Theoretical and methodological approaches to gender and language study*, ed. L. Litosseliti, H. Sauton, K. Segall, and J. Sunderland, 103–118. Houndmills/Basingstoke/Hampshire/New York: Palgrave Macmillan.

Jiménez, R.M., and J. Ojeda. 2009. Girls' and Boys' lexical availability in English as a foreign language. *ITL International Journal of Applied Linguistics* 158: 57–76.

Jiménez, R.M., and M. Terrazas. 2005–2008. The receptive vocabulary of English foreign language young learners. *Journal of English Studies* 5–6: 173–191.

Kiss, C., and M. Nikolov. 2005. Preparing, piloting and validating an instrument to measure young learners' aptitude. *Language Learning* 551: 99–150.

Kissau, S. 2006. Gender differences in motivation to learn French. *Canadian Modern Language Review* 623: 401–422.

Lin, J., and F. Wu. 2003. Differential performance by gender in foreign language testing. *Poster for the 2003 annual meeting of NCME Chicago*, Chicago.

Linn, M.C., and A.C. Petersen. 1985. Emergence and characterization of sex differences in spatial abilities: A meta-analysis. *Child Development* 56: 1479–1498.

Mori, S., and P. Gobel. 2006. Motivation and gender in the Japanese EFL classroom. *System: An International Journal of Educational Technology and Applied Linguistics* 34(2): 194–210.

Ojeda Alba, J., and R.M. Jiménez Catalán. 2007. The worlds children's words build. *Didáctica (Lengua y Literatura)* 19: 155–172.

Reyes Diaz, M.J. 1999. Acerca de la relación entre la variable sexo y el aprendizaje léxico. *In Contribuciones al estudio de la Lingüística Aplicada. Logroño: Asociación Española de Lingüística Aplicada*, ed. J. de las Cuevas, D. Fasla, 387–391. ISBN: 84-605-8800-9.

Samper Hernández, M. 2002. *Disponibilidad léxica en alumnos de español como lengua extranjera*, Colección Monografías, núm. 4. Málaga: ASELE.

Samper Padilla, J.A., and C.E. Hernández Cabrera. 1997. Estudio de la disponibilidad léxica en Gran Canaria. Datos iniciales y variación sociolingüística. *Contribuciones al estudio de la lingüística hispánica* Homenaje a Ramón Trujillo Carreño, Vol. II, 229–239. La Laguna: Universidad de la Laguna.

San Mateo Valdehita, A. 2003–2004. *Aprendizaje de léxico en español como segunda lengua. Investigación sobre tres métodos.* Unpublished M.Phil. thesis, UNED. Available online at http://www.mec.es/redele/biblioteca2005/san_mateo.shtml

Scarcella, R., and C. Zimmerman. 1998. Academic words and gender. ESL student performance on a test of academic lexicon. *Studies in Second Language Acquisition* 20: 27–49.

Schaer, U., and U. Bader. 2003. *Evaluation Englisch an der Primarschule Projekt 012*. http://www.ai.ch/dl.php/de/20031110083112/Evaluationsbericht+07.11.2003.pdf. Accessed 16 Feb 2010.

Spellerberg, S.M. 2011. L3 English acquisition in Denmark and Greenland: Gender-related tendencies. *International Journal of Multilingualism* 8(3): 155–188.

Stumpf, H. 1995. Gender differences in performance on tests of cognitive abilities: Experimental design issues and empirical results. *Learning and Individual Differences* 7: 275–288.

Takala, S. 2000. Test fairness: A DIF analysis of an L2 vocabulary test. *Language Testing* 173: 323–340.

Wen, Q., and R.K. Johnson. 1997. L2 learner variables and English achievement: A study of tertiary-level English majors in China. *Applied Linguistics* 181: 27–48.

Yang, Y. 2001. Sex and language proficiency level in color-naming performance. An ESL/EFL perspective. *International Journal of Applied Linguistics* 112: 238–256.

Young, D.J., and R. Oxford. 1997. A gender-related analysis of strategies used to process written input in the native language and a foreign language. *Applied Language Learning* 81: 43–73.

Chapter 6
Frequency Profiles of EFL Learners' Lexical Availability

Rosa María Jiménez Catalán and Tess Fitzpatrick

6.1 Introduction

Vocabulary researchers under the Anglo-Saxon tradition[1] (Richards 1976, 1985; Laufer 1991, 1998; Nation 1990, 2001) have regarded word frequency as an important dimension in the assessment of learners' lexical competence. It has been adopted as the basis of vocabulary tests to measure learners' vocabulary knowledge. The underlying assumption is that more frequent words are acquired before less frequent ones, and that knowledge of words in different frequency bands can therefore be regarded as a predictor of lexical competence and, in L2, as a proxy measure for proficiency (Laufer 1998; Nation 2001; Morris and Cobb 2004): the higher the number of infrequent words learners know, the higher their lexical richness (Meara et al. 1997; Nation and Waring 1997). Likewise, word frequency has been used in the analysis of high and low-frequency words included in English courses (Meara 1993), reading texts (Gardner 2004) and classroom transcripts (Meara 1993; Meara et al. 1997; Nation 2001; Horst 2010). With regard to English as a second or foreign language learning and teaching, the assumption of word frequency

[1] By Anglo-Saxon tradition we mean vocabulary research on second or foreign language learners entirely conducted in English, and published by publishing houses and academic journals based on English speaking countries. The term is used in opposition to PanHispanic tradition, where research on second or foreign language learners' vocabulary knowledge has also been conducted but from the perspective of lexical availability rather than frequency. This research is published in Spanish by publishing houses and academic journals based on Spain as well as on Latin American countries and other Spanish speaking countries.

R.M. Jiménez Catalán (✉)
Departamento de Filologías Modernas. Facultad de Letras y Educación, Universidad de La Rioja, C/San José de Calasanz s/n, 26004 Logroño, La Rioja, Spain
e-mail: rosa.jimenez@unirioja.es

T. Fitzpatrick
Cardiff School of English, Communication & Philosophy, Cardiff University, Cardiff, UK
e-mail: fitzpatrickT@cardiff.ac.uk

R.M. Jiménez Catalán (ed.), *Lexical Availability in English and Spanish as a Second Language*, Educational Linguistics 17, DOI 10.1007/978-94-007-7158-1_6,

presupposes that the words most frequently used by English native speakers will be the most useful to learners of that language. In this respect, Nation (2001:139) maintains that "There is a small group of high-frequency words which are very important because these words cover a very large proportion of the running words in spoken and written texts and occur in all kinds of uses of the language". However, although this statement sounds commonsense, and it reflects the role of frequent words in any language, in practice there are situations where in addition to frequency, vocabulary selection needs to take other criteria into account. An example of this is the vocabulary that learners of English as a foreign language (EFL), need in classrooms: words such as *pencil*, *chalk* or *blackboard,* which are not within the most frequent words in English language, but are part of the repertoire of any EFL learner in classroom contexts. As Richards (1974:75) noted more than three decades ago, there are words that are important regardless of not being frequent:

> The words chalk, blackboard, pencil, bath, stomach, soap, umbrella, trousers, scissors, camera, leaf, and cushion, are not within the first 2,000 words … yet such words are often needed within the first few lessons of a language course. Without an adequate range of concrete vocabulary, there can be no content to the language lessons.

However, these criticisms do not invalidate the arguments in favour of paying attention to word frequency: among them, there is the fact that knowledge of the most frequent words in English is essential for text comprehension, with a consensus that 2k words account for 95 % of text coverage (Laufer 1992; and Laufer 1989; Nation 1995; Nation and Waring 1997). Another argument is that word frequency predicts repeated encounters, and according to memory theories (Pimsleur 1967; Scarborough et al. 1977; Lynch and Maclean 2000) repetition is an essential condition of learning. From a pedagogical point of view, word frequency offers teachers, researchers and learners an objective instrument for teaching and assessment. As Alderson (2007) notes, word lists and vocabulary tests based on native speakers' use of words, ranked according to frequency of use, may be used as measure of language proficiency. These strengths could explain the proliferation of productive applications of word frequency information found in vocabulary research, among them: (i) the creation of frequency lists for curriculum design, (ii) the compiling of learners' dictionaries, (iii) the production of graded readers, and (iv) the construction of vocabulary tests.

In this chapter we examine word frequency in the context of lexical acquisition. Using these two measures in tandem will enable us to examine both the quantity of lexical output (lexical availability) and its quality (frequency profile). The rest of the chapter will be structured as follows: in Sect. 6.2, we will address the main assumptions underlying the concepts of word frequency and lexical availability and attempt a comparison of both approaches; then we will offer a brief review of productive vocabulary tests based on frequency. It is important to note here that we do not mean to provide the reader with a comprehensive review of the theoretical conceptualization of word frequency or lexical availability, but rather to highlight the strengths and shortcomings of each, bearing in mind their respective perspectives on English language learners' vocabulary profiles. The study we report in Sects. 6.3 and 6.4 draws on these two research strands, following the hypothesis that both can be used in combination to determine EFL learners' vocabulary profiles.

Our study sets out to identify the words retrieved by a group of learners in a lexical availability task and classify them into frequency bands. In Sect. 6.3, then, we describe our sample of informants, the lexical availability task used and the analyses performed by means of VocabProfile, an electronic vocabulary analyser based on word frequency. In Sect. 6.4 we report our preliminary findings and attempt to interpret them in the context of lexical availability and frequency studies. Our chapter closes with some concluding remarks and an outline of further studies to move forward the line of research undertaken in this preliminary study.

6.2 Background

6.2.1 Word Frequency Versus Lexical Availability

Both word frequency and lexical availability deal with lexical characteristics but they do so from apparently opposed stances: whereas in the former the reference is always the oral or written text, in the latter, the reference is the speaker's mental lexicon. Word frequency research centres its methodology on counting the number of occurrences of words in a corpus, and identifying those which occur more and less frequently. The assumptions underlying its application to learning and assessment are (i) word frequency is related to usefulness: vocabulary selection in course books and dictionaries should therefore be organized according to word frequency (ii) knowledge of the most frequent words in English exponentially facilitates text comprehension (Nation and Waring 1997); most frequent words are acquired before least frequent words (Laufer 1994; Nation and Waring 1997); (iii) knowledge of infrequent words is related to lexical richness of language production, and therefore to language proficiency: a greater knowledge of infrequent words will result in more lexically rich language production (Laufer and Nation 1995; Laufer 1998; Meara et al. 1997; Nation and Waring 1997; Morris and Cobb 2004). 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. Lexical availability or word availability can be defined as the propensity for words to come to our mind in response to a word stimulus. This concept is based on three assumptions: (i) the first word responses are the most available in the learner's lexicon; (ii) the responses provided by the learner reflect the organization of the learner's mental lexicon; (iii) these responses are words related to daily situations, as is reflected in the cues conventionally used in lexical availability tasks.

On the face of it, these approaches have nothing in common except the fact that both attempt to quantify lexical knowledge. However, on closer inspection we find they share a number of important characteristics. In the first place, these concepts come from the same origin, since both emerged in recognition of the importance of lexical knowledge and lexical production to vocabulary teaching, vocabulary selection and vocabulary testing; secondly, in lexical availability studies the data are always classified according to internal frequency; that is, word responses are reported in

rankings arranged according to the most and least produced responses to each prompt contained in the task. Thirdly, lexical availability studies report results on the basis of number of responses to each prompt, and compare them to the ones obtained in other lexical availability studies. This use of a base-line, or normative, corpus for comparison, parallels the use of word frequency lists as a tool for assessment; word frequency is derived from a text-based corpus, and lexical availability from a corpus of responses to set lexical stimuli.

6.2.2 Vocabulary Tests Based on Frequency

Research into vocabulary knowledge has paid relatively scant attention to EFL learners' productive vocabulary. This is partly because researchers have tended to focus on attempts to establish vocabulary size estimates, which can be calculated more straightforwardly by testing receptive rather than productive knowledge, and partly because of the difficulties encountered in designing a valid and reliable test capable of assessing the breadth and depth of learners' vocabulary knowledge. Some attempts have been made, though, to apply word frequency information to the evaluation of learners' productive vocabulary. Perhaps the most well-known of these are the productive version of '*Vocabulary Levels Test*' (Laufer and Nation 1999), *Lex30* (Meara and Fitzpatrick 2000), and *Lexical Frequency Profile* (Laufer and Nation 1995). Word frequency as a key predictor of learners' lexical competence is an idea that underlies these tests. However, there are important differences in them regarding format, the frequency lists underpinning the test, and the nature of the test. Following Jiménez Catalán and Moreno's (2005) vocabulary tests review, which in turn followed the frameworks established by Read (1988, 2000) and Schmidtt (2000), we will list the main distinguishing features of each test as follows: the productive version of the Vocabulary Levels Test is a breadth, discrete, selective, completion and context dependent test. Lex30 is a breadth, discrete, comprehensive, word association and context independent test. As for Lexical Frequency Profile, it is a breadth, discrete, comprehensive, free production, context-dependent test. Although all three look at learners' productive vocabulary, it is the Lexical Frequency Profile that better suits the purposes set out in the present study, as it entails a free (i.e. not controlled) productive task, with a profile-based scoring system, allowing for more fine-grained analysis in contrast to the binary scoring of Lex30, for example. With this in mind, the next paragraphs within this section are devoted to survey studies that have used this test to look at the productive vocabulary of EFL learners.

6.2.3 Lexical Frequency Profile

This test designed by Laufer and Nation (1995) is based on a free computer program that can be run online or be downloaded at http://www.vuw.ac.nz/lals/staff/paul_nation/index.html. Further adaptations are found at http://www.lextutor.ca/.

The program requires the researcher to type learners' texts and save them into ASCILL or Text files. Once the text input is loaded into the program, it provides researchers with the count of types and tokens, the frequency of a word in the text, and the classification of words in the text according to three frequency lists: List 1 includes the most 1,000 frequent words of English, List 2 the 2nd 1,000 most frequent words, and finally, List 3 includes words that are not found in the other two lists but are frequent words at the academic level such as in secondary schools and university texts and materials. The final classification is "off-list": words which do not appear in any of the three lists. The sources for these lists are the *General Service List* (West 1953), and the *Academic Word List* (Coxhead 1998). Alternative versions of this profiling tool have been developed, such as *VocabProfile* and *Range* based on the BNC lists.

VocabProfile has attracted the attention of vocabulary researchers due to its potential application to the analysis of the vocabulary produced by learners in free writing (Laufer 1994; Laufer and Nation 1995; Laufer and Paribakht 1998; Muncie 2002; Morris and Cobb 2004; Horst and Collins 2006), the vocabulary contained in course books and other learning and teaching materials (Meara 1993), and the vocabulary used by teachers in interaction with their students (Meara et al. 1997). Regarding learners' vocabulary in free writing, results in the studies mentioned above point to (i) a moderate but still significant relationship between vocabulary profiles and academic performance (Morris and Cobb 2004), (ii) a decrease in most frequent words (2,000 words), (iii) and an increase of least frequent ones (beyond 2,000) as instruction in the target language increases (Laufer 1994), (iv) a considerable increase of most frequent words but moderate increase of least frequent ones (Horst and Collins 2006), (v) an increase in the number of sophisticated words in learners' texts as a result of draft revision, although "the extent of this improvement in vocabulary is not high enough to be reflected in the overall Lexical Frequency Profile of the finished article" (Muncie 2002:233).

6.3 Research Questions and Methodology

In the present study, we propose a complementary approach to the traditional analysis of lexical availability studies by looking at learners' word responses from the perspective of word frequency. By means of VocabProfile, we aim to explore lexical richness in the words retrieved by EFL learners in response to prompts in a lexical availability task. The specific focus of our study is the lexical availability production of EFL learners in Spanish primary and secondary education. There is a well-established body of research literature (see, for example, Nation 2001; Laufer 1994, 1998; Laufer and Nation 1995; Nation and Waring 1997) which claims that knowledge of the most infrequent words presupposes knowledge of the most frequent ones. However, most of the research conducted with VocabProfile has been with university EFL learners rather than with primary or secondary school EFL learners. Likewise, most of the studies have focused on compositions rather than on other types of

lexical production. It remains to be proved whether the pattern of a decreasing scale from most to less frequent words advocated in most studies with VocabProfile is observed in tasks different to compositions. The present study sets out to investigate the following research questions:

1. Does a lexical availability task reveal quantitative differences in the output of EFL learners at two school grades: 6th grade (final year of primary education, age 11–12) and 8th grade (2nd year of secondary education, age 13–14)?
2. Does a lexical availability frequency profile (as measured by VocabProfile) reveal qualitative differences in the output of EFL learners at 6th and 8th grades?
3. What is the nature of any change observed in relation to grades (do 8th grade learners produce more words, and/or more infrequent words)?
4. Do all lexical availability prompts behave in the same way?

Our sample comprises 50 Spanish learners of EFL representing two course groups: 6th and 8th grades. The former was selected at random out of a sample of over two hundred 6th grade primary school EFL learners from four schools in the capital city of La Rioja[2]; the second group comprises an intact group of EFL learners from a bilingual program (CLIL) implemented in a state secondary school in a middle 'Town' in the same region. At the time of data collection, the 6th grade learners had been taught about 629 h of English as a subject, and the 8th grade learners had been taught 839 h of English as a subject plus 350 h Social Sciences in English (CLIL). Both groups were asked to respond to nine prompts in a lexical availability task 'Animals', 'Parts of the body', 'Countryside', 'Food and drink', 'Professions', 'School', 'Sports and Hobbies', 'Town', and 'Transport'). Students were encouraged to write down as many words as possible in response to each prompt. The time given was 2 min per prompt, controlled by a stopwatch. We followed the same methodological steps described in Jiménez Catalán and Ojeda Alba (2009), also adopted by Jiménez, Agustín, Fernández and Canga in Chap. 3 and by Agustín and Fernández in Chap. 5 of this volume.

In order to calculate the lexical frequency profiles of our informants, we typed their word responses to each prompt into a text file, one file per group. The next step was the submission of the word responses to WordSmith Tools as well as to VocabProfile. This double process allowed us to undertake different perspectives in the analysis: the former provided us with the alphabetical lists and word ranks of the words retrieved by each group of learners. The use of this tool gave us the opportunity to identify the words elicited by each prompt and by the number of informants who retrieved each word; this provided us with the lexical frequency profiles of each group according to the external references mentioned earlier in our review (*General Service List and the Academic Word List*). The present study focuses on between-group comparisons; our aim is to determine lexical availability in 6th and 8th grade learners, and to consider the vocabulary profiles of these two groups rather than to look at individual differences within each group.

[2] See Jiménez Catalán and Ojeda Alba (2009) for a detailed description of the whole sample.

6.4 Results

The 8th grade learners had received more instruction in and exposure to English than the 6th grade learners, and therefore we would predict that their lexical proficiency would be quantitatively and qualitatively more advanced. According to advocates of word frequency as a measure of vocabulary growth and lexical richness (Laufer 1994; Laufer and Nation 1995), an increase of low frequent words would be expected as learners advance in language level. On the basis of this assumption, we believe that a considerable increase of number of hours of exposure as well as an increase of course grade would result in more advanced vocabulary. As Laufer (1994: 29) notes: "…the general tendency is for the basic vocabulary to decrease and be replaced by a more advanced vocabulary" (basic and advanced meaning high and low frequency words respectively). Therefore, in the present study the difference between the groups would be characterized by two features: (i) a decrease in the bands of the most frequent words K1 and K2), and (ii) an increase in the bands of the least frequent words AWL and Off-List).

6.4.1 Lexical Availability: Quantity and Quality of Output

The overall lexical availability output of EFL learners at the two school grades was as follows: 6th graders produced a total of 2,257 tokens and 475 types, compared to the 3,702 tokens and 853 types produced by the 8th grade group in the same lexical availability task. The mean number of word types produced in each group was: 109 (6th grade) and 205 (8th grade). According to these figures, the upper grade group generated a considerably higher lexical output than the lower group both in terms of quantity (tokens) and variety (types). As shown in Table 6.1, this pattern is broadly similar across the nine prompts used. However, the increase is by no means equal across all prompts, with the number of words both tokens and types retrieved by each group varying according to the prompts: compare a difference of 252 tokens ('School') with a difference of 99 ('Town'), and for types, a difference of 106 ('Countryside') with a difference of just 6 ('Parts of the body').

A closer inspection of the raw figures reveals slight differences in the most and least productive prompts at 6th grade and at 8th grade. In a ranking from most to least productive prompts by type, in 6th grade we find: (1) 'Town', (2) 'Sports and hobbies', (3) 'Countryside', (4) 'Food and drink', (5) 'Animals', (6) 'Professions', (7) 'School', (8) 'Transport', and (9) 'Parts of the body'. As to 8th grade the prompts ranked in order of productiveness were: (1) 'Countryside', (2)'Sports and hobbies', (3) 'Transport', (4) 'Town', (5) 'School', (6) 'Animals', (7) 'Professions', (8) 'Food and drink', and (9) 'Parts of the body'.

Table 6.1 Lexical availability task output distributed by prompt item

	Tokens			Types		
Prompts	6th grade	8th grade	Difference	6th grade	8th grade	Difference
'Animals'	298	436	138	70	128	58
'Parts of the body'	271	388	117	44	50	6
'Countryside'	181	332	151	74	180	106
'Food and drink'	322	471	149	73	116	43
'Professions'	170	360	190	69	127	58
'School'	283	535	252	57	132	75
'Sports and Hobbies'	294	414	120	84	165	81
'Town'	254	353	99	92	154	62
'Transport'	184	413	229	53	157	104

Table 6.2 Frequency band distribution of types in 6th and 8th form

	K1	K2	AWL	Off-list words
6th grade	161 (34 %)	127 (27 %)	5 (1 %)	182 (38 %)
8th grade	327 (38 %)	166 (19 %)	26 (3 %)	333 (39 %)

6.4.2 Lexical Availability Frequency Profiles: Quality and Quantity of Output

As described in the methodology section, the words retrieved by 6th and 8th grade learners in response to prompts in the lexical availability task were submitted to the VocabProfile program for analysis by frequency. The aim was to determine whether the lexical availability frequency profile of the learners at the higher school grade was qualitatively different from that of the lower grade learners. If, for example, the 8th grade learners produced a higher number of infrequent words, this could be interpreted as a greater level of lexical richness. Table 6.2 displays the frequency profiles of learners' word responses by types). Several tendencies can be observed in the results. In the first place, we can observe that the 8th grade learners produce a greater raw number of types in all four category bands. This difference is most pronounced in the AWL band (an increase of x5.2), followed by the K1 band (x2.03) and the off-lists band (x1.83), with the smallest increase in the K2 band increase of (x1.31). The percentage figures (in brackets in Table 6.2) show that there is also an increase in the K1 words produced as a percentage of total output (an increase of 4 %). The anomaly here seems to be in the K2 band, which is proportionately considerably better represented at 6th grade than at 8th grade.

We wanted to determine whether these qualitative patterns were constant across the nine different cue words. In the analyses below, we again focus on types rather than tokens, and we compare the raw and proportionate profiles for each learner group, by cue word. Whereas the lexical availability task output shown in Table 6.1 can be interpreted as quantitative output number of items produced), we interpret the measures illustrated below as qualitative output (the percentage profiles) and combined qualitative/quantitative output (the raw profiles reflect the numbers of words produced and their frequency).

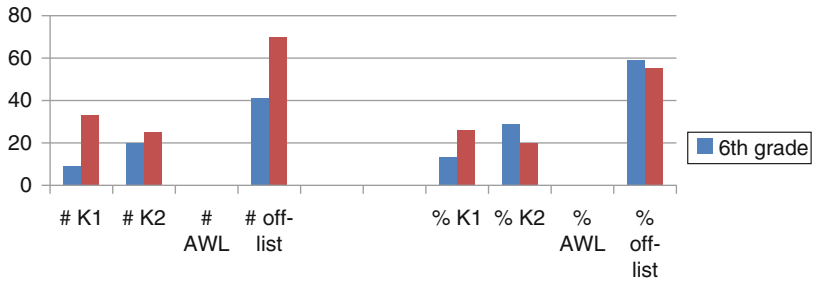


Fig. 6.1 'Animals'

'Animals': the profiles produced by this cue word (Fig. 6.1) do not straightforwardly fit predicted patterns. While the 8th grade learners produce a higher number of words than the 6th graders in each frequency band, the columns on the right show that a larger proportion of their overall output is K1 level. In other words, it is more likely that every next word retrieved by an 8th grade learner is a K1 word, than it is for a 6th grade learner. This runs counter to conventional predictions. Sixth grade learners produce proportionately more K2 and off-list words than 8th graders. There is a predominance of off-list words produced, both in terms of number and percentage, for both groups. This is perhaps to be expected in a domain such as 'animals', the predominance of which in children's literature and course books is usually out of step with frequency of occurrence in the sort of adult text-based corpora which underpin frequency lists.

'Parts of the body': in this prompt, the profiles are more closely matched, both between groups and between measures. The 8th graders broadly mirror the 6th graders' profiles, and for this prompt there is a predominance of responses in the K2 band. The 8th graders, as predicted, produce more off-list words than the 6th graders. Counter to predictions, the number of K1 items also increases (Fig. 6.2).

'Countryside': the available lexicon of 8th graders for this prompt is almost twice that of 6th graders; the difference is particularly strong in K1 and off-lists bands, and interestingly, for both groups the dominant response type is highly frequent items. If we look at the percentage columns, we note a decreasing pattern in the percentages in these bands for the 6th graders. There is some indication of increase in AWL and Off-List bands. The profile for this prompt seems closer to predictions, as there is an increase in AWL and Off-Lists at 8th grade, however, increase is also observed in K1 words (Fig. 6.3.)

'Food and drink': for this prompt, with the exclusion of the AWL, the frequency band profile is the opposite to what we might expect to find in discursive output. Less frequent words are more commonly produced than high frequency words, by both learner groups. As with 'Animals', it is likely that this domain includes specific exemplars which are not reflected in standard corpora (Fig. 6.4).

'Professions': there is a difference between the groups in the raw numbers of items produced in each frequency band, and this is most notable in the K1 band. The percentage counts reveal very equally matched profiles for the two groups, indicating that items from all frequency levels are acquired in this domain between 6th and 8th grades (Fig. 6.5).

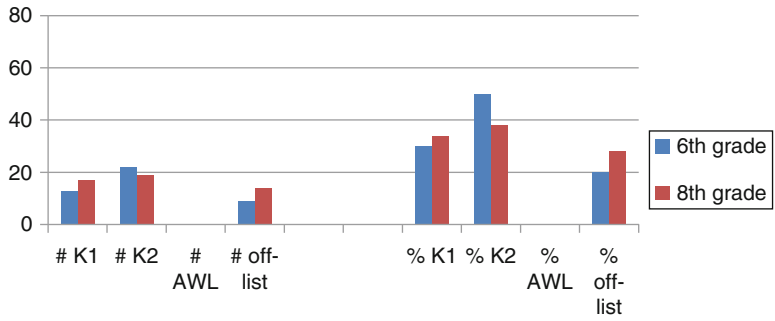


Fig. 6.2 'Parts of the Body'

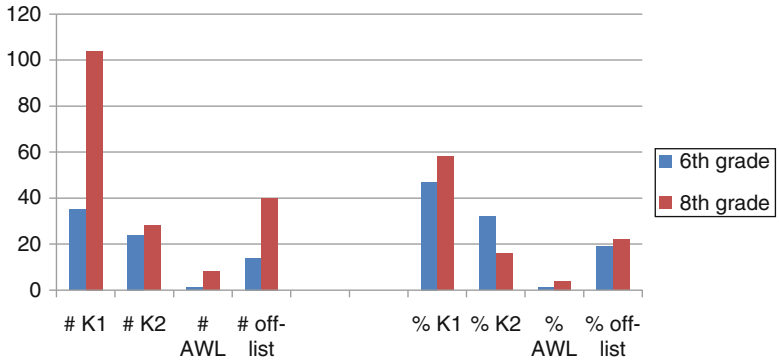


Fig. 6.3 'Countryside'

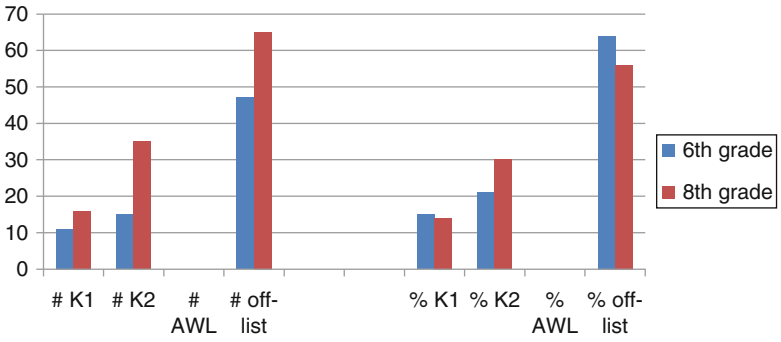


Fig. 6.4 'Food and Drink'

'School': regarding this prompt, we observe a considerable growth in 8th grade at all bands except K2. The increase is particularly pronounced in K1. Although, the pattern we observe is atypical in that there is a higher raw increase in K1 than in any other bands, this is less extreme in the percentage figures. Still, the 6th grade learners produce a higher proportion of K2 words to K1 words than do the 8th graders.

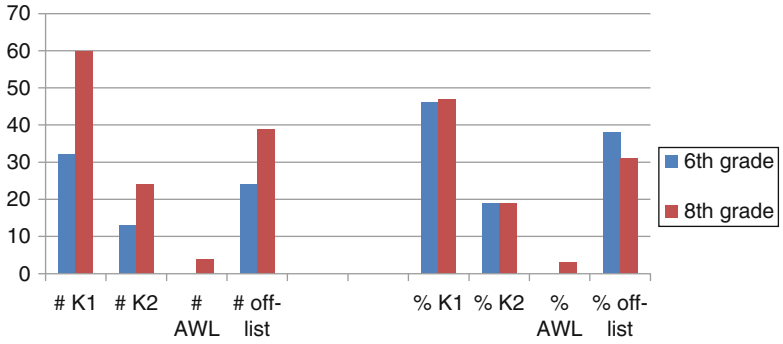


Fig. 6.5 'Professions'

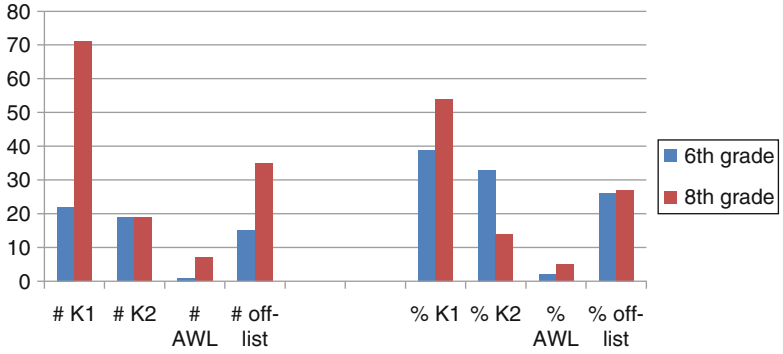


Fig. 6.6 'School'

Lexical availability for this prompt is relatively high, especially in K1 and off-list bands, and this can be explained by the fact that a lot of lexical items in the domain of 'school' are learned during secondary education, particularly, in a situation of content and language integrated learning instruction (Fig. 6.6).

'Sports and hobbies': as with 'professions', there is an increase across the bands in the total output of 8th grade compared to 6th grade. An apparent spurt in acquisition of off-list items between 6th and 8th grade results in this band representing the same proportion of words a K1, by 8th grade (Fig. 6.7).

'Town': Again, there are spikes here for K1 and off-list words at 8th grade. The pattern seen in other prompts of 6th graders producing a higher percentage of K2 words than 8th graders, at the expense of K1 words, is also observed here (Fig. 6.8).

'Transport': As with 'Countryside', there is a considerable increase in the overall output between 6th and 8th grade, and this is attributable to a dramatic jump in the number of K1 words produced. For 6th graders, there is an unusually high number of off-list words produced in comparison to K1 words; this has only previously been seen in the 'food and drink' prompt set (Fig. 6.9).

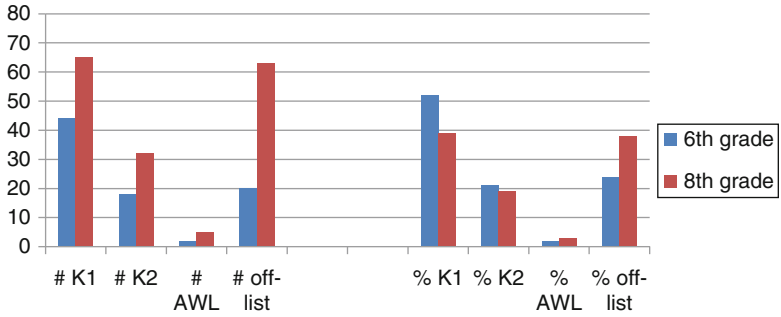


Fig. 6.7 'Sports and Hobbies'

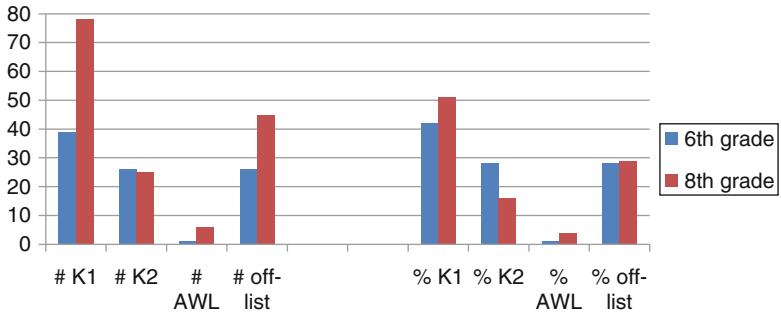


Fig. 6.8 'Town'

Fig. 6.9 'Transport'

6.5 Discussion

The increase in the overall and in the specific (prompts) lexical availability output of the 8th graders is as predicted, given the fact that they had had 560 h more instruction and exposure to English than the 6th graders. In school contexts, an increase in the number of hours of instruction goes hand in hand with an increase in age and school grade. Research on the age factor as well as on vocabulary acquisition and development in second language acquisition has shown a positive correlation between the number of hours of target language instruction and vocabulary growth. The study by Celaya et al. (2001) on Catalan EFL learners, and the study by Horst and Collins (2006) on French ESL learners are particularly relevant here (see also Cenoz 2002). Both were based on data from informants of similar age and course grade as the informants in the present study: the former included two groups of 12-year-old secondary students who had received 200 and 416 h of instruction each, the latter included 11–12-year-old 6th grade primary school students. In both studies, there was an increase in the words used in compositions as the number of hours of instruction increased. *VocabProfile* was used in the Horst and Collins' study, as it was in the one we have reported here.

Our results indicate that there are differences between 6th and 8th graders' output not only in the total numbers of types and tokens produced, but also regarding each individual prompt. Overall, 8th graders produced 1,512 more tokens and 378 more types than 6th graders. This finding points to an increase in lexical fluency as well as in lexical richness, as types stand for different words. The analyses we have presented here are based on group data rather than individual learner data, so that the tokens reported in our study represent the total number of responses to the nine prompts in the lexical availability task, from all the learners in each group.

These global figures tell us that the 8th graders produce more words to each prompt, and that as a group they are able to produce a greater variety of words. However, the lexical diversity of the groups, as measured by their type token ratios, is similar. The 6th graders have a TTR of .21, and the 8th graders .23. The small difference might indicate that there is slightly less repetition in the output of the 8th graders (but note that this refers to repetition within the group, not by individuals). Overall as well as specifically for each prompt, vocabulary growth occurs, then. However, the figures indicate that this increase is not even in all prompts. Both in 6th grade and 8th grade we note that some prompts are more productive than others. Our results corroborate the results in this volume by Ferreira and Echeverría in English as L1 and L2 (Chap. 2), and by Šifrar Kalan regarding Spanish as L1 and L2 (Chap. 8), as well as the findings reported by pioneer researchers in lexical availability studies in English and Spanish as L1 such as Dimitrijévic (1969), Bailey (1971) and Carcedo (1998). If we compare the results, we observe a striking similarity, despite the fact that the informants are from different mother tongues, ages, language levels or learning contexts. As mentioned in Chaps. 2 and 10 in this volume, these similarities suggest the existence of universal processes underlying learners'

lexicons. The lexical availability task elicits a predominance of nouns over other categories such as verbs, adjectives or adverbs: practically all the responses in our study are nouns. There are a number of possible explanations for this. Firstly, the prompts are all nouns, and many are superordinates or hypernyms, which have a propensity to elicit other nouns. Secondly, there is some evidence that the first words acquired in speakers' mother tongues ("labels") usually belong to the noun category, and research indicates that early acquired words are often the most easily accessed, and therefore are likely to be retrieved by prompts such as the ones used in lexical availability tasks. Indeed, in Chap. 3 by Jiménez, Agustín, Fernández and Canga we find evidence to support this interpretation, as they found no differences in the lexical availability of 6th grade primary students out which the sample of EFL learners in the present study was drawn) and 1st year university students. Likewise, evidence of universal processes in the organization of lexicons is found in the fact that some words tend to be learnt and/or stored in association with other words, in conceptual networks. Examples of this would be *pencil-ruler-pen-book-desk*, or *restaurant-waiter-menu-meal-food-drink-main course-bill-tip*. In the lexical availability task we used, these word groups might be elicited by the 'School' and 'Food and drink' prompts respectively.

Another possible interpretation regarding learners' identical responses, within and across languages, to lexical availability tasks might be that the teaching of foreign languages in primary, secondary and tertiary education does not differ as much as we might think at first sight. It may the case that the vocabulary input given to learners does not vary a great deal in foreign language education: in primary, secondary and, even in tertiary education, it continues to refer to thematic issues similar to the prompts used in this lexical availability task. Some words may be internalized in our lexicon as typical examples in western culture such as *cat* or *dog* (rather than *alligator* or *crocodile*) for the prompt 'Animals'. As Hernández Muñoz et al. (2006) showed, typicality is a good predictor of lexical availability see also their chapter in this volume). The results reported elsewhere in this volume (see Chaps. 2, 8 and 10) offer evidence of striking similarities in the responses provided by native and non-native speakers both in English and in Spanish, which again supports the plausibility of the hypothesis of universality of the organization of the lexicon.

It is clear that 8th graders' lexical availability output is characterized by an increase in lexical richness: the greater number of word types provided by learners' responses to the prompts gives evidence of this in the number of different words elicited as well as in the number of learners who retrieved each word. However, it is also evident that this increase in learners' word types does not result automatically in a more advanced frequency profile: if that were the case, the right-hand (percentage) set of columns in the charts above would show a lower proportion of K1 words for 8th graders than for 6th graders. In truth, the opposite is the case for all but two of our nine prompts. A great number of word types, then, means higher lexical richness but not necessarily greater lexical sophistication (as would be indicated by the increased production of less frequent words). In this, our study findings coincide with those reported by Horst and Collins (2006) from French 6th grade primary

school ESL learners in Quebec, where contrary to expectations, rather than integrating words from the least frequent bands, French ESL learners grew their vocabulary in the 1k and 2k bands. We agree with those authors in their interpretation of this unexpected result: far from being negative, it is a positive one, as it proves that, at least at beginner and low intermediate levels, vocabulary still has room to grow in the most frequent bands. We should note too, that the lexical availability task is a productive one, and that differences in the vocabulary performance of 6th and 8th graders might be partly attributable to frequent vocabulary items moving from a state of receptive knowledge to a state of active knowledge. However, our findings differ from those observed in Horst and Collins' French ESL learners, in that both 6th and 8th graders in our study, retrieved a significant number of words from the least frequent band: Off-List. Moreover, in our study a slight increase is observed in this band at 8th grade in five of the prompt sets, something that did not happen in the Quebec study. The difference in the findings of the two studies might be due to the different number of hours of instruction and exposure to the target language. As Horst and Collins remark, their informants, French primary school ESL learners, had received 100, 200, 300 and 400 h of English instruction, and the sample was drawn from an area in Quebec whether there was hardly any contact with English, whereas in our study, 6th primary school EFL learners had received 629 h and 8th graders 839 h plus 350 of English as vehicular language to learn content in a subject. The increase in the Off-List band observed in our students may have to do with the higher number of hours of instruction.

Another difference between the studies, and one which is worthy of discussion, is that they used different instruments for eliciting output from the learners. Whereas here we used lexical availability tasks, Horst and Collins used compositions. Both conceptually and operationally, Lexical Frequency Profile tools were originally designed for use with free writing (specifically, with the written output of EFL learners at university). This may explained the regular patterns found in studies such as Laufer and Nation 1995, particularly regarding the increase of words belonging to UWL (the equivalent of AWL in our study), a list made up of academic words, and created to address university students' academic requirements. In contrast, primary and secondary school EFL learners learn English following a communicative approach rather than an academic approach. Most of the item responses classified by Vocab-Profile, in our study, as Off-List are in fact very familiar to learners throughout primary and secondary education: these are vocabulary items which learners encounter in English lessons under the communicative approach. Just as it is almost impossible to conduct a communicative interaction at a restaurant without knowing words or expressions such as *waiter, order, menu, main course, tip, table, food, drink, water, wine, juice*, etc., without them, communicative interaction in an English lesson entitled 'At the restaurant' or ' What is your favourite meal' would be unthinkable. However, the corpora from which the frequency profile lists and tools are derived consist of written and spoken texts from domains and genres other than this, and the global frequency of these items does not reflect their frequency and usefulness in these communicative contexts.

6.6 Conclusions

The present study has shown that EFL learners' lexical availability increases as the course grade increases. However, regarding the frequency profile of learners' output, we saw that this increase occurs primarily in the most frequent words: particularly, the K1 band, and secondarily in the Off-List band (beyond 2,000). The percentages for the less frequent words are higher than the ones obtained in other lexical frequency studies, and we suggest that this is due to the effect of a non-discursive elicitation task, together with the effect of acquisition through communicative teaching approaches. We also noted that that the prompts used elicit hardly any academic words from the primary and secondary EFL learners.

Our results showed an increase in the lexical availability of 8th graders, compared to 6th graders both overall and specifically for each of the nine prompts used in the lexical availability task. However, as occurs in most lexical availability studies both in L1 and L2, some prompts proved to be more productive than others in retrieving words from learners' lexicons. Likewise, some prompts ('Countryside', 'Sports and hobbies') revealed a greater increase in the responses of 8th graders, whereas for others ('Parts of the body') there was very little increase. In the first case, the result may due to instructional and psychological factors not accounted for in this study. For instance, 8th graders might have been exposed by means of lessons and course books to a higher number of words related to those prompts. This interpretation is very plausible, particularly regarding 'Countryside', as 8th graders were using English to learn social sciences, and 'Countryside' is one of the semantic fields required to discuss social issues. Also, 8th graders' age (adolescents) suggests an increase in their interest towards sports and hobbies, which may explain the increase in lexical resource in this domain. As for 'Parts of the body', at first sight we might think that the low increase may be due to the existence of a top ceiling effect caused by the nature of the prompt: it has the limitation of being a closed system rather than an open system; however, the production of English native speakers (secondary school students) on this same prompt reported by Ferreira and Echeverria in Chap. 2 in this volume tells us that a higher increase might have been potentially possible.

In sum, then, our study has shown that (i) the vocabulary of 6th grade and 8th grade learners differs quantitatively, with the later producing many more tokens and types than the former, (ii) the vocabulary of 6th and 8th graders differs qualitatively in terms of frequency profiles, though not always in the way predicted by earlier studies; (iii) vocabulary retrieval depends on the prompts used; (iv) most vocabulary growth between these two grades occurs at the 1 K band, followed by the OffList band; (v) there is little evidence of knowledge of academic vocabulary; but this is perhaps due to the nature of the prompts used in the lexical availability task.

Importantly, we suggest that by combining a lexical availability task with a frequency profile analysis, this study has demonstrated that other aspects than frequency should be taken into account when measuring the lexical output of learners at this age and level: they almost certainly know more vocabulary than frequency list measures will imply. The evident usefulness of frequency lists to the sort of applications used by other researchers and practitioners materials design, analyses

of written texts, etc.) does not necessarily extend to identifying the vocabulary needed by learners in their daily interactions.

Finally, in closing we acknowledge the limitations of this study. We focused on the lexical availability of learners as a group rather than as individuals. It would be useful to follow up this study with one focusing on correlations between language level (a general proficiency measure), and lexical availability and frequency profile. Further studies are also needed to explore the potential of the lexical availability task to measure vocabulary size: the vocabulary retrieved by learners in this study, in response to certain prompts, provides us with a useful indication of the nature of lexical resources available to learners of this age and level. As the size of the corpus we have elicited demonstrates, the lexical availability task is a very economical way to obtain relatively rich sets of data.

Acknowledgments We would like to acknowledge the financial support of Ministerio de Ciencia e Innovación through grant nº FFI2010-19334.

References

Alderson, C. 2007. Judging the frequency of English words. *Applied Linguistics* 28(3): 383–409.

Bailey, V. J. 1971. *A study of lexical availability among monolingual-bilingual speakers of Spanish and English*. Master thesis, Rice University, Houston.

Carcedo, A. 1998. Tradición y novedad en las aportaciones hispánicas a los estudios de disponibilidad léxica. *Lingüística* 10: 5–68.

Celaya, M.L., M.R. Torras, and C. Pérez-Vidal. 2001. Short and mid-term effects of an earlier start: An analysis of EFL written production. In *EUROSLA yearbook*, ed. S. Foster Cohen and A. Nizegorodcew, 195–209. Amsterdam: John Benjamins.

Cenoz, J. 2002. Age differences in foreign language learning. *ITL- International Journal of Applied Linguistics* 135(136): 125–142.

Coxhead, A. 1998. An academic word list. Occasional Publication Number 18, LALS, Victoria University of Wellington, Wellington.

Dimitrijévic, N. 1969. *Lexical availability*. Heidelberg: Julius Gross Verlag.

Gardner, D. 2004. Vocabulary input through extensive reading: A comparison of words found in children's narrative and expository reading materials. *Applied Linguistics* 25(1): 1–37.

Hernández-Muñoz, N., C. Izura, and A.W. Ellis. 2006. Cognitive aspects of lexical availability. *European Journal of Cognitive Psychology* 18(5): 730–755.

Horst, M. 2010. How well does teacher talk support incidental vocabulary acquisition? *Reading in a Foreign Language* 22(1): 161–180.

Horst, M., and L. Collins. 2006. From faible to strong: How does their vocabulary grow? *The Canadian Modern Language Review* 63(1): 83–106.

Jiménez Catalán, R., and S. Moreno. 2005. Promoting English vocabulary research in primary and secondary education: Test review and test selection criteria. *English Studies ES* 26: 171–188.

Jiménez Catalán, R., and J. Ojeda Alba. 2009. Girls' and boys' lexical availability in EFL. *ITL International Journal of Applied Linguistics* 158: 57–76.

Laufer, B. 1991. Knowing a word: What is so difficult about it? *English Teachers' Journal* 42: 82–88.

Laufer, B. 1992. How much lexis is necessary for reading comprehension? In *Vocabulary and applied linguistics*, ed. Henri Bejoint and Pierre J.L. Arnaud, 126–132. Basingstoke/London: Macmillan.

Laufer, B. 1994. The lexical profile of second language writing: Does it change over time? *RELC Journal* 25(2): 21–33.
Laufer, B. 1998. The development of passive and active vocabulary in a second language: Same or different? *Applied linguistics* 19: 255–271.
Laufer, B., and P. Nation. 1999. A vocabulary size test of controlled productive ability. *Language Testing* 16: 36–55.
Laufer, B., and P. Nation. 1995. Vocabulary size and use: Lexical richness in L2 written productions. *Applied Linguistics* 16(3): 307–322.
Laufer, B., and T.S. Paribakht. 1998. The relationship between passive and active vocabularies: Effects of language learning context. *Language Learning* 48: 365–391.
Lynch, T., and J. Maclean. 2000. Exploring the benefits of task repetition and recycling for classroom language learning. *Language Teaching Research* 4(3): 221–250.
Meara, P. 1993. Tintin and the world service: A look at lexical environments. IATEFL: Annual Conference Report, 32–37.
Meara, P., and T. Fitzpatrick. 2000. Lex30: An improved method of assessing productive vocabulary in an L2. *System* 28: 9–30.
Meara, P., P. Lightbown, and R. Halter. 1997. Classroom as lexical environments. *Language Teaching Research* 1(1): 28–46.
Morris, L., and T. Cobb. 2004. Vocabulary profiles as predictors of the academic performance of Teaching English as a Second Language trainees. *System* 32: 75–87.
Muncie, J. 2002. Process writing and vocabulary development: Comparing lexical frequency profiles across drafts. *System* 30: 225–235.
Nation, P. 1990. What is involved in learning a word? In *Teaching and learning vocabulary*, ed. I.S.P. Nation, 29–50. Rowley: Newbury House.
Nation, P. 1997. The language learning benefits of extensive reading. *The Language Teacher* 21(5): 13–16.
Nation, P. 2001/2007. *Learning vocabulary in another language*. Cambridge: Cambridge Applied Linguistics.
Nation, P., and R. Waring. 1997. Vocabulary size, text coverage and word lists. In *Vocabulary: Description, acquisition and pedagogy*, eds. N. Schmitt and M. McCarthy, 6–19. Cambridge: Cambridge University Press.
Pimsleur, P. 1967. A memory schedule. *The Modern Language Journal* 51(2): 73–75.
Read, J. 1988. Measuring the vocabulary knowledge of second language learners. *RELC Journal* 19: 12–25.
Read, J. 2000. *Assessing vocabulary*. Cambridge: Cambridge University Press.
Richards, J. 1974. Word lists: Problems and prospects. *RELC Journal* 5(2): 84.
Richards, J. 1976. The role of vocabulary teaching. *TESOL Quarterly* 10(1): 77–89.
Richards, J. 1985. Lexical knowledge and the teaching of vocabulary. In *The context of language teaching*, ed. J. Richards, 176–188. Cambridge: Cambridge University Press.
Scarborough, D.L., C. Cortese, and H.S. Scarborough. 1997. Frequency and repetition effects in lexical memory. *Journal of Experimental Psychology. Human Perception and Performance* 31: 1–17.
Schmidtt, N. 2000. *Vocabulary in language teaching*. Cambridge: Cambridge University Press.
West, M. 1953. *A general service list of English words*. London: Longman, Green and Co.

Part II
Lexical Availability in Spanish as L1 and L2

Chapter 7
The Relationship of Language Proficiency to the Lexical Availability of Learners of Spanish

Marta Samper Hernández

7.1 Introduction

The present study looks at the effect of the level of proficiency on the lexical availability of learners of Spanish as a foreign language (SFL) in an immersion context. Our study is part of a research project carried out with students from *Cursos Internacionales* of the University of Salamanca[1] Samper Hernández (2002),[2] which was considered a pioneering investigation in its field as, for the first time, it looked at the lexical availability of learners of Spanish in an immersion context. In this chapter, we will focus on the lexical availability in two levels of Spanish (Basic and Advanced) and we will analyze the results from a quantitative as well as a qualitative perspective.

The variable level of proficiency has turned out to be one of the most influential on the available lexicons of foreign language learners; by specifying the words that learners from different language levels retrieve, researchers can get insights into how lexical knowledge evolves and what kind of lexical difficulties learners encounter on their way towards proficiency in the target language. The results will provide us, therefore, with useful conclusions for vocabulary curricular planning in SFL.

[1] We are very grateful to the teaching and managerial staff from the Cursos Internacionales of this University and, of course, to the students who participated in the study.

[2] This study was awarded the ASELE 2000 and was published by this association.

M. Samper Hernández (✉)
Departamento de Filología Española. Clásica y Arabe,
Universidad de Las Palmas de Gran Canaria, Las Palmas, Spain
e-mail: msamper@dfe.ulpgc.es

R.M. Jiménez Catalán (ed.), *Lexical Availability in English and Spanish as a Second Language*, Educational Linguistics 17, DOI 10.1007/978-94-007-7158-1_7,

7.2 Studies Review

Studies on lexical availability have already had a long tradition in the Hispanic world since the 1970s when López Morales introduced them for the first time into Spanish language research. He did so by following in the footsteps of the pioneering work conducted by the French researchers in the field. The concern in identifying those words which are not activated so frequently, but which cross our minds immediately when a specific theme is presented, led López Morales to develop the PanHispanic project; this project, which aims to gather the available lexicon of non-specialized adult speakers (in the last school year before starting a degree), continues to collate multiple contributions and will culminate with the publication of the available lexicon in Spain in the short term, and that of the whole Spanish-speaking world in the long term.[3] Alongside the PanHispanic project, other investigations have been carried out following the same methodological criteria but with a different focus. The research reported in this chapter is framed within one of these satellite projects, specifically the one which analyzes the lexical availability of learners of Spanish as a foreign language. Investigations conducted with non-native speakers originated in Finland with Carcedo's (1998, 1999, 2000, among others), who established the theoretical and methodological grounds for future studies. Samper Hernández (2002) introduced the methodological novelty of working with a sample of foreign subjects who were studying the language in Spain when the test was administered; that is, in a situation of linguistic and cultural immersion. Subsequent to these initial studies, and starting from them in most cases, the number of scholars who have explored lexical availability in Spanish as well as in English as a foreign language has been on the rise.

Among the extra-linguistic factors considered in this research, the level of proficiency has received considerable attention. Henceforth, we will briefly review the research which analyzes the influence of this variable starting from the initial research projects carried out with foreign language learners.

Carcedo (2000) worked in Finland with a sample made up of 350 subjects who were learning Spanish as a foreign language. Although the main objective of his research was, at first, the lexical availability of students in the last year of secondary education (*bachillerato*), the author decided to increase the sample to include other academic years in order to "determine if there are common characteristics applicable to the whole 'Finnish' group, irrespective of the level of study they are at, and also the evolution and characteristics of each phase" (66).[4] To that end, to the original 150 sample comprising a group of 8th year high school students, another 200 was added; out of this number, 150 4th year high school students and 50 university students from the Spanish Department of the University of Turku, equally distributed into two different levels.

[3] For more information, see for example, Samper et al. (2003) and Samper and Samper Hernández (2006).

[4] The translation of the quotes in the present study is our own.

The other study which is considered pioneering in this field was that of Samper Hernández (2002), which the present chapter is based on. As was mentioned previously, the research project presented the novelty of working with a sample made up of foreign students in a situation of linguistic and cultural immersion. In this chapter, we will study the variable level of proficiency in depth, with a bipartite distribution of the sample, which will be detailed in the methodological section. From the other factors we took into account, the students' mother tongue had a special effect, while sex and knowledge of other languages scarcely had any significance among the various sample subgroups.

Subsequent to these two publications, several studies have looked at the variable under discussion.

Sánchez Gómez (2005) carried out a preliminary qualitative analysis of the lexical availability of 44 subjects from the *International Studies Abroad Centre* in Sevilla, distributed into four distinct competence levels (Intermediate I, Intermediate II, Advanced and Superior). For this purpose, she chose ten traditional lexical domains from the PanHispanic project, to which she added 'Education' and 'Health'.

The results obtained in 5 of the 12 themes tested by Sánchez Gómez were also analyzed by Galloso and Prado (2005), who, in addition, contrasted them with the lexical availability of native speaker students in Huelva. The authors aimed to ascertain, among other hypotheses, whether this comparison revealed differences in groups of different language level proficiency.

Jing Lin (2006) collected data in various centres in China, aiming not only to obtain the lexical availability of Chinese students, but also to determine whether the vocabulary related to food appearing in most frequently used textbooks corresponds to the lexical production of these students as well as to that of Spanish students from Alcalá de Henares. The sample comprised of 263 participants, who were distributed according to their Spanish level as follows: 128 Elementary, 55 Intermediate, and 80 Advanced.

More recently, Sánchez-Saus (2009) analyzed the effect of the Spanish level factor on the lexical availability of 81 university students who were studying Spanish as a foreign language at the *Centro Superior de Lenguas Modernas* of the University of Cádiz. Regarding the distribution of the subjects according to their competence in Spanish, the author reduced the variants to the three levels established by the European Common Reference Framework for Languages, without considering each one of the subgroups these were divided into individually. So, she distinguished between level A (composed of 18 students), level B (42), and level C (21). One of the novelties of this contribution is that it makes use of cue words which, although they coincide to a great extent with those which are traditionally used in studies on available lexicon, also allowed the inclusion of "basic topics which are present in any list of lexical contents for the initial levels of learning Spanish as a foreign language" (145). In addition, she attempted to establish the occurrence of other lexical categories different from the noun.

Also in 2009, Medina Arejita set out to demonstrate "the usefulness of lexical availability tests as basic selection tools of the specific notions of *Cervantes Syllabus* (*Plan Curricular del Instituto Cervantes*) (2006)" (7). With this intention,

she selected a sample of 43 German students who were studying Spanish as a foreign language at three different levels: A (18 subjects), B 19, and C1 (6). The students belonged to the *Universidad Libre* of Berlin and to the *Instituto Cervantes* in the same city.

López González (2010) carried out word association tasks in various Bilingual Sections in six cities of Poland. His sample comprised of 241 subjects, of whom 120 attended the third year of basic secondary education (*Gimnazjum*) and 121 were completing the third year of higher education (*Liceum*).

On her part, Šifrar Kalan (2011) analyzed the lexical availability of 200 subjects who were studying Spanish in Slovenia: 100 were high school students who had been studying Spanish as a second or third foreign language for 3 years and the other 100 were students of Spanish Language and Literature who had B2 level in Spanish. Among the lexical themes under examination, the ones which had turned out to be less productive in previous research projects were not included (e.g., the topics of 'Objects Placed on the Table at Meals'; 'Lighting and Heating'; 'Gardening and Farming'); some of the answers that could overlap were unified ('Parts of the House' and 'Furniture'), and 'Everyday Actions' was introduced as an innovation to traditional prompts.

There are other studies which do not take into consideration the variable level of proficiency, but include the factor of years of studying the language instead. López Rivero (2008) tested 43 students with an Intermediate I level from two Official Language Schools in Madrid and established four subgroups according to the number of years they had been studying Spanish: (i) 1 year or less (32.56 % of the total of the sample); (ii) from 1.1 to 2.9 years (39.53 %); (iii) from 3 to 4.9 years (20.93 %); (iv) 5 years or more (6.98 %). As a complement of the latter, Pérez Serrano (2009) also worked with 43 Intermediate I Level students from one Official Language School in Madrid to evaluate lexical availability in two thematic categories which had not been included in the six categories analyzed by López Rivero: 'Means of Transport', and 'Professions and Jobs'. The sample was divided as in the previous study: (a) 1 year or less (34.9 % of the total); (b) from 1.1 to 2.9 years (53.5 %); (c) from 3 to 4.9 years (9.3 %); (d) 5 years or more (2.3 %).

The studies which examine, among other aspects, the influence of linguistic competence level on lexical availability in SFL continue to grow. Currently, for example, a research project is being undertaken with Romanian students from different course grades (Sandu in preparation) at the University of *Las Palmas de Gran Canaria*. The lexical availability tests were carried out in Romania with 280 subjects distributed into three groups according to their level in Spanish: 103 6th primary school students with an A2 Spanish competence level; 101 students in their first year of high school (9th grade) with a B1 level; and 76 are in the last high school year (12th grade) with a B2 level. Likewise, in *Las Palmas de Gran Canaria* data is now being collected from the Erasmus students who attend the subject *Español, lengua extranjera* at the University, with levels which range from A1 to B2 (Del Pino in preparation).

7.3 Objectives

The main objective of this study is to ascertain whether the variable level of proficiency affects the lexical availability of a group of learners of Spanish as a foreign language in an immersion context. Firstly, by means of a quantitative analysis we will look at the number of words and different units provided by the two groups which comprise our sample, and also at the degree of cohesion in their responses. The quantitative analysis of the data will then be complemented by a qualitative analysis, i.e. the revision and comparison of the most available lexical units in each level.

7.4 Research Questions

The questions that we pose are as follows:

(a) Is there a relationship between Spanish proficiency level and higher word production (word tokens)?
(b) Is there a relationship between Spanish proficiency level and higher production of different units (word types)?
(c) Does the words' cohesion index vary according to the linguistic competence of students?
(d) Is there a relationship between Spanish proficiency level and the type of vocabulary the subjects include in the lists?
(e) Which are the lexical domains that present a higher difference between the groups?

7.5 Method

In order to analyze the incidence of level of language proficiency on the lexical availability of learners of SFL, we shall start with a sample comprising 45 participants who attended the classes organized by *Cursos Internacionales* at the University of Salamanca during the first term of the school year 1999–2000 (they had arrived in Salamanca about 3 weeks before the administration of the availability test). For the present study, we shall distinguish two different groups according to the Spanish linguistic competence of the informants:

(a) Basic Spanish: made up of 22 students at Beginner and Intermediate levels, equivalent to the current A1-B1 levels of the CEFRL;

(b) Spanish advanced: a group consisting of 23 students at Upper Intermediate and Advanced levels, which correspond to the B2-C1[5] levels.

The test administered to each student was divided, as is common in this kind of studies, into two clearly distinguishable parts. The first page requested sociological information and, next, there was the main body of the test, made up of six pages which contained the cue words at the top and a series of numbered lines where the subjects had to write the answers. We opted for the open list system (if the number of lines resulted insufficient, the students could continue writing below or on the back of the corresponding sheet of paper) and for a time limit of 2 min per topic.

Likewise, the lexical domains were the ones usually included in the PanHispanic project (the same used by Carcedo in his research in Finland). The selection of these lexical domains has many advantages: among them it is the fact that they are universal or present in most languages, their frequency in Spanish language and the fact that they have been used in many lexical availability studies, which favours comparative studies. These lexical domains were: 'Parts of the Body', 'Clothes', 'Parts of the House' (without Furniture), 'Food and Drink', 'Objects Placed on the Table at Meals', 'The Kitchen and its Utensils', 'The School: Furniture and Materials', 'Lighting, Heating and Means of Airing Places', 'The City', 'The Countryside', 'Means of Transport', 'Gardening and Farming', 'Animals', 'Games and Entertainment', 'Professions and Jobs'.

In the process of data editing, we followed the criteria adopted by Samper (1998). However, the particular characteristics of this study made it necessary to establish some guidelines specific to the studies on availability in Spanish as a foreign language (Samper Hernández 2001). These guidelines focused on spelling corrections, the treatment of foreign words, unification in the presentation of the lexical units, and the range of the associative relationships, guidelines that we cannot develop further on this occasion due to space constraints.

Regarding data processing, we used the programme Lexidisp (version 2.1)[6]: this program applies the mathematical formula created by López Chávez and Strassburger. The use of this program allows us to find out the percentage of occurrence, relative frequency, accumulated frequency – sum of the relatives – and, of course, the availability of each lexical unit appearing in the list.

[5] The level tests administered to students before the beginning of the course were carried out prior to the publication of the Common European Framework of Reference for Languages (CEFRL), therefore the students from *Cursos Internacionales* of the University of Salamanca were divided into the four groups that we mentioned previously, equivalent to the current levels established by the CEFRL. Though the tests were administered to these four educational stages, in this chapter we have preferred to distinguish between two large groups in order to obtain more significant results, as other authors did; for example Sánchez-Saus (2009) reduced to three the six groups considered initially.

[6] Lexidisp is an application for Windows which can be downloaded online on http://www.linguas.net/Proyectos/LexiDisp/tabid/73/language/es-ES/Default.aspx. The program was created by J. E. Moreno and A. J. García de las Heras, who were in turn assessed by the linguists F. Moreno and Benítez, under the auspices of Alfal, Instituto Cervantes and the University of Alcalá. For further details, see Moreno et al. (1995, 243–249). See also Chap. 11.

In the following section, we will report on the quantitative results obtained in our study and will attempt at their interpretation. Furthermore, we will carry out some comparisons both with the work conducted by Carcedo (2000) in Finland, and with those research papers which are most similar to our study.

7.6 Quantitative Analysis

The total number of words retrieved by our learners is 8,127, which gives a mean of 180.6 words per informant and 507.9 per lexical domain. Regarding word types, the total yielded was 1,840, corresponding to 115 word types per cue.

If we focus on the extra-linguistic variable and the scope of our study, we find, in the first place, that in almost all the lexical fields, the advanced learners produce a larger number of words than learners in the basic level. These results are shown in Table 7.1:

The means of words per subject are, indeed, convincing. If we compare the means obtained by the basic and the advance level we observe an increase of 25 %. In addition to this tendency, we also note two cue words which do not show a difference in favour of advanced learners: 'The City' and 'Games and Entertainment'. In the former, the basic level subjects produce more words than the advanced group. In the latter, total equivalence between the two groups is observed. As can be seen, these topics are less compact and allow the inclusion of more varied lexical units, which contributes to blurring the difference between the two levels of proficiency. Likewise, the type of vocabulary less related to school learning but more directly linked to contact with Spanish reality may have consequences on the results.

Table 7.1 Mean of words according to the variable level of proficiency

	Basic level		Advanced level	
Cue	Mean	Rank	Mean	Rank
01 'Body'	14.4	3	18.1	2
02 'Clothes'	10.2	9	13	6
03 'House'	8.3	11	11.9	10
04 'Furniture'	7.2	12	8.9	12
05 'Food'	19.3	1	21.5	1
06 'Table'	6	14	8.6	13
07 'Kitchen'	7.1	13	8.6	14
08 'School'	10.4	8	12.1	9
09 'Lighting'	4.3	15	6.3	15
10 'City'	18.9	2	16.8	4
11 'Countryside'	9.4	10	12.8	7
12 'Transports'	10.6	7	12.5	8
13 'Garden'	1.8	16	6	16
14 'Animals'	10.9	5	16.9	3
15 'Games'	11.1	4	11.1	11
16 'Professions'	10.9	6	14.4	5
Total	10		12.5	

Similar conclusions were reached by Sánchez-Saus (2009), who, in her study with learners in an immersion context, obtained similar means in lexical domains such as 'The City', 'Leisure and Free Time' and 'Journeys and Holidays' (also 'Food and Drink'). The author claims that these cues "are the most closely related to practical aspects in daily life, and they have become the most familiar topics to students since they arrived in Spain". In the same vein, López González (2010) highlights the quantitative decrease shown in 'Games and Entertainment' at the higher educational level, which he ascribes to "the nature of this lexicon, more related to school life and childhood, vocabulary which is not further developed at high school". Nevertheless, in his research, 'The City' proves to be one of the fields with the highest increase of words, contrary to what happens in our study and in that of Sánchez-Saus.

Without doubt, the greatest differences lie in the responses to the cue word 'Animals', with a mean of 6 words more in the advanced level. This was followed by 'Gardening and Farming', with a distance of 4.2 answers and, with a mean of more than three terms, 'Parts of the Body', 'Parts of the House', 'Professions and Jobs', and 'The Countryside'. The domain of 'Gardening and Farming' is complex for native speakers and, of course, for non- native speakers, which may explain why the basic level students only achieve a mean of 1.8 words. The other two cue words with greater differences, 'Animals' and 'Parts of the Body', are regarded as very cohesive and, besides their specificness, they have a marked school nature. The specific vocabulary related to school learning is also generated in 'Professions and Jobs' and, to a lesser extent, in 'The Countryside', which can explain the lexical increase in these domains of much more diffuse nature. In any case, in order to confirm the school nature of this vocabulary, we would need to ascertain from a qualitative point of view which are the word types that make up a difference between the two levels. The contrast in the type of words is particularly notorious in 'Parts of the House' which *a priori* does not seem so closely linked to school vocabulary.

The conclusions reached coincide to a great extent with those reported in Sánchez-Saus (2009), who claims marked divergences between the two extreme levels of proficiency in cues such as 'Human Body', 'The Countryside' and 'Animals' (with a difference of more than 8 words). In her study, however, the cue word that draws most attention regarding the development of lexical productivity is 'School and University' (in which the difference exceeds the mean of 9 words). Likewise, remarkable coincidences can be established between our results and the ones obtained by López González (2010); yet in his study, 'Human Body', 'The Countryside', 'Animals', and 'Professions and Jobs' are some of the lexical domains with the greatest divergences between the groups. On the contrary, in 'Gardening and Farming', we only find a difference of one word, a very distant figure from the mean of 4.2 we observe in our results.

Regarding the position occupied by each prompt according to the mean number of words given by the informants, we can establish a parallel between the extreme poles: 'Food and Drink' holds the top ranking position in both groups, whereas the less productive domains are held by 'Gardening and Farming', and 'Lighting'. The progression in the evolution of the prompt 'Parts of the Body' stands out: rank 3 in basic level, rank 2 in advanced level. This tendency also appears in other studies with learners of Spanish such as Carcedo's (2000), where it goes from the

Table 7.2 Total of word types according to the variable level of proficiency

	Basic level		Advanced level	
Cue	Word types	Rank	Word types	Rank
01 'Body'	52	9	67	10
02 'Clothes'	46	13	60	12
03 'House'	51	10	63	11
04 'Furniture'	47	12	52	14
05 'Food'	123	1	165	1
06 'Table'	35	15	52	15
07 'Kitchen'	56	7	57	13
08 'School'	71	6	78	9
09 'Lighting'	48	11	50	16
10 'City'	122	2	135	3
11 'Countryside'	89	5	109	5
12 'Transports'	44	14	83	7
13 'Gardening and Farming'	30	16	82	8
14 'Animals'	55	8	94	6
15 'Games'	92	4	114	4
16 'Professions'	103	3	139	2

fourth position at fourth grade of high school to the first position at university level, in a gradual progression across educational levels. A similar trend was observed in the study conducted by Samper Hernández (2009) with Gran Canarian schoolchildren, native speakers of Spanish.

Finally, we should point out the existing difference with regard to cue word 15, 'Games and Entertainment': although basic and advanced learners retrieve the same number of words, it holds very different positions in each level. This may be due to the open nature of this vocabulary domain, as well as to its limited link to formal education.

The total number of word types reveals once more the superiority of the advanced level group, as can be seen in Table 7.2:

There is an obvious parallel in those cue words which hold the top positions according to the total number of different units retrieved. That is to say, 'Food and Drink' displays the first position in both lists, 'The City' and 'Professions and Jobs' alternate between the second and the third place, and 'Games and Entertainment' and 'The Countryside' occupy fourth and fifth positions respectively in the two groups under examination. Once again, these findings coincide to a great extent with the ones reached by Carcedo (2000) in Finland.

As to the cue words that occupy the lowest positions, we found some disparities between the two levels. The fact that 'Gardening and Farming' and 'Means of Transport' hold such extreme positions in the basic level group, whereas in the advanced one their position is intermediate is something remarkable; yet it points to a considerable vocabulary growth in the higher level. Likewise, though the distance is not so marked, the difference in the cues 'The Kitchen and its Utensils' and 'Lighting' shall be highlighted, as they hold much lower positions in the advanced

Table. 7.3 Cohesion index according to the variable level of proficiency

	Basic level		Advanced level	
Cue	CI	Rank	CI	Rank
01 'Body'	0.27	1	0.27	1
02 'Clothes'	0.22	3	0.22	2
03 'House'	0.16	6	0.19	3
04 'Furniture'	0.15	8	0.17	5
05 'Food'	0.16	7	0.13	10
06 'Table'	0.17	5	0.16	6
07 'Kitchen'	0.13	11	0.15	7
08 'School'	0.15	9	0.15	8
09 'Lighting'	0.09	15	0.13	11
10 'City'	0.15	10	0.12	12
11 'Countryside'	0.11	13	0.12	13
12 'Transports'	0.24	2	0.15	9
13 'Garden'	0.06	16	0.07	16
14 'Animals'	0.2	4	0.18	4
15 'Games'	0.12	12	0.1	14
16 'Professions'	0.1	14	0.1	15

group. Except for the former, this tendency is also observed with slight variations in the Finnish participants (Carcedo 2000).

If we focus on the number of different units, we find that the two latter categories are precisely the ones that reveal a greater similarity between the two groups, whereas the most marked differences can be seen in 'Gardening and Farming', 'Food and Drink', 'Means of Transport', 'Animals', and 'Professions and Jobs'. Different results are attained by López González (2010) when considering the relative increase in the various lexical fields; yet according to this parameter, the prompt 'Objects Placed on the Table at Meals' stands out (+120.5 %), followed by 'The Countryside' (+100 %). However, 'Gardening and Farming' and, to a lesser extent, 'Food and Drink', also show a notable increase of word types in each group.

In the present study, we also set out to ascertain whether the relationship between word tokens and word types turns out to be similar for the two groups. With this purpose in mind, we looked at the cohesion degree, a concept created by Max S. Echeverría, which is calculated by dividing the words' average by the number of different units, and that aims to find which are the domains with the greatest coincidence among the words given by the informants (i.e. the most cohesive prompts). Table 7.3 displays the indices obtained for the various cue words, as well as their position according to this parameter. The cue words in the highest positions are the most compact and cohesive; in other words, they represent the word responses shared by a great percentage of students.

As can be observed, there is total coincidence in the highest and lowest cue words: in both groups, the most compact is 'Parts of the Body' and the most diffuse is 'Gardening and Farming'. Overall, this result corroborates the tendency reported in most studies on lexical availability. There are, however, some aspects that must

be pointed out. On the one hand, in our study the cohesion indices for both levels are systematically higher than the ones reported in other studies; this could be explained by the lower number of participants in our sample, since there is an inversely proportional relationship between number of participants and cohesion degree. On the other hand, in our study the low index and ranking of 'Means of Transport' at the advanced level is particularly notorious; this cue word is usually among the most cohesive as shown in the basic group, where it holds an even higher position than the one reached by 'Clothes' or 'Animals'. Finally, the group of advanced learners demonstrated a greater degree of agreement regarding the cue words 'Parts of the House' and 'Furniture', which normally hold more intermediate positions; higher cohesion in the words retrieved by the advanced group compared to the basic level group is also observed regarding 'Lighting'.

The findings according to the cohesion index henceforth bear witness to some aspects that differentiate the higher level students both from their colleagues with a lower linguistic competence and from native speaker subjects.

To sum up, the three aspects considered in this quantitative analysis– number of words, number of different units and cohesion index – have allowed us to confirm the effect of this factor in order to explain intergroup differences.

7.7 Qualitative Analysis

In this section, we compare the word types retrieved by the two groups. So that the comparison is reliable, we do not contrast the whole set of different units, but establish a borderline of 75 % of the accumulated frequency (sum of relative frequencies). This decision allows us, on the one hand, to gather a number of types relevant to each cue word; on the other hand, being a relative index, it gives us the possibility of working with a number of different lexical units appropriate to the input received by students for each cue word and to the distribution regarding the availability index of the units (for more information on this aspect, see Samper and Samper Hernández 2007).

In the qualitative analysis, we look at both the common lexicon within the 75 % accumulated frequency limit of the two lists, and at those units that are within the specified limit of a lexical list, but are not present in that of the contrasting group.[7]

Table 7.4 shows the number of different units selected for each cue word taking into account the 75 % accumulated frequency index.

Accordingly, we shall present the percentage of word types shared by the two groups under comparison. In calculating the percentage, we considered the maximum possible amount of coincidences, which corresponds to the number of word types retrieved by the group with the lowest figures within the 75 % accumulated frequency limit (see Table 7.4).

[7] We also include two appendices with those words which appear within the 75 % accumulated frequency limit in one of the levels and which are placed in lower positions in the other level.

Table 7.4 Number of word types that reach 75 % of the accumulated frequency in each level of proficiency

Cue	Basic level	Advanced level
01 'Body'	16	21
02 'Clothes'	14	17
03 'House'	16	17
04 'Furniture'	15	12
05 'Food'	45	67
06 'Table'	9	14
07 'Kitchen'	21	21
08 'School'	21	24
09 'Lighting'	25	16
10 'City'	44	51
11 'Countryside'	37	41
12 'Transports'	11	22
13 'Garden'	24	48
14 'Animals'	18	32
15 'Games'	38	53
16 'Professions'	45	59
Total	399	515

Table 7.5 Common words within the 75 % accumulated frequency limit according to the factor of level of proficiency

Cue	Rank	Common words	Percentage of coincidence
01 'Body'	1	16	100
02 'Clothes'	5	12	85.7
03 'House'	8	12	75
04 'Furniture'	9	9	75
05 'Food'	6	37	82.2
06 'Table'	4	8	88.9
07 'Kitchen'	12	13	61.9
08 'School'	7	17	80.9
09 'Lighting'	15	8	50
10 'City'	10	30	68.2
11 'Countryside'	14	20	54
12 'Transports'	2	11	100
13 'Garden'	16	6	25
14 'Animals'	3	17	94.4
15 'Games'	13	23	60.5
16 'Professions'	11	28	62.2
Total		267	66.9

In the first place, it emerges that there is a high degree of coincidence between the two groups; a great percentage of words are shared. In 13 domains, this convergence exceeds 60 % and in only one cue word, 'Gardening and Farming', falls below 50 %. This outcome is even more important if we consider that we tested learners of very different characteristics on the basis of their different cultural backgrounds, mother tongues and educational systems. These learners had not had the opportunity to form a close group in the weeks they spent in Salamanca. This is a very different situation from the one reported in lexical availability studies on Spanish native

speakers as well as on learners of Spanish as a foreign language such as Carcedo's (2000) with Finnish students or Sandu (2009) with Romanian students, in which the participants who make up the sample form more closely-related groups.

The cue words which offer the greatest coincidence are, undoubtedly, 'Parts of the Body' and 'Means of Transport', both with a compatibility of 100 %. The absolute convergence in the second lexical field is particularly notorious; yet, though generally quite compact, in our data the cohesion index is below 'Parts of the Body', 'Clothes', 'Animals', and 'Parts of the House'. Therefore, we may assume that there is a basic vocabulary related to 'Means of Transport', which is acquired in the first levels of Spanish and is maintained with a high degree of availability throughout the consecutive stages of the learning process. These lexical items hold the first positions, whereas the rest of the lexical units are more diverse. This characteristic could also explain the differences between the two groups as far as lexical cohesion index is concerned; this prompt turns out to be very compact at the basic level – where the participants fundamentally activate the essential lexicon of the thematic category – and more diffuse in the advanced level – where they also include other words which do not have that high degree of coincidence among the subjects' responses.

The lexical fields that show less coincidence are 'Gardening and Farming', 'Lighting, Heating and Means of Airing Places', and 'The Countryside'. The three cue phrases are characterized by a high lexical dispersion, something that can be observed from the very first units on the lists. In the case of 'Lighting', however, a notable difference is found when calculating the cohesion index for each group; yet the index is higher in the advanced group than in the basic group (see Table 7.3). Such divergence can be explained by learners' deficit on knowledge on words related to this domain on the part of students with the lower proficiency level. The learners in the basic level include wider associative relationships as shown by a low coincidence in their responses to the same prompt. In contrast, the advanced group with a similar number of word types (see Table 7.2) produce responses more related to the cue word, hence the higher cohesion. We will turn again to this aspect later on when we analyze the lexical units which stand for the difference between both groups.

Next we move on to present the 267 words within the 75 % accumulated frequency shared by basic and advanced Spanish learners. This is, therefore, the common lexicon for both levels of proficiency, which can be considered as essential because it corresponds to the lexical items which are learnt first and, what is more, maintained as students progress towards proficiency in Spanish. These words should hence form part of the vocabulary which should be taught in the first lessons of Spanish as a foreign language (Table 7.6).

The common words in the first positions could be considered essential in the various thematic categories in which they appear. This characteristic is more evident in the case of lexical fields where a reduced number of words make up 75 % of the accumulated frequency, yet they are retrieved by the majority of the informants and are in high positions on the lists. An example is found in 'Objects Placed on the Table at Meals', in which both groups retrieve the fundamental words from this cue word. Moreover, on both lists the first six positions are held by words referring to objects, whereas words such as *sal* (salt) and *pimienta* (pepper) – here used with an

Table 7.6 Word types which in both levels of proficiency are placed in the 75 % accumulated frequency limit

01 'Parts of the Body'	*brazo* (arm), *boca* (mouth), *cabeza* (head), *cara* (face), *dedo* (finger), *diente* (tooth), *espalda* (back), *estómago* (stomach), *labio* (lip), *mano* (hand), *nariz* (nose), *ojo* (eye), *oreja* (ear), *pelo* (hair), *pie* (foot), *pierna* (leg)
02 'Clothes'	*abrigo* (coat), *bufanda* (scarf), *calcetín* (sock), *camisa* (shirt), *camiseta* (T-shirt), *chaqueta* (jacket), *falda* (skirt), *jersey* (jumper), *pantalón* (trousers), *vaquero* (jeans), *vestido* (dress), *zapato* (shoe)
03 'Parts of the House'	*balcón* (balcony), *baño* (bathroom), *cocina* (kitchen), *comedor* (dining room), *cuarto de baño* (bathroom), *dormitorio* (bedroom), *habitación* (room), *puerta* (door), *salón* (living room), *suelo* (floor/ground), *techo* (ceiling), *ventana* (window)
04 'Furniture (House)'	*armario* (wardrobe), *cama* (bed), *escritorio* (desk), *lámpara* (lamp), *mesa* (table), *silla* (chair), *sillón* (armchair), *sofá* (sofa), *televisión* (television)
05 'Food and Drink'	*aceite* (oil), *agua* (water), *arroz* (rice), *azúcar* (sugar), *café* (coffee), *carne* (meat), *cerveza* (beer), *chocolate* (chocolate), *chorizo* (hard pork sausage), *coca-cola* (coca-cola), *ensalada* (salad), *fanta* (fanta), *fruta* (fruit), *jamón* (ham), *leche* (milk), *lechuga* (lettuce), *lenteja* (lentil), *manzana* (apple), *naranja* (orange), *pan* (bread), *pasta* (pasta), *patata* (potato), *pescado* (fish), *piña* (pineapple), *pizza* (pizza), *plátano* (banana), *pollo* (chicken), *queso* (cheese), *sangría* (sangria), *sopa* (soup), *té* (tea), *tomate* (tomato), *tortilla* (omelette), *verdura* (vegetables), *vino* (wine), *zanahoria* (carrot), *zumo* (juice)
06 'Objects Placed on the Table at Meals'	*cuchara* (spoon), *cuchillo* (knife), *plato* (plate), *pimienta* (pepper), *sal* (salt), *servilleta* (serviette), *tenedor* (fork), *vaso* (glass)
07 'The Kitchen and its Utensils'	*cuchara* (spoon), *cuchillo* (knife), *frigorífico* (fridge), *grifo* (tap), *horno* (oven), *lavadora* (washing machine), *mesa* (table), *microondas* (microwave), *nevera* (fridge), *plato* (plate), *refrigerador* (refrigerator/fridge), *tenedor* (fork), *vaso* (glass)
08 'School: Furniture and Materials'	*bolígrafo* (pen), *borrador* (eraser), *cuaderno* (notebook), *diccionario* (dictionary), *goma de borrar* (rubber), *lápiz* (pencil), *libro* (book), *luz* (light), *mesa* (table), *ordenador* (computer), *papel* (sheet/piece of paper), *pizarra* (blackboard), *pluma* (fountain pen), *silla* (chair), *television* (television), *tiza* (chalk), *ventana* (window)
09 'Lighting, Heating and Means of Airing Places'	*aire acondicionado* (air conditioning), *calefacción* (heat), *estufa* (heater), *fuego* (l*ámpara* lamp), *luz* (light), *radiador* (radiator), *ventana* (window)
10 'The City'	*árbol* (tree), *autobús* (bus), *avenida* (avenue), *banco* (bench/bank), *bar* (bar), *café (*café), *calle* (street), *casa* (house), *catedral* (cathedral), *coche* (car), *discoteca* (disco/night club), *edificio* (building), *escuela* (school), *farmacia* (pharmacy), *gente* (people), *hotel* (hotel), *iglesia* (church), *metropolitano* (tube), *museo* (museum), *oficina* (office), *parque* (park), *piso* (flat), *plaza* (square), *policía* (police/policeman), *restaurante* (restaurant), *semáforo* (traffic light), *supermercado* (supermarket), *taxi* (taxi/cab), *tienda* (shop), *universidad* (university)

(continued)

Table 7.6 (continued)

11 'The Countryside'	*animal* (animal), *árbol* (tree), *bosque* (forest), *caballo* (horse), *campesino* (farmer), *campo* (field), *casa* (house), *conejo* (rabbit), *flor* (flower), *hierba* (grass), *jardín* (garden), *lago* (lake), *montaña* (mountain), *nieve* (snow), *perro* (dog), *río* (river), *sol* (sun), *tierra* (soil/land/earth), *toro* (bull), *vaca* (cow)
12 'Means of Transport'	*autobús* (bus), *avión* (plane), *barco* (ship), *bicicleta* (bike/bicycle), *caballo* (horse), *camión* (lorry), *coche* (car), *motocicleta* (motorbike), *pie* (foot), *taxi* (taxi/cab), *tren* (train)
13 'Gardening & Farming'	*campesino* (farmer), *cortar árboles* (to cut down trees), *jardinero* (gardener), *plantar* (to plant), *poner agua* (to put water), *tierra* (soil/land)
14 'Animals'	*caballo* (horse), *cerdo* (pig), *elefante* (elephant), *gato* (cat), *león* (lion), *mono* (monkey), *mosca* (fly), *oso* (bear), *oveja* (sheep), *pájaro* (bird), *perro* (dog), *pez* (fish), *pollo* (chick), *ratón* (mouse), *serpiente* (snake), *toro* (bull), *vaca* (cow)
15 'Games & Entertainment'	*bailar* (dancing), *baloncesto* (basketball), *béisbol* (baseball), *correr* (running), *deporte* (sport), *escribir* (writing), *escuchar música* (listening to music), *esquiar* (skiing), *fútbol* (soccer), *fútbol americano* (American football), *hockey* (hockey), *ir al cine* (going to the cinema), *jugar a las cartas* (playing cards), *leer* (reading), *música* (music), *nadar* (swimming), *natación* (swimming), *rugby* (rugby), *televisión* (television), *tenis* (tennis), *tenis de mesa* (table tennis/ping pong), *ver televisión* (watching TV), *voleibol* (volleyball)
16 'Professions and Jobs'	*abogado* (lawyer), *actor* (actor), *camarero* (waiter), *cantante* (singer), *conductor* (driver), *dentista* (dentist), *dependiente* (shop assistant), *director* (director), *doctor* (doctor), *enfermero* (nurse), *escritor* (writer), *estudiante* (student), *fontanero* (plumber), *ingeniero* (engineer), *jefe* (boss), *jugador* (player), *maestro* (teacher), *médico* (doctor), *peluquero* (hairdresser), *periodista* (journalist), *piloto* (pilot), *policía* (policeman), *portero* (doorman/goalkeeper), *presidente* (president), *profesor* (teacher/professor), *secretario* (secretary/clerk), *taxista* (taxi driver/cab driver), *traductor* (translator)

obvious metonymic value - appear further down on the list. In the domain of 'Means of Transport', where only 11 different words – with a very high presence in the group of basic level – are compared, the basic nature of this lexicon stands out. In the two lists, the first position is occupied by *coche* (car) as expected, whereas the following three positions are held by *autobús* (bus), *avión* (plane), and *tren* (train), though not in the same order. The comparison of these first 11 words with the 12 shared – in the limits of 0.1 availability – in the study by López González (2010) allows us to corroborate the existence of an essential lexicon in the field of 'Means of Transport', yet nine out of ten common words retrieved by the group of Spanish learners at Salamanca are in the first positions on lists reported by López González on Spanish learners in Poland from two educational levels. In the latter lists, the word *camión* (lorry) does not appear with such high availability, whereas in our

research *metro* (tube) and *tranvía* (trolley) are not shared in these first positions. The seven lexical units shared among the four levels analyzed by Carcedo (2000) in Finland – *coche* (car), *autobús* (bus), *tren* (train), *avión* (plane), *bicicleta* (bike) (bicycle), *barco* (ship), *taxi* (taxi) – confirm, once again, the fundamental nature of this vocabulary.

In order to establish the qualitative comparison between the different levels of proficiency, it is important to know both the shared and the non-shared words by the two groups analyzed. This comparison reveals objective differences concerning the available lexicons of different language groups. Tables 7.7 and 7.8 display the exclusive words retrieved in each group distributed by cue words.

On the basis of this parameter, three tendencies are revealed:

1. Domains in which all the words included within the 75 % accumulated frequency limit are present on the list of the opposing group. These are: 'Parts of the Body', 'Clothes', and 'School: Furniture and Materials'.
2. Domains in which all the absent units are concentrated on the basic level list: 'Parts of the House', 'Objects Placed on the Table', 'Means of Transport' and 'Animals'.
3. Cues in which there are words absent on both lists. Most of the lexical domains are included in this group: 'Furniture' (House), 'Food and Drink', 'The Kitchen and its Utensils', 'Lighting', 'The City', 'The Countryside', 'Gardening and Farming', 'Games and Entertainment', and 'Professions and Jobs'.

Logical space constraints oblige us to select, from each group, some lexical fields which are worth commenting on:

Regarding the first tendency, from the three cue words that have no absent terms, two are more compact lexical fields; indeed, 'Parts of the Body' is the most compact domain and 'Clothes' can be found among the first three areas in both levels of proficiency. On the contrary, 'School' is an intermediate cue according to the cohesion index, hence greater divergences could be expected to be found when collating. We should, then, assume that there is a wide range of shared vocabulary, which is soon learnt and which is maintained as very available when talking about cues to school-related activities.

As to the second tendency, we would like to highlight what happens with the cue 'Animals': in this case, it can be seen that three of the five lexical units absent on their list – *pitón* (python), *iguana* (iguana) and *foca* (seal) – are clearly of an encyclopedic nature, and show the superiority of the advanced group. This, together with a high qualitative convergence in the first positions on both lists, allows us to confirm adequate development in the acquisition of vocabulary related to 'Animals': there is a group of terms which students of both levels know and use naturally when needed in conversation; these words refer to animals with a high frequency occurrence in daily life, the first positions are held by *perro* (dog), *gato* (cat), *pájaro* (bird), *vaca* (cow). At the same time, the students with a higher competence gradually produce new lexical units, beyond the basic communicative needs.

The last tendency gathers various cue words which present absent units on both lexical lists. However, some questions must be clarified: (i) These are cues which are generally less compact, hence the greater differences between the sample subgroups. (ii) Naturally, there is a clear quantitative difference between the absent

Table 7.7 Words which do not show up on the basic level group lists, being within the limits of 75 % on the advanced ones

Cue word	Words not listed in the basic level
03 Parts of the House	*piso* (floor)
04 Furniture (house)	*puerta* (door), *estantería* (shelf)
05 Food and Drink	*chupito* (shot), *vinagre* (vinegar), *soda* (soda water), *ajo* (garlic), *hamburguesa* (hamburger), *mayonesa* (mayonnaise), *calabacín* (courgette/zucchini), *lomo* (loin), *tarta* (cake), *postre* (dessert)
06 Table	*copa* (cup), *mantel* (tablecloth), *pañuelo* (kerchief)
07 The Kitchen	*cazuela* (cooking pot), *bombona* (gas bottle), *cazo* (casserole pot), *lavavajillas* (dishwasher)
09 Lighting, Heating and Means of Airing Places	*neón* (neon), *enchufe* (socket), *hoguera* (bonfire), *vela* (candle), *hogar* (fireplace), *chimenea* (fireplace)
10 The City	*paseo* (promenade), *ayuntamiento* (town hall), *basura* (rubbish), *barrio* (suburb), *palacio* (palace), *lámpara* (lamp), *fuente* (fountain), *librería* (bookshop)
11 The Countryside	*camino* (road/route), *planta* (plant), *cazador* (hunter), *coche* (car), *toro de anuncio* (advertisement bull), *encina* (oak), *prado* (field/meadow/lawn), *gallina* (hen), *parque* (park), *hoja* (leaf), *insecto* (insect), *iglesia* (church), *colina* (hill), *hormiga* (ant), *tractor* (tractor)
12 Means of Transport	*aeropuerto* (airport), *tráfico* (traffic), *carretera* (road), *estación de tren* (train station), *tranvía* (tram/streetcar), *parada de autobús* (bus stop)
13 Gardening and Farming	*cultivar* (to cultivate), *planta* (plant), *flor* (flower), *tractor* (tractor), *pala* (shovel), *cortar* (to cut off/shop), *agua* (to water), *regar* (to water), *hacer verdura* (to cook vegetables), *cortar* plantas (to cut off/shop plants), *hacer fruta* (to make fruit), *tijeras* (scissors), *coger* (to take/pick up), *cortar hierba* (to cut off/shop grass), *recoger* (to pick up/to collect), *crecer* (to grow up), *semilla* (seed), *vender* (to sell), *caballero* (horse rider), *camponista* (*lexical creation (as happens in"harvestar" y "hortinero")), *vaso* (glass/vase), *campana* (bell), *florecer* (to bloom/blossom), *crecer animales* (to grow animals), *trabajar* (to work), *hoz* (sickle), *pastor* (shepherd), *dar agua* (to give water), *coger flores* (to pick up flowers), *meter* (to prepare), *barbacoa* (barbecue), *limpiar* (to clean), *cultivo* (crop/planting), *jardín* (garden), *agrícola* (agriculture), *cazar* (to hunt), *jarrón* (vase), *ganadería* (ranching), *machete* (machete), *barro* (mud), *montar a caballo* (to ride a horse), *harvestar* (*lexical creation)
14 Animals	*cabrón* (Billy goat), *mariposa* (butterfly), pitón (python), *iguana* (iguana), *foca* (seal)
15 Games and Entertainment	*ir de compras* (go shopping), *nintendo* (nintendo), *billar* (billiards), *charlar* (chatting), *discoteca* (disco), *piragüismo* (canoeing), *gimnasia* (gymnastics), *ajedrez* (chess), *maratón* (marathon), *juego de mesa* (board game), *dar paseo* (going for a walk), *tocar guitarra* (playing guitar), *charlar con amigos* (chatting with friends), *hacer deporte* (doing sports), *juego artificial* (artificial game), *karaoke* (karaoke), *juego* (game)
16 Professions and Jobs	*zapatero* (shoemaker/shoe repairer), *carnicero* (butcher), *azafata* (stewardess), *intérprete* (interpreter), *entrenador* (coach), *carpintero* (carpenter), *músico* (musician), *cartero* (postman/mailman), *empleado* (employee), *emperador* (emperor), *obrero* (worker), *basurero* (garbage collector), *panadero* (baker), *limpiador* (cleaner), *despachador* (shop assistant), *bailarín* (dancer), *pescadería* (fish shop), *cazador* (hunter), *paro* (unemployment), *soldado* (soldier), *cultivador* (farmer), *mendigo* (beggar), *agente de bolsa* (broker), *futbolista* (soccer player)

Table 7.8 Words which do not show up on the advanced level group lists, being within the limits of 75 % on the basic ones

Cue word	Words not listed in the advanced level
04 Furniture (house)	*espejo* (mirror), *armarillo* (bathroom cabinet)
05 Food and Drink	*legumbre* (legume), *espagueti* (spaghetti)
07 The Kitchen	*agua* (water), *comida* (food), *lavaplatos* (dishwasher)
09 Lighting, Heating and Means of Airing Places	*calor* (heat), *frío* (cold), *aire fresco* (fresh air), *calefacción central* (central heating), *nieve* (snow), *luna* (moon), *calefacción de gas* (gas heating), *abrir ventana* (to open the window), *pared* (wall), *climatizor* (*lexical creation)
10 The City	*apartamento* (small flat), *río* (river), *aeropuerto* (airport), *centro* (downtown), *teatro* (theatre), *flor* (flower)
11 The Countryside	*oso* (bear), *piedra* (stone), *serpiente* (snake), *paz* (peace), *pasto* (pasture), *autopista* (motorway/highway), *niebla* (fog)
13 Gardening and Farming	*jardinería* (gardening), *vaquero* (cowboy), *florista* (florist), *coger frutas* (to pick up fruit), *cortar césped* (to mow the grass), *cortar* to (mow the pasture), *esclavo* (slave), *hortinero* (*lexical creation), *conducir tractor* (to drive a tractor), *aguar flores* (to water plants), *poner flores* (to plant), *regar césped* (to water the grass), *conducir camión* (to drive a lorry), *preparar tierra* (to prepare the soil), *arquitecto de jardines* (architect of gardens)
15 Games and Entertainment	*caminar* (walking), *mirar televisión* (watching television), *dibujar* (drawing), *fotografía* (photography), *básquetbol* (basketball), *salir por la noche* (going out at night)
16 Professions and Jobs	*trabajador* (worker), *hombre de negocios* (businessman), *pastor* (shepherd), *trabajar en un banco* (to work at a bank), *jugador de deporte* (sport player), *recepcionista* (receptionist), *chófer* (driver), *toreador* (bullfighter)

lexical units of each group of learners. The only domain which shows more absences in the responses of the students with a higher level is 'Lighting'. (iii) In many cases, a qualitative difference can be appreciated regarding the lexical terms which are not present in each level of proficiency and which appear within the 75 % accumulated frequency limit in the other one: the units absent from the lists of the advanced level students are usually less concrete and not so directly related to the stimulus in question. The best example can be found precisely in 'Lighting', with the absence of *neón* (neon), *enchufe* (socket), *hoguera* (bonfire), *vela* (candle), *hogar* (fireplace), and *chimenea* (fireplace) on the beginner students' list, opposed to units such as *calor* (heat), *frío* (cold), *aire fresco* (fresh air), *nieve* (snow), *luna* (moon), or *abrir ventana* (to open the window), which do not appear on the higher level list.[8] In The Kitchen and its Utensils, taking this case as another example, the learners from the basic group do not write *cazuela* (cooking pot), *bombona* (gas bottle),

[8] The qualitative contrast between the types of lexical units given by one and another group of learners explain why subjects from the advanced level register a greater amount of lexical absences, as was stated in Sect. ii.

cazo (casserole pot) and *lavavajillas* (dishwasher), whereas the subjects from the high level do not include *agua* (water), *comida* (food) and, this one more specific, *lavaplatos* (dishwasher).

The cue word with most absences is 'Gardening and Farming'. However, some of these lexical deficiencies could be justified by the nature of this particular semantic stimulus. Firstly, given its complexity, the participants retrieve less concrete words, and vocabulary which is not specific to the thematic category, hence there is a much lesser probability of shared vocabulary between the groups. On the other hand, many unique lexical creations show up with this cue word, due to the linguistic deficit of the students. What is more, it must be taken into account that this cue favours the appearance of complex lexical units -mostly verb + direct object –,which makes the intergroup convergence even more difficult. If we revise the absences of different units in each group, a divergence between the levels of proficiency can be seen: complex lexical units represent only 19 % of the total for the learners with a lower proficiency level, whereas for their counterparts with a higher Spanish level this percentage reaches 66.7 %. This objective distance is intensified by the nature of the different units which are absent on each list; if we leave lexical creations and compound words aside, whereas for the participants in the advanced group the only absence regarding simple units are *jardinería* (gardening) but they do include *jardinero* (gardener), *vaquero* (cowboy), *florista* (florist), and *esclavo* (slave), the students from the basic level do not include on their lists words which are more directly related to the thematic domain and also more linked to formal teaching, such as *pala* (shovel), *semilla* (seed), *hoz* (sickle), *pastor* (shepherd), or *machete* (machete).

7.8 Conclusions

The results of this study allow us to confirm the importance of 'level of proficiency' on the lexical availability of learners of Spanish as a foreign language (in an immersion context). The students from the advanced level of our sample systematically overperform their counterparts regarding the production of word tokens and word types. A greater similarity between the two groups can be seen in the cohesion index, as happens in most lexical availability studies. However the present study also shows some divergence on the cue phrases 'Lighting, Heating and Means of Airing Places' and 'Means of Transport'.

The different lexical units retrieved by the two groups show a high level of coincidence: 66.9 %, when taking into account the terms which represent 75 % of the accumulated frequency for each cue word. The domains 'Parts of the body' and 'Means of Transport' present total compatibility, whereas 'Gardening and Farming', 'Lighting', 'Heating and Means of Airing Places' and 'The Countryside' reveal the greatest divergences. Both the shared and non shared words in the two lists contain relevant material for the teaching of Spanish as a foreign language, particularly suitable to programme vocabulary learning according to objective criteria.

Appendices

Table 7.9 Word types retrieved by the basic group at 75 % of accumulated frequency and which appear out of this limit in the advanced level

02 Clothes	*suéter* (sweater) (21), *zapatilla* (slipper) (34)
03 Parts of the House	*sala* (living room) (45), *escalera* (stairs) (19), *garaje* (garage) (33), *servicio* (toilet) (21)
04 Furniture (house)	*alfombra* (rug/carpet) (18), *luz* (light) (38), *ducha* (shower) (52), *horno* (oven) (16)
05 Food and Drink	*melocotón* (peach) (111), *uva* (grape) (147), *ternera* (veal) (73), *tapa* (snack) (134), *pera* (pear) (76), *fresa* (strawberry) (97)
06 Objects Placed on the Table at Meals	*taza* (cup) (31)
07 The Kitchen and its Utensils	*silla* (chair) (23), *botella* (bottle) (56), *taza* (cup) (30), *congelador* (freezer) (28), *batidora* (mixer) (43)
08 School: Furniture and Materials	*puerta* (door) (33), *servicio* (toilet) (38), *pupitre* (desk) (39), *aula* (classroom) (28)
09 Lighting, Heating and Means of Airing Places	*sol* (sun) (41), *viento* (wind) (43), *ventilador* (fan) (17), *manta* (blanket) (40), *agua* (water) (50), *electricidad* (electricity) (19), *gas* (gas) (48)
10 The City	*cine* (movie theatre) (56), *estación de autobuses* (bus station) (63), *correo* (post office) (53), *estación de tren* (train station) (109), *camión* (lorry) (110), *tren* (train) (130), *tráfico* (traffic) (76), *puente* (bridge) (75)
11 The Countryside	*pájaro* (bird) (56), *mar* (sea) (90), *pueblo* (town/village) (45), *agua* (water) (93), *cielo* (sky) (97), *carretera* (road) (44), *estrella* (star) (73), *pollo* (chicken) (66), *gato* (cat) (75), *playa* (beach) (42)
13 Gardening and Farming	*granjero* (farmer) (69), *agricultura* (agriculture) (77), *agricultor* (farmer) (65)
14 Animals	*rana* (frog) (34)
15 Games and Entertainment	*golf* (golf) (110), *leer libros* (reading books) (55), *pintar* (painting) (99), *escribir cartas* (writing letters) (81), *dormir* (sleeping) (77), *salir* (going out) (109), *atletismo* (athletics) (85), *tocar instrumento* (playing an instrument) (80), *beber* (drinking) (79)
16 Professions and Jobs	*artista* (artist) (82), *actriz* (actress) (60), *negocios* (business) (65), *bombero* (fireman) (100), *jugador de fútbol* (football player) (98), *arquitecto* (architect) (121), *manager* (manager) (73), *pintor* (painter) (115), *cientista* (*lexical gap) (64)

The exact position held by each of these words in the contrasting group is shown as well

Table 7.10 Word types retrieved by the advanced group at 75 % of accumulated frequency and which appear out of this limit in the basic level

01 Parts of the Body	*cuello* (neck) (48), *uña* (nail) (27), *pecho* (chest/breast) (21), *culo* (bottom) (25), *ceja* (eyebrow) (29)
02 Clothes	*corbata* (tie) (17), *guante* (glove) (20), *sombrero* (hat) (16), *braga* (knickers) (34), *gorra* (cap) (25)
03 Parts of the House	*pared* (wall) (17), *pasillo* (corridor) (18), *jardín* (garden) (23), *terraza* (balcony) (42)
04 Furniture (house)	*mesita* (small table) (31)
05 Food and Drink	*cerdo* (pork) (55), *helado* (ice cream) (114), *cebolla* (onion) (57), *sal* (salt) (107), *vino tinto* (red wine) (93), *limón* (lemon) (117), *pimiento* (pepper) (71), *yogur* (yoghurt) (62), *huevo* (egg) (47), *caña* (small glass of beer) (73), *paella* (paella) (113), *bocadillo* (sandwich made with French bread) (94), *sprite* (sprite) (61), *mantequilla* (butter) (50), *pastel* (cake/pie) (100), *tequila* (tequila) (70), *güisqui* (whisky) (82), *merluza* (hake) (68), *galleta* (cookie) (121), *vodka* (vodka) (63)
06 Objects Placed on the Table at Meals	*botella* (bottle) (15), *agua* (water) (32), *pan* (bread) (34)
07 The Kitchen and its Utensils	*sartén* (frying pan) (27), *olla* (pot) (43), *cucharón* (ladle) (34), *lavabo* (sink) (25)
08 School: Furniture and Materials	*mochila* (backpack) (25), *radio* (radio) (28), *escritorio* (desk) (23), *texto* (text) (61), *bolso* (bag) (49), *reloj* (clock) (46), *clase* (classroom) (36)
09 Lighting	*calefactor* (heater) (35), *bombilla* (light bulb) (33)
10 The City	*monumento* (monument) (56), *jardín* (garden) (59), *hospital* (hospital) (51), *quiosco* (kiosk) (78), *carretera* (road) (89), *carnicería* (butchers) (102), *plaza mayor* (main square) (61), *biblioteca* (library) (54), *autopista* (highway) (77), *atasco* (traffic jam) (67), *empresa* (business) (121), *perro* (dog) (60), *panadería* (bakery) (85)
11 The Countryside	*naturaleza* (nature) (49), *verdura* (vegetable) (67), *césped* (grass) (47), *finca* (country house) (48), *oveja* (sheep) (40)
12 Means of Transport	*metro* (tube) (12), *moto* (motorcycle) (13), *autostop* (hitch hiking) (16), *autocar* (coach/bus) (17), *carro* (car/cart/trolley) (39)
14 Animals	*gallina* (hen) (23), *conejo* (rabbit) (22), *tigre* (tiger) (21), *gallo* (rooster) (20), *cabra* (goat) (35), *burro* (donkey) (27), *mosquito* (mosquito) (38), *pavo* (turkey) (40), *pato* (duck) (51), *cucaracha* (cockroach/roach) (36)
15 Games and Entertainment	*monopoly* (monopoly) (47), *pasear* (going for a walk) (76), *gimnasio* (gym) (64), *montar a caballo* (riding) (42), *patinar* (roller skating) (41), *ver película* (watching a film) (57), *cantar* (singing) (45), *hacer ejercicio* (doing exercise) (89), *esquí* (skiing) (63), *comer* (eating) (79), *hacer footing* (jogging) (56), *llamar por teléfono* (phoning) (81), *viajar* (travelling) (43)
16 Professions and Jobs	*ama de casa* (housewife) (49), *cocinero* (cook) (85), *vendedor* (salesman) (64), *campesino* (farmer) (55), *banquero* (banker) (48), *político* (politician) (83), *ministro* (minister/secretary) (103)

References

Carcedo, A. 1998. Sobre las pruebas de disponibilidad léxica para estudiantes de español como lengua extranjera. *RILCE* 14(2): 205–224.

Carcedo, A. 1999. Desarrollo de la competencia léxica en español LE: análisis de cuatro fases de disponibilidad. *Pragmalingüística* 5–6: 75–94.

Carcedo, A. 2000. *Disponibilidad léxica en español como lengua extranjera: el caso finlandés*. Turku: Turun Yliopisto.

Galloso, M.V., and J. Prado. 2005. La estructura estadística del léxico disponible de informantes nativos e informantes de ELE. In *Las gramáticas y los diccionarios en la enseñanza del español como segunda lengua: deseo y realidad. Actas del XV Congreso Internacional de ASELE*, coords. María Auxiliadora Castillo et al., 370–375. Sevilla: Universidad de Sevilla.

Lin, J. 2006. *Disponibilidad léxica de alumnos chinos estudiantes de español como lengua extranjera*. Final master's report, Universidad de Alcalá.

López González, A.M. 2010. La evaluación del desarrollo de la competencia léxica en L2 por medio de la disponibilidad léxica. redELE Red Electrónica de Didáctica del Español como Lengua Extranjera18.http://www.educacion.gob.es/dctm/redele/MaterialRedEle/Revista/2010_18/2010_redELE_18_01Lopez.pdf?documentId=0901e72b80dd3177/. Accessed 28 Nov 2011.

López Rivero, E. 2008. *Estudio de disponibilidad léxica en 43 estudiantes de ELE*. Final master's report, Universidad Antonio de Nebrija.

Medina Arejita, E. 2009. *Las nociones específicas del PCIC y la disponibilidad léxica como instrumento de selección del vocabulario*. Final master's report, Instituto Cervantes and Universidad Internacional Menéndez Pelayo.

Moreno, F., J. Enrique Moreno, and A.J. García de las Heras. 1995. Cálculo de la disponibilidad léxica. El programa LexiDisp. *Lingüística* 7: 243–249.

Pérez Serrano, M. 2009. *Estudio de disponibilidad léxica en estudiantes de E/LE en los centros de interés "Medios de transporte" y "Profesiones y oficios"*. Final master's report, Instituto Cervantes and Universidad Internacional Menéndez Pelayo.

Samper, J.A. 1998. Criterios de edición del léxico disponible. *Lingüística* 10: 311–333.

Samper Hernández, M. 2001. Dificultades de los estudios de disponibilidad léxica en ELE: los criterios de edición de los materiales. In *Nuevas aportaciones al estudio de la lengua española. Investigaciones filológicas*, ed. José Antonio Bartol et al., 277–285. Salamanca: Luso-Española de Ediciones.

Samper Hernández, M. 2002. *Disponibilidad léxica en alumnos de español como lengua extranjera*. Málaga: ASELE.

Samper Hernández, M. 2009. *Evolución de la disponibilidad léxica en estudiantes grancanarios de enseñanza primaria y secundaria*. Las Palmas de Gran Canaria: Cabildo de Gran Canaria/ Casa de Colón.

Samper, J.A., and M. Samper Hernández. 2006. Aportaciones recientes de los estudios de disponibilidad léxica. *LynX: Panorámica de estudios lingüísticos* 5: 5–95.

Samper, J.A., and M. Samper Hernández. 2007. El proyecto panhispánico de disponibilidad léxica y los cotejos dialectales. *Lingüística* 19: 88–116.

Samper, J.A., J.J. Bellón, and M. Samper Hernández. 2003. El proyecto de estudio de la disponibilidad léxica en español. In *Pautas y pistas en el análisis del léxico hispanoamericano*, ed. Gerd Wotjak, 27–140. Frankfurt/Madrid: Vervuert/Iberoamericana.

Sánchez Gómez, C. 2005. Naturaleza gramatical del léxico disponible en informantes de español como lengua extranjera. *Interlingüística* 16(2): 977–986.

Sánchez-Saus, M. 2009. La variable 'nivel de español' en el léxico disponible de los estudiantes de español como lengua extranjera. *Pragmalingüística* 17: 140–153.

Sandu, B. 2009. *Disponibilidad léxica en español como lengua extranjera de los alumnos rumanos Estudio exploratorio*. Diploma of Advanced Studies' report, Universidad de Las Palmas de Gran Canaria.

Šifrar Kalan, M. 2011. *Leksikalna Razpoložljivost V Španščini Kot Tujem Jeziku*. Ph.D. thesis, Univerza v Ljubljani.

Chapter 8
Slovene Students' Lexical Availability in English and Spanish

Marjana Šifrar Kalan

8.1 Introduction

Lexical availability (LA) has a prominent tradition in Spanish as a mother tongue and some aspects of LA research can definitively be applicable to foreign languages. After having compared the LA of Slovene students, learners of Spanish as a foreign language (SFL) with other learners Šifrar Kalan (2009), in the present study I wanted to explore the quantitative and qualitative differences and similarities in learners' LA in two foreign languages, English and Spanish. The present study compares the LA in eight semantic categories in two groups: Slovene students of English and Spanish, and presents the most available words and semantic prototypes to these groups. The results are also observed from the point of view of language proficiency and the years of study of English and Spanish.

8.2 Lexical Availability in English as a Second/Foreign Language

So far, most LA research has been done in Spanish as L1; although some important studies have also been carried out in English. In Chap. 1 of this volume, the pioneering research by the Yugoslav educator Dimitrijevic is mentioned. As early as 1969, this author published the results of LA of 185 Scottish secondary school students in Edinburgh. He focused on LA in English as L1 and looked at the effect

M. Šifrar Kalan (✉)
Department of Romance Languages and Literatures, Faculty of Arts,
University of Ljubljana, Ljubljana, Slovenia
e-mail: marjana.kalan@siol.net

R.M. Jiménez Catalán (ed.), *Lexical Availability in English and Spanish as a Second Language*, Educational Linguistics 17, DOI 10.1007/978-94-007-7158-1_8,

of intelligence and gender on the LA availability of Scottish secondary school students: the former proved to be an important variable while the latter did not. Dimitrijevic used open lists for 11 semantic fields (as opposed to a closed list consisting of 20 words). The same technique was followed closely by the American linguist Bailey (1971), whose study is also mentioned in Chap. 1. His investigation, on the LA of monolinguals and bilinguals of English and Spanish, is perhaps the only one of its kind. He compared the LA of 16–18-year-old high school students who were distributed into three groups: 33 monolinguals in Spanish (from Monterey, Mexico), 33 monolinguals in English (from Houston, Texas), and 33 bilinguals in Spanish and English (also from Houston, Texas). He chose ten semantic categories: 'Animals', 'City', 'Entertainment', 'Family', 'God', 'Jobs and Professions', 'Medicine', 'Music', 'Space' and 'War'. These were slightly different from the 16 categories used by Gougenheim et al. (1967), as well as by PanHispanic researchers (Samper Padilla 1998). The 5-min time period allotted for the responses was shortened to 2 min in the study conducted by Bailey. The students who took the lexical availability test in English produced 12 % more lexical items than the students who took the test in Spanish. The monolingual English group ranked highest in total items, followed by the bilingual group in Spanish, the bilingual group in English and finally, the monolingual group in Spanish. As to the different lexical items, however, the bilingual group in Spanish yielded the highest results, while the monolingual group in English was second. The English-speaking monolinguals obtained a higher mean in lexical output by centre of interest ('Animals' 25.84, 'City' 22.09, 'Entertainment' 12.79). The Spanish-speaking monolinguals produced a lower mean (*Animales* ('Animals') 22.76, *Ciudad* ('City') 19.45, *Diversión* ('Entertainment') than the Spanish-speaking bilinguals (*Animales* 22.76, *Ciudad* 19.45, *Diversión* 13.03). The bilingual group produced a very similar total lexical output in Spanish and English. These students took both tests on the same day; first in Spanish, then in English.

During the last two decades, lexical availability has focused almost exclusively on Spanish; hardly any research has been conducted on English as L1 or English as L2. Some exceptions are the studies by Germany and Cartes (2000) and Ferreira and Echeverria (2010) on Chilean English foreign language learners; the studies on primary and secondary school EFL learners (Jiménez-Catalán and Ojeda Alba 2009; Fernández Fontecha 2010), and the research conducted by Gallardo and Martínez Adrián on senior Basque EFL learners. Except for Ferreira and Echeverría (2010), these studies are descriptive and correlational. They focus on learners' lexical production according to the different semantic categories and on the effect of age, sex or type of instruction on the lexical availability output of these learners.

As mentioned in the introduction to this chapter, the present study looks at Slovene students' lexical competence in two foreign languages: English and Spanish. As far as I know, except for Bailey's study on bilinguals of English and Spanish in immersion context, no comparative studies have been carried out on learners' lexical competence in two languages, and certainly not on Slovene English and Spanish FL learners.

8.3 Methodology

The research questions addressed in the present study are the following:

1. Which semantic categories are the most and the least productive in foreign language learners of English and Spanish?
2. Is lexical output comparable in English and Spanish?
3. Is lexical output (much) higher in students with more years of studying a foreign language?
4. Are the most available words (top ten) in English and Spanish similar?
5. Are the semantic prototypes the same in English and Spanish?

8.3.1 Sample

This study was carried out at the Faculty of Arts of the University of Ljubljana, Slovenia. A lexical availability test was administered to 40 Slovene students. The sample was equally distributed into learners of English and learners of Spanish as a foreign language: 20 English and 20 Spanish. Out of the 20 learners of English, 10 were in their first year of study and 10 were in their second year; their ages ranged from 20 to 24 years old, 2 were males and 18 females. All of them had been studying English for at least 9 years; in addition to English, they also spoke other foreign languages such as German, Spanish, French, Italian, Croatian and Japanese. As to the group of learners of Spanish, they were all attending the last year of a Spanish language and literature programme, 16 had been studying Spanish for 4 years at university and 4 had already studied Spanish for 4 years at secondary school, making a total of 8 years of learning Spanish. For these 4 students, Spanish was the second FL, whereas for the other 16 students it was the third FL. For all 20 Spanish students, English was the first FL, followed by other foreign languages: German (10 students), Italian (7), Portuguese (3), French (2), Russian (2), Chinese (1), Croatian (1) and Japanese (1). All the students were (22–30 years old) females.

8.3.2 Data Collection and Procedures

Taking as reference the traditional LA task used in Spanish lexical availability studies, eight semantic categories (or centres of interest) were chosen for both groups of learners, these were: 'Parts of the body'(*Partes del cuerpo*), 'House'(*La casa*), 'Food and drink'(*Comida y bebida*), 'School'(*La escuela*), 'City'(*La ciudad*), 'Countryside'(*El campo*), 'Animals'(*Animales*), 'Games and Entertainment'(*Juegos y distracciones*).

Students were allowed 2 min to write down their word responses for each semantic category. The lexical statistical programme Dispolex (Bartol Hernández and Hernández Muñoz 2003) was then used for data processing and calculation.

8.4 Results of Quantitative Analysis

8.4.1 Lexical Availability in English FL

As shown in Table 8.1, among the 20 Slovene English FL learners, the highest lexical output was produced in the category 'Animals'. In comparison, the other categories had a much lower output, the lowest being 'Games and Entertainment' and 'School'. Regarding lexical diversity, we note that three categories, 'Country', 'Animals' and 'City', are the categories that elicit the highest number of different items. In comparison, 'Parts of the body' triggered very similar word responses. This fact explains the cohesion index found in 'Parts of the body': higher and, as a result, much more compact than in the other semantic categories. In contrast, differences in the cohesion index are smaller in the other semantic categories. The lowest cohesion index is found in 'Country' and 'City', probably due to the nature of these categories; open categories are capable of bringing up a greater diversity of word responses, and therefore, a greater diversity in word associations.

The overall production for each category runs parallel with the means obtained by the students. Table 8.1 indicates that the highest mean of words per student falls into the category 'Animals', whereas the lowest is found in 'Games and Entertainment'.

Table 8.2 summarizes the results obtained by two groups of EFL learners with different years of exposure to the target language. Compared to first year students, EFL learners in their second year of university produced a higher number of word responses in the five semantic categories. However, both groups obtained equal results in one category, and even more: first year students outperformed second year students in two semantic categories. The highest difference between the two groups was found in the category 'School', where, on average, the second year students wrote 3.6 words more than the first year students on average.

Table 8.1 Overall results of lexical availability in EFL learners

Semantic category	Total	Types	Mean of items per student	Cohesion index
'Parts of the body'	379	72	18.95	0.26
'House'	363	127	18.15	0.14
'Food and drink'	360	144	18.00	0.13
'School'	326	138	16.30	0.12
'City'	345	156	17.25	0.11
'Country'	329	167	16.45	0.10
'Animals'	406	163	20.30	0.12
'Games and Entertainment'	317	135	15.85	0.12

Table 8.2 Comparison of means for 1st and 2nd university students EFL learners

Semantic category	Mean of items per student	
	1st year students	2nd year students
'Parts of the body'	19.0	18.9
'House'	16.4	19.5
'Food and drink'	18.0	18.0
'School: Furniture and Material'	14.5	18.1
'City'	16.5	18.0
'Country'	15.5	17.4
'Animals'	20.2	20.4
'Games and Entertainment'	16.1	15.6
Average	17.02	18.24

Table 8.3 Overall results of lexical availability in Spanish FL

Semantic category	Tokens	Types	Mean of items per student	Cohesion index
'Parts of the body'	337	73	16.85	0.23
'House'	314	143	15.70	0.11
'Food and drink'	385	170	19.25	0.11
'School'	241	118	12.05	0.10
'City'	339	181	16.95	0.09
'Countryside'	327	168	16.35	0.10
'Animals'	357	153	17.85	0.12
'Games and Entertainment'	272	141	13.60	0.10

8.4.2 Lexical Availability in Spanish FL

As shown in Table 8.3, the Slovene university student learners of Spanish FL produced the highest lexical output in the category *Alimentos y bebidas* ('Food and drink'), followed by *Los animales* ('Animals'). The lowest lexical output was in *Juegos y distracciones* ('Games and Entertainment') and *La escuela: muebles y materiales* ('School: Furniture and Material'). As to the different lexical items, the category *La ciudad* ('City') shows the highest diversity, followed by *Alimentos y bebidas* ('Food and drink') and *El campo* ('Countryside'). The lowest lexical diversity and the highest cohesion index are found in the category *Partes del cuerpo* ('Parts of the body'). Other differences in the cohesion index only range from 0.09 to 0.12. The lowest index is observed in *El campo* ('Countryside').

Regarding the distribution of means of words per student, Table 8.4 displays the means of lexical items per categories distributed into two groups of learners: those who had already studied Spanish in secondary school (8 years of learning) and those who started to study Spanish at university (4 years of learning). The 16 students in their fourth year of Spanish obtained a lower mean than the 4 students who had been

Table 8.4 Mean of items per student in Spanish FL

Semantic category	Mean of items per student	
	4 years of Spanish	8 years of Spanish
'Parts of the body'	16.6	18.0
'House'	14.4	21.0
'Food and Drink'	18.3	23.5
'School: Furniture and Material'	11.1	16.0
'City'	15.9	21.0
'Country'	15.1	21.3
'Animals'	17.0	21.3
'Games and Entertainment'	12.8	17.0
Average	15.15	19.89

studying Spanish for 8 years. The latter group is superior in all eight semantic categories, with the highest difference (about six words of difference) shown in the categories 'House' and 'Country' and the smallest in 'Parts of the body'.

8.4.3 Comparison of Lexical Availability in English and Spanish as FL

Both groups, EFL and SFL learners, were very productive in the semantic category 'Animals'. However, differences were observed concerning two other categories: the former was more productive in 'Parts of the body', whereas the latter was in 'Food and drink'. The two groups behave similarly regarding the least productive semantic categories: 'Games and Entertainment' and 'School'. These results corroborate the ones obtained by Samper Hernández (2002) on the lexical availability of learners of Spanish in an immersion context.

The comparison of the number of different lexical items shows no similarities regarding the semantic categories with the highest lexical diversity; however, it does prove that, in both groups, 'Parts of the body' concentrated the smallest range of diverse vocabulary. As a semantic category, this is much more compact than 'Countryside' or 'City'. The lexical diversity in LA tests therefore depends a great deal on the specific semantic category.

For each semantic category, the average number of words is slightly higher among EFL learners than among SFL learners. Nevertheless, both groups produced a greater number of words than other groups of SFL and EFL learners as for instance: Finnish students SFL learners (Carcedo González 2000a), international students at Salamanca (several nationalities) (Samper Hernández 2002, 2003), Chilean EFL learners (Germany and Cartes 2000). Closest to our results is the mean of total lexical output reported for the bilingual school in Chile (English and Spanish), ('Body' 19.65, 'House' 12.65, 'Food' 17.4).

In spite of the high mean of words retrieved per student, the results of the Slovene students of English and Spanish are still inferior to the average results of native speakers, which are usually over 20 words per student (Samper Hernández 2002). Advanced students of Spanish (8 years) wrote the highest number of words per student (20 words) and category, followed by the advanced students of English (second year at university) with an average of 18 words per student, then the students of English from the first year of a university programme with 17 words, and finally the students of Spanish with 4 years of study with 15 words. As a conclusion, I would like to refer to Norbert Schmitt's (2000: 42) observation: "It seems that native-like association behavior, and by implication native-like lexical organization, is something that is not easy to acquire."

8.5 Results of the Qualitative Analysis

A comparison of the top ten available words in English FL and Spanish FL shows whether the two groups of FL learners produced similar or different associations in the same eight semantic categories, and whether we can observe any common semantic prototypes. The semantic prototype is the most ideal, or the best representative example in a given category (Kleiber 1995). The top ten words were chosen for the qualitative analysis because the most available words show collective results, while the least available words show individual results.

The English and Spanish words are displayed in Tables 8.5, 8.6, 8.7, 8.8, 8.9, 8.10, 8.11, 8.12, and 8.13 in the appendix, together with their index of availability. The Spanish words have been translated into English. The shared words on both lists are presented in bold and their number is written next to the semantic category in the first row.

In the semantic category 'Parts of the body', seven words are the same in both languages: *head*, *leg*, *eye*, *finger*, *nose*, *ear*, *mouth*. The most available word (the highest availability index) in the English group is *head*. The same word is in second place in the Spanish group, but with a slightly lower index. The word *head* can probably be defined as a semantic prototype of 'Parts of the body'. Other indexes are very similar in both language groups. The three words that do not coincide in the list of the top ten available words appear soon after the tenth place.

As to the semantic category 'House', again, seven words concur in both languages: *kitchen*, *bathroom*, *garden*, *window*, *door*, *bed*, *room*. 'Kitchen' is the first and the most common association, with a very similar index in both language groups; therefore, it may be considered as the best representative or semantic prototype of the category 'House'.

Regarding 'Food and drink', there are five words in common in both languages: *juice*, *meat*, *vegetable*, *milk*, *water* or actually six words because the Spanish list includes two expressions for juice: *zumo* and *jugo*. The most available word with the highest index is *water* (*agua*) in the Spanish group, while the same word has a much lower index in the English group. Due to this disparity, we cannot speak

of a shared semantic prototype for this category. However, one characteristic that is common to both lists is that the students included drinks as hyponyms (*juice, milk, beer, water, coffee*) and superordinates (*meat, vegetable, fruit*) as the most available words.

In the semantic category 'School: Furniture and Materials', there are six words shared by the two groups of FL learners: *blackboard*, *chair*, *table*, *teacher*, *book*, and *window*. The first three words are identical in both groups and all show a high index. Although the subtitle of the semantic category refers to objects ('Furniture and materials'), the students nevertheless wrote the word teacher (*professor*) in strong association with 'School'.

Concerning the semantic category 'City', there are, again, seven words in common in both languages: *car*, *people*, *bus*, *shop*, *park*, *traffic* and *street*. The indexes are not alike in these common words, with the highest index being attributed to the word *car* in English, while *coche* in Spanish has a much lower index. Given that this semantic category is very open and, hence, gives rise to many different associations, the similarity of the most available words in English and Spanish is surprisingly high.

In the semantic category 'Country', six words appear in both languages: *green*, *cow*, *peace*, *animal*, *nature*, *house*. 'Country' is the only category where an adjective is found within the most available words, both in English (*green*) and Spanish (*verde*). This word can also be defined as a semantic prototype within the category under examination. There are actually seven words in common if we count *farm* and *farmer* (*campesino* in Spanish). Some words display a very similar index in both language groups: *cow*, *nature* or *house*. The prototype *green* in this category is an example of how associations and prototypes depend on the specific geographical context. In her prototype theory, Aitchison (1994) mentions a "mental model" that is based on the combination of several factors such as observation and experience, cultural influence, memory and imagination. These factors definitively do have an influence. Different cultures have different prototypes.

If we focus on the semantic category 'Animals', we also find six words shared by both languages: *dog*, *cat*, *cow*, *bird*, *elephant*, *horse*. The semantic prototypes in this category are *dog* and *cat*, both with a very high index in English and Spanish. Domestic animals prevail, but there are also many exotic or non-domestic animals within the top ten in both groups. The English list includes *bull* among top ten, while on the Spanish list *toro* appears in 12th place.

In comparison, if we look at the semantic category 'Games and Entertainment', we observe that there are only four words shared by both languages: *sport*, *reading*, *basketball*, *football*. This semantic category shows the fewest similarities in the word responses retrieved by the two language groups. No semantic prototype is observed here, only a tendency to associate games and entertainment with different sports.

Most available words are nouns in all the semantic categories. However, some slight differences are noted that have to do with the specific semantic categories and languages. For instance, as far as 'Food and drink' is concerned, the first ten available words in English are nouns, while in Spanish we find eight nouns and

two verbs: *comer*, *cocinar* (eat, cook). Regarding the category 'School', we find one verb in the Spanish list: *estudiar* (*study*) but none in the English list. In 'Games and Entertainment', two verbs were retrieved in English in the form of gerunds: *reading*, *swimming*, and three verbs in Spanish: *jugar*, *correr*, *leer* (*play*, *run*, *read*). The only adjective in the learners' lexical availability output is found in the category 'Country' in both languages (*verde*, green) together with a nominal phrase (*fresh air*) in the English group. This tendency suggests that word responses depend on the part of speech the semantic category belongs to. The stimulus word seems to trigger associations within the same part of speech. In a study conducted by Šifrar Kalan (2011) using a new semantic category entitled 'Actions Carried Out Every Day' (*Acciones que se realizan todos los días*), 72 % of the available words were verbs.

Table 8.13 presents the ten most available words produced by students of English FL and Spanish FL on the basis of eight semantic categories. However, the words with the highest availability index are not from all of the eight categories but only from four categories in Spanish and five in English, where the indexes are the highest: 'Parts of the body', 'Animals', 'School: Furniture and Materials', 'House', 'City'. The two groups of foreign language learners produced a very similar list of the most available words; eight words (presented in bold in Table 8.13) out of ten are the same and their respective indexes are very similar. The words with the highest index over (0.7) are in the English group *head*, *dog* and *cat*). The same words are also the most available in the Spanish group but their indexes are a little lower. The top ten available words in both groups are all nouns, which is in accordance with the word class the semantic categories belong to, and also with the prevalence of nouns in language in general (Pastora Herrero 1990). These words are usually learnt in beginner FL courses, so the students of the present study, whose level of foreign language is at least Level B according to the European Framework, show that their most available words are from Level A. At the same time, these words are prototypes, such as *dog* and *cat* for animals, *head* for body, *chair* for furniture, *kitchen* for house, and *blackboard* for school.

8.6 Conclusion

The study of lexical availability of eight semantic categories in 20 Slovene students of English FL and 20 students of Spanish FL at the University of Ljubljana has demonstrated a lot of similarities in the two groups under examination. The most productive semantic category in English FL was 'Animals'. In comparison, this category was in second position in the Spanish group, for whom the most productive category was 'Food and drink'. Both groups also coincide in the least productive categories: 'Games and Entertainment' and 'School: Furniture and Material' as well as in the category with the lowest lexical diversity: 'Parts of the body'. These results are similar to the ones reported by Germany and Cartes (2000) with Chilean EFL learners. The cohesion index for this semantic category

was the highest in both cases. In the light of these findings, it is possible to conclude that lexical diversity in LA tests depends a great deal on the specific semantic category.

Both in the Spanish and in the English group, advanced learners are the ones who obtain the highest means in word responses, and within these groups, the highest means are achieved by the groups with higher exposure to the language. However, they still fall behind native speakers in their potential to retrieve words out of prompts related to specific domains, what seems to suggest that lexical competence is closely related to language proficiency. A similar conclusion was reached by van Ginkel and van der Linden: "[…] there is a correlation between the proficiency of the subjects and the number of association responses that they produce" (quoted in Schmitt 2000: 42). This can be further connected with the organisation of the mental lexicon: "On the assumption that a greater number of responses indicates more words connected to the stimulus word in the lexicon, this also suggests a greater level of organization" (Schmitt 2000: 42).

The comparison of the top ten available words for the eight semantic categories under study in English FL and Spanish FL shows a surprisingly high degree of similarity as shown by the great number of shared words in most semantic categories: seven words in 'Parts of the body', 'House' and 'City', and six in 'Food and drink', 'School: Furniture and Materials', 'Country' and 'Animals'. The lowest similarity was found in 'Games and Entertainment'. Similarities were not only found in the number of shared words but also regarding word classes. As it is typical in lexical availability research, in our study, learners' most available words were nouns, with only a few word responses being verbs and adjectives. Among the top ten available words with the highest index, regardless of the semantic category, there are only nouns, with eight words out of ten being the same in both languages.

There is also a great deal of agreement among our sample of SFL and EFL students concerning the best representatives of each semantic category. The common semantic prototypes that stand out in the most available words in English FL and Spanish FL in the present study were: *head* for 'Parts of the body'; *kitchen* for 'House'; *blackboard*, *chair*, *table* and *book* for 'School: Furniture and Materials'; *green* and *farm* or *farmer* for 'Country'; and *dog* and *cat* for 'Animals'. 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). Nevertheless, it is necessary to be added that people have uniform ideas of prototypes or "best examples" only within a certain culture. A good example of cultural or environmental influence is the prototype 'green' for 'Country' in the present research, which was conducted in Slovenia, a country surrounded by mountains.

The great consistency found in the word responses retrieved by the two groups, Slovene university students learners of English FL and Spanish FL, suggests that these groups make similar connections between words and that mental lexicons are very alike in different foreign languages.

Appendix

Table 8.5 Top ten available words: 'Parts of the body'

'Parts of the body' (7)			
English FL	Availability index	Spanish FL	Availability index
1. ***head***	0.78195	*mano* (hand)	0.66900
2. ***leg***	0.63450	***cabeza*** (head)	0.64026
3. *arm*	0.60822	***ojo*** (eye)	0.59931
4. ***eye***	0.59596	***pierna*** (leg)	0.58263
5. ***finger***	0.50792	***dedo*** (finger)	0.53270
6. ***nose***	0.44517	***nariz*** (nose)	0.47264
7. ***ear***	0.44402	***oreja*** (ear)	0.43169
8. *hair*	0.34888	*cuello* (neck)	0.36383
9. ***mouth***	0.32607	*uña* (nail)	0.34085
10. *knee*	0.30164	***boca*** (mouth)	0.33954

Table 8.6 Top ten available words: 'House'

'House' (7)			
English FL	Availability index	Spanish FL	Availability index
1. ***kitchen***	0.56616	***cocina*** (kitchen)	0.55364
2. *living room*	0.49632	***ventana*** (window)	0.48765
3. ***bathroom***	0.44058	***puerta*** (door)	0.41494
4. ***garden***	0.41691	***cama*** (bed)	0.40426
5. ***window***	0.39124	***cuarto de baño*** (bathroom)	0.30845
6. *roof*	0.34165	*familia* (family)	0.29841
7. ***door***	0.29833	***habitación*** (room)	0.27868
8. ***bed***	0.28303	*baño* (bath/bathroom)	0.26287
9. *home*	0.27791	***jardín*** (garden)	0.25350
10. ***room***	0.27519	*silla* (chair)	0.24239

Table 8.7 Top ten available words: 'Food and drink'

'Food and drink' (5–6)			
English FL	Availability index	Spanish FL	Availability index
1. ***juice***	0.37017	***agua*** (water)	0.52929
2. ***meat***	0.30507	*fruta* (fruit)	0.35818
3. *pizza*	0.29020	*pan* (bread)	0.34466
4. ***vegetable***	0.27344	***verdura*** (vegetable)	0.33287
5. *potato*	0.26952	***carne*** (meat)	0.31800
6. ***milk***	0.25165	***zumo*** (juice)	0.31119
7. *beer*	0.25119	***leche*** (milk)	0.23684
8. ***water***	0.24361	*comer* (to eat)	0.20935
9. *cheese*	0.21604	*cocinar* (to cook)	0.19859
10. *coffee*	0.20746	***jugo*** (juice)	0.19642

Table 8.8 Top ten words: 'School: Furniture and Materials'

'School: Furniture and Materials' (6)			
English FL	Availability index	Spanish FL	Availability index
1. ***blackboard***	0.66731	***silla*** (chair)	0.62404
2. ***chair***	0.57000	***pizarra*** (blackboard)	0.60567
3. ***table***	0.52947	***mesa*** (table)	0.58300
4. ***teacher***	0.50947	***libro*** (book)	0.34057
5. ***book***	0.31430	*lápiz* (pencil)	0.30148
6. *chalk*	0.30418	***ventana*** (window)	0.25163
7. *television*	0.20567	*estudiar* (study)	0.16164
8. ***window***	0.19206	***profesor*** (teacher)	0.16008
9. *desk*	0.18716	*bolígrafo* (pen)	0.15665
10. *classroom*	0.18509	*papel* (paper)	0.14737

Table 8.9 Top ten words: 'City'

'City' (7)			
English FL	Availability index	Spanish FL	Availability index
1. ***car***	0.55144	***calle*** (street)	0.50101
2. ***people***	0.49825	***parque*** (park)	0.32936
3. ***bus***	0.36819	**coche** (car)	0.32034
4. ***shop***	0.26856	***tráfico*** (traffic)	0.27921
5. ***park***	0.26143	***gente*** (people)	0.27339
6. ***traffic***	0.25480	*casa* (house)	0.24100
7. ***street***	0.25065	***tienda*** (shop)	0.23037
8. *crowd*	0.24581	***autobús*** (bus)	0.20035
9. *restaurant*	0.19011	*ruido* (noise)	0.20019
10. *building*	0.18581	*metro* (metro)	0.16598

Table 8.10 Top ten words: 'Country'

'Country' (6)			
English FL	Availability index	Spanish FL	Availability index
1. *farm*	0.38895	***verde*** (green)	0.42650
2. ***green***	0.36451	***animal*** (animal)	0.41859
3. *grass*	0.35830	*campesino* (farmer)	0.39583
4. ***cow***	0.31915	***vaca*** (cow)	0.30564
5. ***peace***	0.29697	*árbol* (tree)	0.29598
6. ***animal***	0.27866	***naturaleza*** (nature)	0.25677
7. *forest*	0.25551	*sol* (sun)	0.24921
8. ***nature***	0.25232	*pueblo* (village)	0.21892
9. ***house***	0.24538	***casa*** (house)	0.21573
10. *fresh air*	0.23928	***tranquilidad*** (peace)	0.17013

Table 8.11 Top ten available words: 'Animals'

'Animals' (6)			
English FL	Availability index	Spanish FL	Availability index
1. ***dog***	0.75681	***perro*** (dog)	0.69874
2. ***cat***	0.71201	***gato*** (cat)	0.64531
3. *cow*	0.44352	***caballo*** (horse)	0.49899
4. *tiger*	0.36080	*vaca* (cow)	0.49618
5. *mouse*	0.30911	***pájaro*** (bird)	0.34291
6. ***bird***	0.29376	*oveja* (sheep)	0.31504
7. ***elephant***	0.28629	***elefante*** (elephant)	0.30590
8. *snake*	0.26424	*cerdo* (pig)	0.20134
9. *horse*	0.22176	*león* (lion)	0.20254
10. *bull*	0.21673	*pez* (fish)	0.19369

Table 8.12 Top ten available words: 'Games and Entertainment'

'Games and Entertainment'			
English FL	Availability index	Spanish FL	Availability index
1. *television*	0.35553	*jugar* (play)	0.50105
2. *sport*	0.32146	*amigo* (friend)	0.31061
3. *reading*	0.28014	*fútbol* (football)	0.27261
4. *volleyball*	0.27756	*deporte* (sport)	0.27053
5. *swimming*	0.27032	*ajedrez* (chess)	0.23233
6. *basketball*	0.24389	*cartas* (cards)	0.22268
7. *tennis*	0.21234	*baloncesto* (basketball)	0.20434
8. *dance*	0.21127	*cine* (cinema)	0.18550
9. *monopoly*	0.21023	*correr* (run)	0.15593
10. *football*	0.19110	*leer* (read)	0.15554

Table 8.13 Top ten available words

English FL	Availability index	Spanish FL	Availability index
1. *head*	0.78195	*perro* (dog)	0.69874
2. *dog*	0.75681	*mano* (hand)	0.66900
3. *cat*	0.71201	*gato* (cat)	0.64531
4. *blackboard*	0.66731	*cabeza* (head)	0.64026
5. *leg*	0.63450	*silla* (chair)	0.62404
6. *arm*	0.60822	*pizarra* (blackboard)	0.60567
7. *eye*	0.59596	*ojo* (eye)	0.59931
8. *chair*	0.57000	*mesa* (table)	0.58300
9. *kitchen*	0.56616	*pierna* (leg)	0.58263
10. *car*	0.55144	*cocina* (kitchen)	0.55364

References

Aitchison, J. 1994. *Words in the mind: An introduction to the Mental Lexicon*. Oxford: Blackwell.

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

Bartol Hernández, J.A., and N. Hernández Muñoz. 2003. *Dispolex: Base de datos de la disponibilidad léxica*. Salamanca: Universidad de Salamanca. http://www.dispolex.com. Accessed 4 Dec 2011.

Carcedo González, A. 2000a. Disponibilidad léxica en español LE: análisis de cuatro fases de disponibilidad. *Pragmalingüística* 5–6: 75–94.

Dimitrijević, N. 1969. *Lexical availability*. Heidelberg: Julius Gross Verlag.

Fernandez Fontecha, A. 2010. Gender and motivation in EFL vocabulary production. In *Gender perspectives on vocabulary in foreign and second languages*, ed. Jiménez Catalán, 93–117. New York: Palgrave Macmillan.

Ferreira, R., and M.S. Echeverría. 2010. Redes semánticas en el léxico disponible de inglés L1 e inglés LE. *Onomázein* 21(1): 133–153.

Germany, G.P., and E.N. Cartes. 2000. Léxico disponible en inglés como segunda lengua en instrucción formalizada. *Estudios pedagógicos* 26: 39–50.

Gougenheim, G., R. Michéa, P. Rivenc, and A. Sauvageot. 1967. *L'élaboration du français fondamental 1er degré. Étude sur l'élaboration d'un vocabulaire et d'un grammaire de base*, Nouvelle édition refondue et augmentée. Paris: Didier.

Jiménez-Catalán, R.M. 2010. Gender tendencies in EFL across vocabulary tests. In *Gender perspectives on vocabulary in foreign and second languages*, ed. Jiménez Catalán, 117–139. New York: Palgrave Macmillan.

Jiménez-Catalán, R.M., and J. Ojeda Alba. 2009. Girls' and boys' lexical availability in EFL. *ITL International Journal of Applied Linguistics* 158: 57–76.

Kalan, Šifrar. 2009. Disponibilidad léxica en español como lengua extranjera: el cotejo de las investigaciones en Eslovenia, Salamanca y Filandia. Verba hispánica: Anuario del Departamento de la Lengua y la Literatura Españolas de la Facultad de Filosofía y Letras de la Universidad de Ljubljana.

Kleiber, G. [1990] 1995. *La semántica de los prototipos. Categoría y sentido léxico*. Madrid: Visor.

Pastora Herrero, J.F. 1990. *El vocabulario como agente de aprendizaje*. Madrid: La Muralla.

Samper-Hernández, M. 2002. *Disponibilidad léxica en alumnos de español como lengua extranjera*, Colección Monografías nº 4. Málaga: ASELE.

Samper-Hernández, M. 2003. Datos comparativos entre el léxico disponible de estudiantes de español como lengua extranjera y el de la variedad grancanaria. In *Estudios sobre el español de Canarias*, ed. C. Díaz Alayón, M. Morera, and G. Ortega, 961–987. Islas Canarias: Academia Canaria de la lengua.

Samper Padilla, J.A. 1998. Criterios de edición del léxico disponible: sugerencias. *Lingüística* 10: 311–333. University Press.

Schmitt, N. 2000. *Vocabulary in language teaching*. Cambridge: Cambridge University Press.

Šifrar Kalan, M. 2011. Leksikalna razpoložljivost v španščini kot tujem jeziku. Unpublished doctoral thesis, Univerza v Ljubljani, Ljubljana.

Chapter 9
The Effect of Instruction on Polish Spanish Learners' Lexical Availability

Antonio María López González

9.1 Introduction: The Educational Context of the Research

The Spanish Bilingual Sections programme in Poland is based on the May 1997 agreement between the Ministry of National Education and Sport for the Republic of Poland and the Ministry of National Education and Science for Spain, and the later 2005 and 2010 appendices, as its legal basis. These relate to the creation and running of Spanish Bilingual Sections in middle and high schools in the Republic of Poland. To be admitted onto the programme, candidates have to perform linguistic ability tests, which examine general understanding and command of the language. The selected students then follow a programme reinforced with classes in Spanish. At the end of this programme, in addition to getting the Polish baccalaureate certificate, they have the option to obtain the Spanish baccalaureate certificate, after passing the corresponding exams and complying with the requirements necessary to issue said documentation.

The Spanish Bilingual Sections programme is divided into two educational stages. The first stage, during which students devote 630 h of classes to studying Spanish as a Foreign Language (SFL), is completed during Middle school or at the beginning of high school. The second stage is a specific programme consisting of "Spanish Language and Literature" and "Spanish History and Geography". Students dedicate approximately 1,100 h to classes, which are given in Spanish, over the three baccalaureate years.

The Bilingual Sections model in Poland implements what is usually referred to as "Content and Language Integrated Learning" (CLIL), as part of the 2004–2006 European Commission Action Plan to promote language learning and linguistic diversity in the European Union. Through this type of teaching, students learn

A.M. López González (✉)
Associate professor of Spanish linguistics, Katedra Filologii Hiszpanskiej, University of Lódz, Lódz, Poland
e-mail: antoniomlg@hotmail.com

R.M. Jiménez Catalán (ed.), *Lexical Availability in English and Spanish as a Second Language*, Educational Linguistics 17, DOI 10.1007/978-94-007-7158-1_9,

curriculum contents whilst exercising and perfecting their linguistic competences, combining contents and language. CLIL emerges "with the aim of better preparing students for life in a Europe in which mobility is becoming increasingly widespread and should be within reach of everyone" (Eurydice 2006: 3).

With regards to learning a foreign language, CLIL objectives in Poland are threefold: (i) promoting the values of tolerance and respect towards other cultures through use of the CLIL target language (socio-cultural objectives); (ii) developing linguistic abilities with an emphasis on efficient communication, motivating students to learn languages by using them for real, practical purposes (linguistic objectives); (iii) developing subject-related knowledge and learning ability, stimulating the assimilation of content by means of different and innovative methods (educational objectives) (Eurydice 2006: 22).

As to the educational objectives, the idea of an innovative methodological approach is intrinsic to CLIL. Both the language and the non-linguistic content constitute teaching tools, with neither one being predominant over the other. "Achieving this twofold aim calls for the development of a special approach to teaching in that the non-language subject is not taught in a foreign language but with and through a foreign language (Eurydice 2006: 7)". Arnau (2001) highlights the following characteristics of this innovative approach: (1) language is learnt within a contextualised use; (2) the learner employs the language in a meaningful way, using it to communicate himself or herself; (3) the student learns forms whilst using them, and uses them whilst learning them.

Therefore, the CLIL methodology responds to the need for students to be exposed to situations that require authentic communication, because "learning a language is learning to communicate oneself" (Ellis 1992). This point is related to the CLIL linguistic objectives' concept of *effective communication,* which promotes the teaching of a foreign language by means of a communicative approach.

Regarding CLIL socio-cultural objectives, it should be highlighted that in the Spanish Bilingual Sections model in Poland, language is strongly related to culture. In addition to achieving high-level linguistic capabilities, a great amount of knowledge about Spain is also acquired in bilingual classes. Therefore, language is not only a tool for communication, but also, and most importantly, an instrument to relate and convey culture (Tatoj et al. 2008).

9.1.1 First Stage: Year 0 vs. Middle School (Gimnazjum)

Taking the CLIL approach in the bilingual programme into consideration, it is imperative that the students gain a level of linguistic competence in Spanish in order to tackle the non linguistic subjects with success. For this reason, when defining the linguistic competence objectives and contents at different stages of the bilingual programme, each stage of the Bilingual Sections curriculum is related to the Council of Europe's Common European Framework of Reference for Languages (CEFR) scale. Furthermore, basic functional, grammatical and socio-cultural contents required at different stages of the Cervantes Institute Spanish Diploma qualifications are also

Table 9.1 Comparison of curriculum levels to Spanish sections

Curriculum of Spanish bilingual sections	Common European Framework of Reference	D.E.L.E.	Age range
First stage	B1	Initial (++)	16 13–15
Years: 1, 2, 3 High school and Matura	B1+, B2, B2+	Intermediate	16–19

integrated within the programme. This means students should reach the level of 'independent user' (B1) by the end of the first stage, and have achieved the level of B2+ by the end of the bilingual programme, approaching the stage classified by the CEFR as 'competent user' (Table 9.1).

Within this context of language learning, an emphasis is placed on the first stage of the bilingual programme, with 630 teaching hours devoted to studying Spanish as a Foreign Language (SFL). The objective of this stage is to prepare students for learning non-language subjects in Spanish, and the language is treated as an essential vehicle for acquiring knowledge in Literature, History, Geography and Spanish Art History. This stage also gives teachers the chance to evaluate the linguistic level achieved by students in depth. It also allows them to evaluate the application of this knowledge in both language and non-language subjects during the following stages.

Poland's Spanish Bilingual Sections programme offers two learning modalities for this initial stage:

(a) Modality I: *Year 0* – language immersion, with intensive Spanish as a Foreign Language (SFL) classes, to which at least 18 h are dedicated per week. This course is given in high schools in Year 0, before the first year of the Baccalaureate.
(b) Modality II: A 3-year course of extensive SFL classes. These classes are taught in *gimnazjum* (Middle school), and the 18 weekly hours of classes seen in Modality I are shared out over the 3 years of Polish Middle school, therefore becoming six teaching hours a week.

In Year 0, (high school), classes are taught by three or four teachers, both Polish and native Spanish speakers. With regard to the 3-year Middle school course, students have one main Polish teacher, supported by a native Spanish speaker who gives 1 h of conversation classes per week and two additional hours of Introduction to Spanish Literature classes from year 2 onwards.

Concerning continuity in the programme, there is normally one class in Year 0, where students continue to follow the bilingual programme in the first year of high school, except in very special cases. In Middle school, there are normally two classes, reduced to one in high school. This reduction is on one hand due to students voluntarily dropping the subject after the 3 years of Middle school and on the other, due to a selection process based on linguistic competence in Spanish gauged by a final exam and the student's academic performance shown in their school report. Students who voluntarily leave the course mostly do so because either they choose to study non-humanities subjects or they have a purely linguistic interest in the

Table 9.2 Bilingual programme stages in the Spanish sections in Poland

First stage preparatory (Spanish as a foreign language B1): 630 h		**Second stage bilingual** (non-language subjects in Spanish): 1,100 h
Modality I: Year 0 High school: a year, age 16 **Modality II: Middle school** Years: 1, 2, 3; age 13–15		**High school:** years: 1, 2, 3; age 16–19

programme. To a lesser extent, students leave the programme as a result of the difficulties they experience in their learning of the Spanish language (Table 9.2).

Regardless of which modality is chosen, the student must complete the same SFL curriculum objectives, content, tasks, etc.). The SFL curriculum, which serves as a reference for "Year 0" and the three Middle School years, was developed by the Education Office of the Embassy of Spain in Poland (Consejería de Educación de España en Polonia 2005).

However, each method has pros and cons, and various factors (time, funding, motivation, effort, or psychological development) tend to favour one modality or the other. These factors were analysed by Tatoj et al. (2008) in the *Evaluation of Poland's Spanish Bilingual Sections Report.* Based on interviews carried out with teachers in Bilingual sections, the authors concluded that students who followed the Year 0 course showed a higher linguistic and learning level than those who had followed the bilingual Middle school programme. The report states that teachers highlighted the fact that choosing the bilingual course in high school is a more conscious decision, and as such, students are more motivated to study. According to the authors, in general, high school students who take the Year 0 course gain extra time to concentrate almost exclusively on the study of the Spanish language. Furthermore, it has been claimed students can learn the grammar more easily at the age of 16 than when they are 13 years old (Tatoj et al. 2008).

On the other hand, in their opinion, bilingual classes during the 3 years of Middle school does not adequately prepare students for the demands of a bilingual class in high school. Students have neither sufficient knowledge nor linguistic abilities to allow them to participate fully in classes given entirely in Spanish, by a teacher who does not speak Polish (Tatoj et al. 2008).

Evidently, these conclusions have been fiercely challenged by teachers and educators working in bilingual Middle schools in Poland. In defence of Middle school teaching, it has been argued that both the Council of Europe and the European Union encourage bilingual teaching from the beginning learning stages (pre-school and primary school), and that as a result of the early start and partial immersion method, these programmes have been incredibly successful in terms of language performance in Canada and America (Eurydice 2006). The official stance of Poland's Ministry of Education is to favour the homogenisation of the first stage of the Bilingual Sections around the Middle school model, and as such, is opting for the abolition of Year 0 and the establishment of collaboration ties with "satellite" Middle schools, where future students would be prepared for the second stage of the bilingual programme.

9.2 Theoretical Foundations: Lexical Availability and Evaluation of Lexical Competence

As it has been shown, it is the first stage of the bilingual programme which develops and establishes the basic linguistic abilities in Spanish necessary to tackle non-language subjects in the bilingual Baccalaureate curriculum. Lexical competence is one of these competences.

Even with the limitations of the methodology (Hernández Muñoz 2006; Higueras García 2008), a study of lexical availability is suggested as an exceptional tool to evaluate the control of fundamental vocabulary which ensures the ability of basic communication in a foreign language.

As it is well known, in addition to basic and common linguistic expressions related to our physical surroundings (*head, window, food)* or basic conceptual distinctions (*sleep, leave, enter*), the fundamental lexicon also includes other, more abstract terms expression of possibility, how close or far away the concept is from reality, etc. The fundamental lexicon includes two easily-distinguishable lexical sub-groups (Michéa 1950, 1953): (a) *Basic lexicon* – commonly- used and non-subject specific. Mainly grammatical words and words which continuously appear in any conversation or written text, regardless of the topic being discussed (e.g., *to, the, not, many, there is, give, person, put, etc.*), and (b), *Available lexicon* – topic-related, comprising specific semantic content and words that whilst commonly-used, are only employed in relation to a topic (e.g., *frying pan* and *fork* in relation to 'kitchen', or *letter* and *stamp* in relation to 'post').

It is precisely the need to select words which must be taught in foreign language classes which gave rise to the birth of lexical availability, as explained in the introductory chapter to this book (Chap. 1).

From a pedagogical perspective, the benefits of studying lexical units – *lexías*, using Pottier's terminology (1971) – must be highlighted. This helps to evaluate adequate learning of vocabulary as well as to determine the group of widely-available words which shape the active lexicon. Effectively, lexical availability is designed to evaluate the school's efficacy in its aim to educate the students in their command of the fundamental lexicon, both in L1 and L2. This evaluation has been carried out by means of the monitoring of students' development of lexical competence at specific points in the learning process.

In the Spanish-speaking world, the results of such an evaluation of the native tongue have provided researchers with different results. López Morales (1973, 1978) in San Juan de Puerto Rico, and López Chávez (1993, 1995) in Mexico measured the development of lexical availability in primary education. Both parties detected significant irregularities in the pupils' lexical competence and a lack of gradual qualitative and quantitative progression in the lexical acquisition process. However, Román-Morales (1985), in Dorado, Puerto Rico, Mena Osorio (1986) in Concepción, Chile, Echeverría (1991), in Chile, and Alba (1995), in the Dominican Republic found positive results. After conducting an analysis of lexical availability at three different stages of primary school, they all noticed a fairly regular, gradual increase

in the number of word types, as well as in the average of word responses as school grade increased.

Similarly, lexical availability in L2 also allows for the examination of different phases of the lexical learning-acquisition process of Spanish as a Foreign Language. This was done by Carcedo González (1998, 2000) in his studies devoted to studying Finnish students' lexical availability in Spanish. In his conclusions, Carcedo González (2000: 213–216), recorded a very uneven development of vocabulary in different subject areas, and a gradual evolution of lexical richness parallel to the rise in the level of study, with a qualitative leap from high school to university level.

Carcedo González's monograph (2000) looks at the lexical availability of a sample of 350 Finnish students, learners of Spanish as a Foreign Language. Without doubt, this constitutes the work of reference for any exhaustive analysis of learners' lexical availability and of the effect that extra-linguistic variables can have on it. The variables considered by Carcedo González are 'type and course grade' (4th and 8th year of high school, first and second year of university), 'gender', 'mother tongue' (Finnish and Swedish) and 'knowledge of other Romance languages'.

Following this, Samper Hernández (2002) devoted a monograph to the examination of the lexical availability of 45 students of different nationalities who attended Spanish courses at the University of Salamanca, adopting methodological guidelines similar to those of Carcedo González (2000). In her study, she found out a clear decrease in learners' lexical development once students reached the highest level. This was explained by students being poorly grouped according to their command of the Spanish language, or by the belief that the use of more complex or less common lexical units – and not only the number of words – implies a better mastery of a foreign language Samper Hernández (2002: 85–86).

Using these studies as a model, I conducted an analysis of the development of lexical availability in Polish students, learners of Spanish attending bilingual sections in Poland (López González 2010). In this study, using two identical samples of 120 students studying Spanish in Middle school and High school, I found an evident enrichment of lexical competence in Bilingual Sections students as they advanced in their studies, both in total words (+27.8 %) – quantitative – and in different words (+49.6 %) – qualitative – together with the existence of a solid common base in both educational levels with regard to easily-available vocabulary.

Lexical availability studies therefore allow for the identification and understanding of the vocabulary which is actually available to a group of language learners. As such, they become an instrument for the evaluation of lexical competence in a foreign language as well as for the study of the effect of educational methods on the development of lexical knowledge.

Germany and Cartes (2000) carried out a study to determine the effect of the factor 'type of educational setting bilingual, (private, state school) on the lexical availability of learners of English as a foreign language in Chile. Using a sample of 60 students in the first year of Middle school, and working with three cue words, 'Body', 'Food' and 'House', they found out that the teaching methodology used in each institution proved to be decisive. Students in the bilingual educational proved to have a higher degree of lexical availability than students in other educational institutions. This was so because they used the target language as a means of

communication in 80 % of the core subjects in the curriculum. Behind the bilingual school, it was the private school, in which English was taught by way of a communicative methodology based on functions of the language. The state school appeared in last position, with lower results due to vocabulary being taught out of a communicative context and following a traditional programme based on the teaching of grammar rather than on a communicative approach.

With regard to Spanish as a Foreign Language, the study by Higueras García (2008) is outstanding. This author looks at the lexical availability of 43 adult students learners of Spanish as a Foreign Language in the metropolitan zone of Madrid, in six semantic categories ('Body parts', 'Clothing', 'Food and drinks', 'The kitchen and its utensils', 'Games and entertainment', and 'City'). For this purpose, initially, she follows Carcedo González's (2000) and Samper Hernández's (2002) methodology, bearing in mind the extralinguistic factors of 'sex', 'age', 'sociocultural level', 'mother langue' and 'knowledge of other languages'. However, given the main characteristic of the group – Intermediate level (B1) students from two Official Language Schools in Madrid, in a programme of immersion in the Spanish language and culture – this author includes two new variables: the 'Teaching-learning method' (regulated methods versus non regulated methods – both methods) and 'Years of study of the Spanish language' (with intervals of a year). Regulated methods include universities, official language schools, or private language academies; non regulated methods are non systematic methods and self-taught learning.

In the results, "the variable 'Years of study of the Spanish language' has a significant impact […], to such a degree that a general directly proportional relationship between the years of study of the Spanish language and a larger number of words provided by informant can be noted" (Higueras García 2008: 202), with "a consequent upward trend between the average number of responses given by the informant and the number of years that the informant has learnt Spanish", up to 3 years of study (Higueras García 2008: 205). In the variable 'Teaching-learning method', in every semantic category "the highest quantitative rates […] were provided by those students who had not followed any official system of learning and teaching Spanish", noting a slight superiority in the informants who had combined both methods regulated and non regulated) (Higueras García 2008: 203).

9.3 Research Objectives

As seen above, at the end of the first bilingual stage, students are required to have a good command of the Spanish language, reaching B1 level of the CEFR at least. There are two modalities for the first stage, and a controversy, when it comes to designating which modality should prevail in the model of bilingual teaching. However, the preference for one or the other is based more on subjective notions than on objective data.

Given this state of affairs, the present study aims to ascertain which of these modalities obtains better results. With this purpose in mind I set out to achieve the following objectives: (1) to provide objective data, based on lexical availability tests

and on a lexical-statistical study; (2) to describe quantitatively and qualitatively the available lexicon of learners of Spanish in the two instructional programs in the 16 sampled semantic categories; (3) to compare quantitatively and qualitatively the available lexicons from Middle school and Year Zero; (4) to analyse the influence of the type of school on the lexical availability of two instructional programmes; (5) to determine the structure of the most available lexicon (active vocabulary), distinguishing both common and exclusive lexicon for each type of instructional programme; (6) to analyse the composition of learners' active vocabulary, according to the level of difficulty of the lexical units.

9.4 Methodology

This study adopted the methodological steps followed in previous research in Spanish as L2. We also adopted a quantitative and qualitative approach to the study of lexical competence in representative samples of students in the two instructional programmes within the initial bilingual stage. As such, it follows the methodological guidelines of the PanHispanic Project on lexical availability, supervised by López Morales. Material was gathered by way of a written semantic fluency task in which the informants have to produce all the words that they come to their minds about a specific topic – also known as centre of interest or semantic category- for 2 min. The number of categories in the test was up to 16 categories. These categories were: (1) 'Parts of the human body', (2) 'Clothing', (3) 'Parts of the house', (4) 'House furniture', (5) 'Food and drink', (6) 'Objects on the table for the meal', (7) 'The kitchen and its utensils', (8) 'School furniture and materials', (9) 'Heating and lighting', (10) 'The city', (11) 'The 'Countryside', (12) 'Means of transport', (13) 'Farm and Garden Work', (14) 'Animals', (15) 'Games and entertainment', (16) 'Jobs and professions'.

The study was carried out in six bilingual Polish schools: Poznan, Lublin, Wroclaw (Middle school); Bydgoszcz, Warsaw-Cervantes, Warsaw-Marti (Year 0). The tests were administered at the end of the school years 2005/06 and 2006/07 in the case of the Middle school and 2010/11 in the case of Year 0. An additional group of 30 tests is scheduled to be done in the Lodz Bilingual Section for the year 2012. That will make the sample for Year 0 up to 120 participants. The distribution of the sample under studied is shown in Table 9.3 as follows:

Table 9.3 Sample distribution

	Middle school			Year Zero			Total
City	Lublin	Poznan	Wroclaw	Bydgoszcz	Warsaw Cervantes	Warsaw Marti	–
City total	43	43	34	33	27	30	210
Level total	120			90			210

In relation to the terminology used, it is important to point out that when speaking of words, in practice it is *lexical units* that are being dealt with, which can consist of more than one word. An example is found in the answers produced by informants in respond to prompts in lexical availability tasks (e.g., *flat plate, pull up weeds, mow the lawn*, etc.), an observation that had already been made by Dimitrijévic (1969).

In a statistical study of lexical availability as the present one it is important to distinguish between token and type. *Token* refers to all the informants' computable word occurrences, whereas *type* refers to each different lexical unit.

In line with lexical availability studies in Spanish L1 and L2, we followed the editing criteria suggested in Samper Padilla (1998). The data was electronically processed and recorded on the website Dispolex.com, which provided us with the tools needed to carry out the most common calculations in lexical availability studies: total number of words (tokens) and different words (types) counts, the average number of responses given by informant, and the lexical availability index. This last measure accounts for the number of informants who generated a given word within a semantic category and the position in which they produced the word.

We also provide the cohesion index by applying Max Echeverría's formula (1991). This index relates the values obtained in tokens and types by dividing the average number of responses given by the informant in each centre of interest by the number of different words. In this way, it can be determined which semantic categories are compact (or closed); that is to say, the degree of coincidence in informants' word responses.

In order to carry out a qualitative comparison of the Middle school and Year 0 lexicons, I have restricted the comparable lexical units to those with an availability index (a.i.) equal to or higher than 0.1, as was done by Carcedo González (2000). These units are those which, being mentioned more frequently by the participants and being placed higher in the lists, correspond to widely available lexicon: active vocabulary. The need for this limitation is justified by Samper Padilla (1999: 554) in the following way:

> After a determined point on the records of availability, a group of words appear which are mentioned by very few participants and which, furthermore, are included in the answer sheets in positions of little relevance. Therefore, we could find ourselves comparing phenomena particular to the language of a specific group of participants or even only one informant, rather than comparing general facts.

As a result, I have obtained lists of widely available lexicon (active vocabulary) from both Year 0 and Middle school. I have listed the common types with an a.i. of >0.1, in the initial ten positions (more active vocabulary), the remaining common lexicon which falls into this interval of availability, and the types with an a.i. of >0.1, (those found on one list are not repeated on the other).

For each centre of interest I classified word responses according to the CEFR levels. To make this classification, I contrast the types obtained to the guidelines included in the chapter of "Specific notions" of the *Niveles de referencia para el español*, belonging to *Plan curricular del Instituto Cervantes* (Instituto Cervantes 2007), and

the glossaries of *Aula Virtual de Español, AVE, del Instituto Cervantes* http://www.ave.cvc.cervantes.es), at the levels A1-A2 and B1-B2. When none of these words are found on these lists, I have contrasted specific materials to teach lexicon, in which lists of words according to the CEFR levels are offered, such as *Vocabulario*, *Elemental A1-A2,* and, *Medio B1,* by Baralo et al. (2008, 2009), and textbooks used at the first bilingual stage, such as *Club Prisma A1, A2, A2-B1* and *B1* (Equipo Club Prisma 2008/2010).

9.5 Results

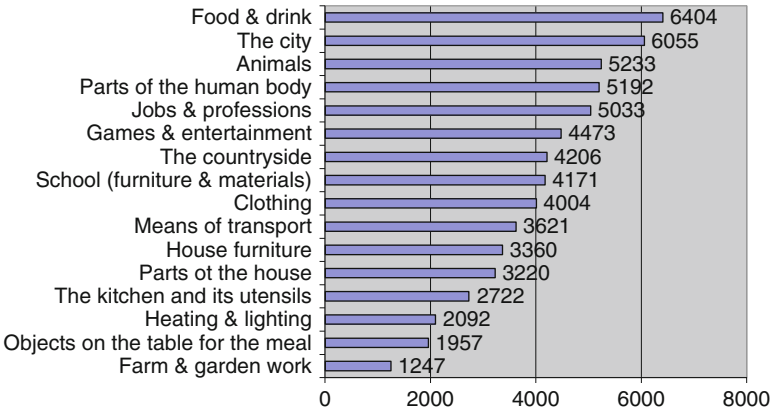
9.5.1 Overall Results: Lexical Availability in the Spanish Bilingual Sections

The general statistical results counted 62,990 words in total, giving an average of 190 words per student and more than 3,936 per category of interest. However, distribution relating to center of interest shows some important differences. 'Food and drink' (6,404) and 'The city' (6,055) are the most productive categories in terms of tokens, followed by 'Animals' (5,233), 'Body parts' (5,192) and 'Jobs and professions '(5,033); students retrieved the least amount of words in the centers of interest 'Heating and lighting' (2,092), 'Objects on the table for the meal' (1,957) and 'Farm and garden work' (1,247) (Graph 9.1).

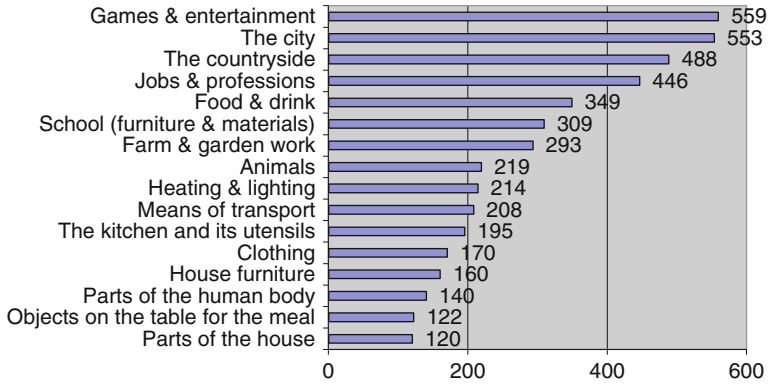
Now taking types, or different words, into consideration, the sample produced a total of 4,545 types, with an average of 284 types per center of interest. The distribution varies according to the specific category. Those with the highest number of types are 'Games and entertainment' (559), 'The city' (553), 'The 'Countryside' (488) and 'Jobs and professions' (446); with a lower number of types appearing in 'House furniture' (160), 'Body parts' (140), 'Objects on the table for the meal' (122) and 'Parts of the house' (120). Furthermore, we can also observe the closeness of the interval dispersion of lexical richness below the average (91 units) as opposed to the widespread interval above the average (208 units). We also can see the categories containing a low degree of lexical richness, 'House furniture' (141), 'The kitchen' (160), 'Heating and lighting' (183), 'Means of transport' (184) and 'Animals' (187) (Graph 9.2).

Comparing tokens and types, it can be seen that, in line with other studies, the ranges which correspond to the different categories of interest for total words and different words are not proportionate. In this sense, it is striking that categories which are very productive in terms of tokens – 'Animals' (3), 'Body parts' (4), have considerably smaller type ranges – 8 and 14 respectively; or that 'Farm and garden work', with a range of 16 for tokens increases to a range of seven for types (Table 9.4).

The cohesion index indicated that the most compact category is 'Body parts' (0.11), followed by 'Parts of the house' (0.8), and 'Animals', 'Clothing' with (0.7).



Graph 9.1 Total words in the general distribution



Graph 9.2 Total of word types in the general distribution

On the other hand, the most open, that is to say, the category with the widest variety of replies is 'Farm and garden work' (0.01), followed by 'Games and entertainment' (0.02), and a wide group ('The 'Countryside', 'Jobs and professions', 'The city', 'Heating and lighting') with an index of (0.03) (Table 9.5).

The words which appear most frequently among Polish students – mentioned by more than 75 % of the participants – are, in decreasing order: *coche* 'car' (97.3 %), *perro* 'dog' (97.3 %), *gato* 'cat' (96.9 %), *ojo* 'eye' (96.4 %), *bicicleta)* 'bike/bicycle' (92.7 %), *plato* 'plate' (92.4 %), *(auto)bús* 'bus' (91.8 %), *lámpara* 'lamp' (90.6 %), *cocina* 'kitchen' (89.7 %), *cabeza* 'head' (89.4 %), *mesa* 'table' (89.4 %; 84.6 %; 76.4 %), *profesor* 'professor' (89.1 %), *mano* 'hand' (88.2 %), *pantalón* 'trousers' (86.4 %), *silla* 'chair' (85.8 %; 81.6 %), *camisa* 'shirt' (85.2 %), *cama* 'bed' (83.4 %), *nariz* 'nose' (83.1 %), *avión* 'plane' (82.2 %), *calle* 'street' (79.5 %),

Table 9.4 Comparison of ranges for tokens and types

Tokens	Range	Types
05 'Food & drink'	1	15 'Games & entertainment'
10 'The city'	2	10 'The city'
14 'Animals'	3	11 'The 'Countryside'
01 'Parts of the human body'	4	16 'Jobs & professions'
16 'Jobs & professions'	5	05 'Food & drink'
15 'Games & entertainment'	6	08 'School' (furniture & materials)
11 'The 'Countryside'	7	13 'Farm & garden work'
08 'School' (furniture & materials)	8	14 'Animals'
02 'Clothing'	9	09 'Heating & lighting'
12 'Means of transport'	10	12 'Means of transport'
04 'House furniture'	11	07 'The kitchen and its utensils'
03 'Parts of the house'	12	02 'Clothing'
07 'The kitchen and its utensils'	13	04 'House furniture'
09 'Heating & lighting'	14	01 'Parts of the human body'
16 'Objects on the table for the meal'	15	16 'Objects on the table for the meal'
13 'Farm & garden work'	16	03 'Parts of the house'

Table 9.5 Cohesion index according to centre of interest

N.	Centre of interest	Cohesion index
01	'Parts of the human body'	0.11
03	'Parts of the house'	0.08
02	'Clothing'	0.07
14	'Animals'	0.07
04	'House furniture'	0.06
05	'Food & drink'	0.06
06	'Objects on the table for the meal'	0.05
12	'Means of transport'	0.05
07	'The kitchen and its utensils'	0.04
08	'School' (furniture & materials)	0.04
09	'Heating & lighting'	0.03
10	'The city'	0.03
16	'Jobs & professions'	0.03
11	'The Countryside'	0.03
15	'Games & entertainment'	0.02
13	'Farm & garden work'	0.01

caballo 'horse' (77.6 %), and *pierna* 'leg' (75.2 %). All of these words – 22 in all- represent things immediately connected to daily life. By center of interest, the representation is the following: 5 words from 'Body parts' (*ojo* 'eye', *cabeza* 'head', *mano* 'hand', *nariz* 'nose', *pierna* 'leg'), 4 from 'Means of transport' (*coche* 'car', *bici(cleta)* 'bike/bicycle', *(auto)bús* 'bus', *avión* 'plane), 3 from 'Animals' (*perro* 'dog', *gato* 'cat', *caballo* 'horse') and 3 from 'House furniture' (*mesa* 'table', *silla* 'chair', *cama* 'bed'), 2 from 'Clothing' (*pantalón* 'trousers', *camisa* 'shirt') and 2 from 'School'

(*mesa* 'table', *silla* 'chair'), and 1 from 'Heating and lighting', 'The city', 'The kitchen', 'Objects on the table for the meal', 'Parts of the house' and 'Jobs and professions'. More open fields – 'Food and drink', 'The 'Countryside', 'Games and entertainment' – and more specialised fields – 'Farm and garden work' – are not represented.

9.5.2 Comparison of Year 0 vs. Middle School

Quantitative Analysis

The statistical analysis applied to the data yields a total of 20,407 words in Middle school and 16,498 in Year 0. However, given that the samples compared are not identical – 120 Middle school students to 90 high school students, the average of responses per student in both groups is compared.

As can be gathered from Table 9.6, the average per category of interest is higher in Year 0 (11.46) in comparison to Middle school (10.63) by nearly one word more. In general, to a greater or lesser extent, word production is higher in Year 0 than Middle school in 13 of the 16 centers of interest. The biggest differences favouring Year 0 can be found in "Body parts" (+4.15), 'Clothing' (+1.68), 'Objects on the table for the meal' (+1.67) and 'Jobs and professions' (+1.57). The three centers of interest with a higher production in Middle school are 'The city' (−0.34), 'Food and drink' (−0.71) and 'School' (−1.45). This last category shows a significant difference.

The distribution of the centers of interest coincides quite a lot, as in both educational modalities, the centers of interest found above and below the average are the same, with the ranges coinciding in seven categories: 'Food and drink' (1), 'The city' (2), 'Means of transport' (10), 'House furniture' (11), 'Clothing' (12), 'The kitchen' (13) and 'Farm and garden work' (16). In the remaining cases, the ranges are very close, with differences of between 1 and 3 points, although the case of 'Body parts' should be highlighted, with a range of 7 for Middle school and 2 for Year 0 (Table 9.7).

If we focus on a similar comparison of word types, the average per center of interest is also higher in Year 0 (134.24) in comparison to middle school (120.47). The vocabulary is therefore more varied in Year 0. However, some changes must be noted. Here, the number of categories with a higher production in Year 0 reduces to 11, with the biggest advantages in 'The city' (+43), 'Animals' (+36), 'Clothing' (+30) and 'Objects on the table for the meal' (+30). After these, 'The kitchen' with +25 and the three categories 'Body parts', 'Food and drink' and 'The 'Countryside' with +24. The advantages favouring Middle school are minimal and of little significance 'School' (−8), 'House furniture' (−4), 'Farm and garden work' (−4) and 'Heating and lighting' (−2).

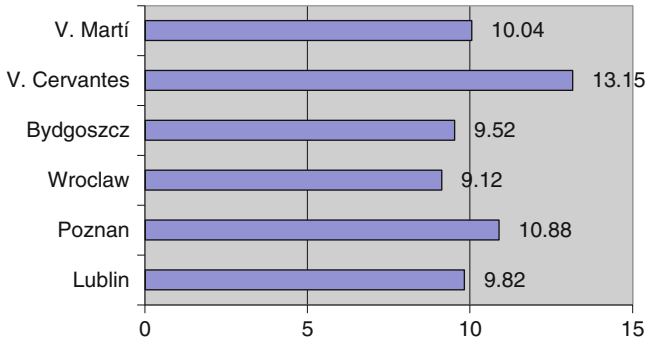
The distribution of the categories also coincide here, as the center of interest which are found above and below the average is the same. With regard to ranges, they are identical in half of the cases: 'Games and entertainment' (1), 'The city' (2), 'Jobs

Table 9.6 Comparison of word average in the initial bilingual stage

Middle school	Interval	Year Zero
'Food & Drink' – 18.34	16.5–18.5	17.63 – 'Food & Drink'
'City' – 16.46	14.5–16.5	16.12 – 'City'
		16.12 – 'Body'
		14.69 – 'Professions'
'Animals' – 13.83	12.5–14.5	14.22 – 'Animals'
'Entertainment' – 13.59		14.18 – 'Entertainment'
'School' – 13.21		
'Professions' – 13.12		
'Body' – 11.97	10.5–12.5	12.31 – 'Clothing'
'Countryside' – 11.04		11.76 – 'School'
'Clothing' – 10.63		
Average – 10.63		11.48 – 'Countryside'
		11.46 – Average
		11.06 – 'Transport'
'Transport' – 10.09	8.5–10.5	10.36 – 'Furniture'
'Furniture' – 9.58		9.68 – 'House'
'House' – 8.57		
'Kitchen' – 6.94	6.5–8.5	8.42 – 'Kitchen'
'Heat & Light' – 5.54	4.5–6.5	5.90 – 'Table'
		5.79 – 'Heat & Light'
'Table' – 4.23	2.5–4.5	3.60 – 'Garden work'
'Garden work' – 2.92		

Table 9.7 Comparison of word type production in the initial bilingual stage

Middle school	Interval	Year Zero
'Entertainment' – 271	264–297	281 – 'Entertainment'
	231–264	264 – 'City'
'City' – 221	198–231	213 – 'Professions'
'Professions' – 203		
'Countryside' – 172	165–198	196 – 'Countryside'
'Food & drink' – 171		195 – 'Food & Drink'
'School' – 158	132–165	150 – 'School'
		142 – 'Animals'
		134 – Average
Average – 120	99–132	115 – 'Clothing'
'Transport' – 110		106 – 'Kitchen'
'Animals' – 106		101 – 'Garden work'
'Garden work' – 105		
'Heat & Light' – 97		
'Furniture' – 94	66–99	95 – 'Heat & Light'
'Clothing' – 85		95 – 'Body'
'Kitchen' – 81		91 – 'Transport'
'Body' – 71		90 – 'Furniture'
		75 – 'House'
		69 – 'Table'
'House' – 64	33–66	
'Table' – 39		



Graph 9.3 Word average according to educational context

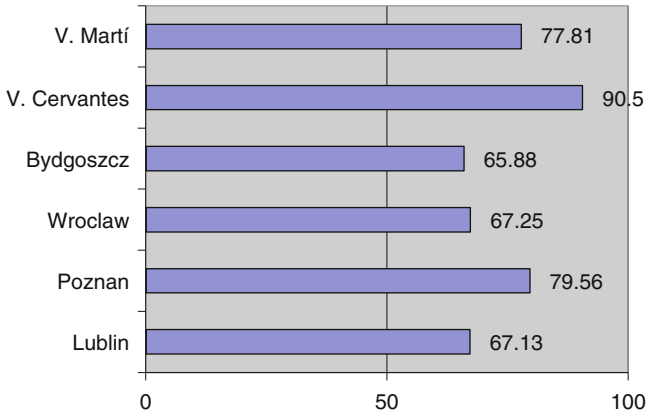
and professions' (3), 'The 'Countryside' (4), 'Food and drink' (5), 'School' (6), 'Parts of the house' (15), and 'Objects on the table for the meal' (16). Only the ranges in 'Clothing' (12–8) and 'The kitchen' (13–9) are different, by 4 points, favouring Middle school, and 'Means of transport' (7–13) by 6 points, favouring Year 0.

Until now, an advantage towards the Year 0 group has been demonstrated, but are the differences shown based on a homogenous distribution of data within each group? As can be seen in Graph 9.3, this is not the case. Whilst Middle school gives a normal distribution, around an average of 9.94 words with a standard deviation of 0.89, with a maximum of 10.88 (Poznan) and a minimum of 9.12 (Wroclaw); in Year 0, with the average being 10.90 words, the standard deviation is 1.96, with a maximum of 13.15 words (Warsaw Cervantes) and a minimum of 9.52 (Bydgoszcz).

As such, in relation to the word average, a much higher production can be observed in Warsaw Cervantes with regard to Year 0, and a notable advantage in Poznan as well, in the Middle school. If we disregard these two schools, the other four Warsaw Marti, Bydgoszcz, Wroclaw and (Lublin), either Middle school or Year 0, fairly similar results are found, with an average of around 9.62 words. Two of the schools with Year 0 have a lower average than that of Poznan Middle school with the highest average, (and one of them, Bydgoszcz), has lower results than two of the Middle schools.

A similar situation can be seen regarding the word types in relation to centers of interest and city. Here, the distribution continues to be normal in Middle school, with an average of around 71.3 types and a standard deviation of 7.14, with a maximum of 79.56 (Poznan) and a minimum of 67.25 (Wroclaw). In Year 0, the average is 78.06 types and the standard deviation is 12.31, with a maximum of 90.5 (Warsaw Cervantes) and a minimum of 65.88 types (Bydgoszcz) (Graph 9.4).

In the case of type distribution in relation to the educational stage, the situation is found to be quite similar to that seen for the word average. More than one division in the two groups, Middle school and Year 0, can be observed, with three heterogeneous groups distinguished: a trio formed by Bydgoszcz (Year 0), Wroclaw and Lublin (Middle school), with around an average of 78.68 word types, and significantly above the others, Warsaw Cervantes with an average of 90.5.



Graph 9.4 Word type production according to educational context

9.5.3 Qualitative Analysis: Structure of Active Vocabulary

As I have already mentioned in the methodology section, in order to analyse the structure of the most active vocabulary of the two groups under study as well as the convergences and divergences within lexical inventories of educational stages, I have restricted comparable lexical units to those whose availability index is equal or higher than 0.1, discarding any *lexias* of inferior values. Thus the standard-related facts are compared.

After collecting all available vocabulary with the indices exceeding 0.1, the proportion of types to be compared is relatively small. As a result, the total number of 2,048 and 2,282 lexical units which form the respective inventories of Middle school and Year Zero has been reduced to 286 (14.0 %) and 297 (13.0 %) units (Table 9.8).

As a result of the examination and one-by-one comparison of lexical units which form the inventories of highly available vocabulary: (active vocabulary) in Middle school and Year Zero different lists were obtained. These lists present common types with a.i. >0.1 up to position 10 (the most active vocabulary), the remaining common vocabulary within this interval of availability as well as the types with a.i. >0.1, which are present on one list and not on the remaining ones- highly available vocabulary which is exclusive to Middle school and Year Zero. All those lists are presented as appendix attached to this chapter. Table 9.9 below displays a quantitative summary of the lists included in the appendix.

A close inspection of such Table reveals a strong uniformity in the lexical units which produce the highest levels of availability for each educational stage; this provides us with a solid common base for comparison. A large amount of these available words coincide on both models for the initial stage, as can be observed in the column "Top 10 common types". Here, word types completely coincide for 'Clothing' (10) and are very similar for 'Body parts', 'The 'Countryside', 'Means

Table 9.8 Distribution of word types (a.i >0.1), according to centre of interest

	Gimnazjum		Year 0	
C.I.	Absolute frequency	Relative frequency (%)	Absolute frequency	Relative frequency (%)
01.	21:71	29.6	24:95	25.3
02.	19:85	22.3	23:115	20.0
03.	15:64	23.4	17:75	22.7
04.	14:94	14.9	17:90	18.9
05.	33:171	19.3	33:195	16.9
06.	7:39	17.9	8:69	11.6
07.	11:81	13.6	19:106	17.9
08.	19:158	12.0	16:150	10.7
09.	8:97	8.2	9:95	9.5
10.	26:221	11.8	24:264	9.09
11.	20:172	11.6	20:196	10.2
12.	16:110	14.5	13:91	14.3
13.	4:105	3.8	8:101	7.9
14.	29:106	27.4	24:142	16.9
15.	20:271	7.4	20:285	7.0
16.	24:203	11.8	22:213	10.3
Total	286:2,048	14.0	297:2,282	13.0

Table 9.9 The most available word types on the lists of Middle school and Year Zero

C. I.	Top ten common types with a.i. >0.1	Remaining common types with a.i. >0.1	Types with a.i. >0.1 exclusive to Gimnazjum	Types with a.i. >0.1 exclusive to Year Zero
01.	9	10	2	5
02.	10	8	1	5
03.	7	6	2	4
04.	8	3	3	6
05.	8	18	7	7
06.	5	0	2	3
07.	5	5	1	9
08.	7	6	6	3
09.	6	0	2	3
10.	7	10	9	7
11.	9	6	5	5
12.	9	3	4	1
13.	2	0	2	6
14.	9	10	10	5
15.	8	4	8	8
16.	6	9	9	7
Total	115	98	73	84

of transport and 'Animals' (9 types), and 'House furniture', 'Food and drink' and 'Games and entertainment' (8 types). This convergence of lexical repertoires is equally evident in "Remaining common types with an a.i. of >0.1", with 18 common word types in 'Food and drink' and 10 in 'Body parts', 'The city' and 'Animals'.

Within common vocabulary with a.i. > 0.1 results show important differences in productivity for the different categories. As opposed to categories with a varied vocabulary – 'Food and drink' (26), 'Body parts' (19), 'Animals' (19), 'Clothing' (18), 'The city' (17), 'The Countryside' (15) and 'Jobs and professions' (15) – others present a very limited widely available vocabulary, and even the absence of common types in the positions after the initial 10 with an a.i. of >0.1. – 'The kitchen' (10), 'Heating and lighting' (6), 'Objects on the table for the meal' (5), 'Farm and garden work' (2).

In both groups – Middle school and Year Zero–, and for each subject stimulus, the types with an a.i. of >0.1 specific to one of the educational modalities appear, as well as the richest lexicon given by one group or the other, according to category of interest. Year 0 leads in eight categories, showing a significant difference in six of these: 'The kitchen' (9), 'House furniture' (6), 'Farm and garden work' (6), 'Body parts' (5), 'Clothing' (5), 'Parts of the house' (4). On the other hand, the widely available vocabulary in Middle school is higher in five categories: 'Animals' (10), 'The city' (9), 'Jobs and professions' (9), 'School' (6), 'Means of Transport'(4). In each of these cases, the difference is significant, though fields with more varied associations like 'The city' and 'Jobs and professions' tend to favour a higher number of word types also specific to Year 0. The same circumstances influence on the seven exclusive types in both groups for the category 'Food and drink'.

If we look at the data from the perspective of the percentage of students who retrieved each word, we find that in Middle school 22 types were produced by more than 75 % of the participants; in the case of lyceum this number increases to 27 types. The types common to both groups are as follows: *perro* (dog), *gato* (cat), *coche* (car), *ojo* (eye), *mesa* (table), *bicicleta* (bike/bicycle), *cabeza* (head), *plato* (plate), *(auto)bús* (bus), *profesor* (professor), *mano* (hand), *pantalón* (trousers), *lámpara* (lamp), *cocina* (kitchen), *camisa* (shirt) and *cama* (bed). As can be inferred from the above, these were produced by more than 75 % of informants. Next to these word types other five exclusive category types appear on the Middle school list, such as *silla* (chair), *mesa* (table), *habitación* (room) and *calle* (street); while in the case of lyceum other ten types have been listed, such as *avión* (plane), *nariz* (nose), *dormitorio* (bedroom), *libro* (book), *oreja* (ear), *camiseta* (t-shirt), *cuchillo* (knife), *bolígrafo* (ball-pen), *agua* (water) and *caballo* (horse).

The percentage of highly available word types retrieved by 75 % or more informants is quite low: 7.7 % in Middle school, and 9.1 % in Year Zero, with significant advantage on the part of Year Zero. However, if we extend the studied interval up to 50 %, the percentages become equal, as a result of which both in Middle school and in Year Zero half of the students produced a little more than a quarter of the active vocabulary: to be exact, 26.5 % in Middle school and 26.6 % in Year Zero.

All in all, it should be remembered that many types are repeated in different categories. Therefore, if we consider the totality of the vocabulary with a.i. >0.1, in

the case of Middle school we can find 79 entries corresponding to 35 actual types (27.6 % of the active vocabulary), while in the case of Year Zero we can find 83 entries corresponding to 39 truly different types (27.9 % of the active vocabulary). Therefore the percentage values of the repeated vocabulary within semantic categories are similar.

We can distinguish three groups within such active repeated vocabulary:

(a) Common repeated types
(*auto)bús* (bus), *árbol* (tree), *caballo* (horse), *casa* (house), *cerdo* (pig), *cine* (cinema), *coche* (car) *cuchillo* (knife), *frigorífico* (fridge), *gallina* (hen), *gato* (cat), *lámpara* (lamp), *lavadora* (washing machine), *mesa* table), *ordenador* (computer) *pájaro* (bird), *perro* (dog), *plato* (plate), *pollo* (chicken), *puerta* (door), *silla* (chair), *teatro* (theatre), *televisión* (televisión), *tienda* (shop), *tranvía* (tramway), *vaca* (cow), *ventana* (window).
(b) Repeated types exclusive to Middle school:
armario (wardrobe), *bicicleta* (bicycle), *bufanda* (scarf), *discoteca* (discotheque), *flor* (flower), *fregadero* (sink), *gente* (people), *iglesia* (church).
(c) Repeated types exclusive to Year Zero:
agua (water), *baño* (bathroom), *cocina* (kitchen), *cuchara* (spoon), *escritorio* (desk), *gallo* (rooster), *jardín* (garden), *jardinero* (gardener), *libro* (book), *metro* (subway), *nevera* (fridge), *olla* (pot), *oveja* (sheep), *tenedor* (fork), *tren* (train).

The repeated vocabulary can be grouped in the following categories: 18 word types in the category 'The 'Countryside', 14 in 'The city', 13 in 'The kitchen', 12 in 'House furniture', 10 in 'Animals', 10 in 'School', 7 in 'Means of transport', 6 in 'Games and entertainment', 5 in 'Parts of the house', 5 in 'Objects on the table', 3 in 'Heating and lighting', 2 in 'Food and drink', 2 in 'Farm and garden work', 1 in 'Jobs and professions' and 1 in 'Clothing'.

As for the characteristics of the highly available word types, their comparison with the inventories of "Specific notions" listed in *Niveles de referencia para el español* (The levels of reference for the Spanish language) under *Plan curricular del Instituto Cervantes* (the Curriculum Plan of the Cervantes Institute) the Cervantes Institute (2007), as well as their comparison with the vocabulary lists of the most frequently used Spanish language manuals (Equipo Club Prisma 2008/2010) at the initial bilingual stage and with specific materials for teaching vocabulary (Baralo et al. 2008, 2009) shows that this classification is a reliable reflection of the complexity level of active vocabulary and can serve as a manner of establishing correspondence between the level of planned and real lexical command.

With reference to the above, with the exception of 'Objects placed on the table for the meal' and 'Farm and garden work', in the other 14 thematic centres, the common active vocabulary until position 10 belongs to level A. This situation is repeated in the rest of common types with a.i. of >0.1. in 10 centres of interest ('Clothing', 'Parts of the house', 'Food and drink', 'The kitchen, School', 'The city', 'The 'Countryside', 'Means of transport', 'Games and entertainment' and 'Jobs and professions').

On the other hand, there appears most words of level B1 in 'Objects placed on the table for the meal' and 'Farm and garden work' in the first ten positions, as well as in the rest of common types with a.i. of >0.1 in 'Body parts', 'House furniture' and 'Animals'.

As refers to the level of difficulty of the word types, exclusive to each group, also the majority belongs to level A. However, also level B types are included on both lists; still, those types of superior level are more frequently present on Year Zero lists. Thus the presence of level B word types is significant (at least two types) in 3 centres of interest ('The animals', 'Games and entertainment') when compared with 7 centres in Year Zero ('Parts of the human body', 'Clothing', 'The kitchen', 'The 'Countryside', 'Farm and garden work', 'Animals' and 'Jobs and professions').

Table 9.10 summarizes the lexical structure of the lists of learners' widely available lexicon according to the CEFR levels. Shaded are the levels with higher contribution of vocabulary.

9.6 Discussion

The objective of this investigation was to determine which of the modalities of the initial bilingual stage guarantees more efficient and consolidated command of the Spanish language. Objective data based on the results of surveys of lexical availability seem to indicate better lexical competence of Year Zero over Middle school. This was demonstrated by the mean values of the average number of responses given by the informants and the production types with the average advantage of +0.83 words per informant and +13.77 types per centre of interest.

When interpreting these results, it is necessary to take into account a varied range of factors (age, motivation, cognitive capacity, or world experience). All these factors are related to the selected modality at the initial bilingual stage. As for the learners' age, the results of this study are similar to those obtained by Gallardo del Puerto and Martínez Adrián (reported in Chap. 4) in the case of English vocabulary acquisition, and support the hypothesis held in studies on age-related differences in SLA (e.g., Lightbown 2008; Muñoz 2008), in which it is claimed that older learners are better and more efficient in vocabulary knowledge, especially in school settings. It is also in agreement with late starters' faster rate of acquisition at the initial stage (Gallardo del Puerto 2007; García Mayo and García Lecumberri 2003; Muñoz 2006). The above examples also seem to support the arguments of Tatoj et al. (2008), concerning the benefit of greater maturity of lyceum students at the moment of learning Spanish as a foreign language (SFL).

The advantage of Year Zero over Middle school is not systematic nonetheless, as slight advantages occur for Middle school in the average number of word responses given by informants for 'Food and drink' (+0.71), 'The City'(+0.34) and especially 'School' (+1.45), as in words (types) for 'School' (+8), 'Farm and garden work' (+4), 'Heating and lighting' (+2), 'House furniture' (+4), and especially meaningfully in 'Means of transport' (+19).

Table 9.10 Widely available lexicon distributed according to levels

C.I.	Top ten common types with a.i. > 0.1		Remaining common types with a.i. > 0.1		Types with a.i. > 0.1 exclusive to gimnazjum		Types with a.i. > 0.1 exclusive to Year Zero	
01	A1: 4	A2: 5	A1: 2	A2: 3	A1: 0	A2: 1	A1: 0	A2: 0
	B1: 0	B2: 0	B1: 5	B2: 0	B1: 1	B2: 0	B1: 2	B2: 3
02	A1: 6	A2: 4	A1: 1	A2: 6	A1: 0	A2: 0	A1: 0	A2: 3
	B1: 0	B2: 0	B1: 1	B2: 0	B1: 1	B2: 0	B1: 2	B2: 0
03	A1: 7	A2: 0	A1: 3	A2: 3	A1: 0	A2: 1	A1: 3	A2: 1
	B1: 0	B2: 0	B1: 0	B2: 0	B1: 1	B2: 0	B1: 0	B2: 0
04	A1: 8	A2: 0	A1: 0	A2: 1	A1: 2	A2: 1	A1: 4	A2: 1
	B1: 0	B2: 0	B1: 2	B2: 0	B1: 0	B2: 0	B1: 1	B2: 0
05	A1: 5	A2: 3	A1: 9	A2: 9	A1: 3	A2: 4	A1: 4	A2: 2
	B1: 0	B2: 0	B1: 0	B2: 0	B1: 0	B2: 0	B1: 1	B2: 0
06	A1: 0	A2: 2	A1: 0	A2: 0	A1: 0	A2: 1	A1: 0	A2: 2
	B1: 3	B2: 0	B1: 0	B2: 0	B1: 0	B2: 0	B1: 0	B2: 1
	C1: 0	C2: 0	C1: 0	C2: 0	C1: 1	C2: 0	C1: 0	C2: 0
07	A1: 2	A2: 2	A1: 0	A2: 3	A1: 1	A2: 0	A1: 0	A2: 2
	B1: 1	B2: 0	B1: 2	B2: 0	B1: 0	B2: 0	B1: 6	B2: 1
08	A1: 5	A2: 2	A1: 5	A2: 0	A1: 5	A2: 1	A1: 3	A2: 0
	B1: 0	B2: 0	B1: 0	B2: 1	B1: 0	B2: 0	B1: 0	B2: 0
09	A1: 4	A2: 2	A1: 0	A2: 0	A1: 0	A2: 1	A1: 1	A2: 1
	B1: 0	B2: 0	B1: 0	B2: 0	B1: 0	B2: 1	B1: 0	B2: 1
10	A1: 7	A2: 0	A1: 8	A2: 1	A1: 3	A2: 5	A1: 6	A2: 0
	B1: 0	B2: 0	B1: 1	B2: 0	B1: 1	B2: 0	B1: 1	B2: 0
11	A1: 3	A2: 6	A1: 3	A2: 1	A1: 3	A2: 1	A1: 1	A2: 1
	B1: 0	B2: 0	B1: 2	B2: 0	B1: 1	B2: 0	B1: 2	B2: 1
12	A1: 6	A2: 2	A1: 2	A2: 1	A1: 2	A2: 2	A1: 0	A2: 0
	B1: 1	B2: 0	B1: 0	B2: 0	B1: 0	B2: 0	B1: 1	B2: 0
	C1: 0	C2: 0	C1: 0	C2: 0	C1: 0	C2: 1	C1: 0	C2: 0
13	A1: 0	A2: 0	A1: 0	A2: 0	A1: 0	A2: 1	A1: 2	A2: 1
	B1: 2	B2: 0	B1: 0	B2: 0	B1: 0	B2: 0	B1: 3	B2: 0
14	A1: 0	A2: 7	A1: 0	A2: 2	A1: 2	A2: 4	A1: 0	A2: 0
	B1: 2	B2: 0	B1: 7	B2: 1	B1: 4	B2: 0	B1: 3	B2: 2
15	A1: 7	A2: 0	A1: 4	A2: 0	A1: 5	A2: 0	A1: 6	A2: 1
	B1: 1	B2: 0	B1: 0	B2: 0	B1: 3	B2: 0	B1: 1	B2: 0
16	A1: 5	A2: 0	A1: 6	A2: 2	A1: 1	A2: 5	A1: 1	A2: 3
	B1: 1	B2: 0	B1: 1	B2: 0	B1: 3	B2: 0	B1: 1	B2: 2

About this particular, the influence of schooling hours and the continued practice of this school-related vocabulary seem to be evident in the case of the Middle school's advantage in this subject. Similarly, a very plausible explanation for the four average additional words which a student in Year 0 is able to produce relating to the stimulus 'Body parts' could be attributed to the interest in the

semantic category shown by 16 year old adolescents in comparison to 13 year old pupils age (at which the vocabulary is learnt).

The influence of the 'sex' factor, which interacts with the 'age' factor, can also be noted on the results, since the sample is mostly female. In my opinion, this explains the greater number of words and vocabulary variety about 'Clothing' in Year Zero: +1.68 words per students and +30 types. The interest in 'Clothing' and the world of fashion is also stronger in a female student of 16 than in pupils of 13–14 years old age (at which this subject is dealt with in Middle school). The greatest production and variety of the vocabulary in the class entitled 'Clothes', which is traditionally associated with the social female role has also been noted in the majority of investigations on lexical availability in the Spanish language (Samper Padilla y Samper Hernández 2006: 51).

In the interpretation of the quantitative data, one must take into account the anomalies shown in the distribution of tokens and types, in the analysis of the sampled schools. Significant differences could be noted in data distribution, which placed the Cervantes Lyceum in Warsaw with +2.27 words per student before the following centre, which surprisingly was a Middle school from Poznań. This tendency was also observed regarding types, in which the same schools ranked first and second: the Cervantes Lyceum in Warsaw with 90.5 types, and Poznań Middle school with 79.56 types. In the lower distribution range low results of the Lyceum in Bydgoszcz stood out: 9.52 words per student and 65.88 types on average, which situated it in the result range which was characteristic for other centres from the Middle school group. These results suggest the positive influence of additional factors, such as the teaching work itself, the used textbooks or the selection of students.

From the point of view of the objectives of teaching Spanish at the initial bilingual stage, it attempts to equip the student with lexical competence which will allow him to engage in effective communication on the level of an independent language user. This is the reason for the importance of analysing active vocabulary of a student learning Spanish as a foreign language at the end of the first bilingual stage, which in this case is identified with the available vocabulary, the indices of which exceed 0.1.

In this study, the reduced values of the magnitudes of highly available vocabulary in the Middle school −286 (14.0 %) and Year Zero −297 (13.0 %) point to what was already noted by Carcedo González (2000: 160): "much more limited magnitudes of Spanish vocabulary acquired by foreigners are accompanied by major coincidence of answers". In fact, the percentage values in Poland, especially in the case of Year Zero, show striking similarity to the value of 13 % which was obtained by Carcedo González (2000) upon having limited the vocabulary of pre-university students in Finland to the a.i. range >0.1.

In Polish bilingual sections, the productivity differences between lexical categories within the common vocabulary with the a.i. > 0.1 are characterised by the same distribution of the words: those centres of interest which appear with increased productivity have been located above the average and those with less common vocabulary have been located below the average. If we look at the indices of cohesion obtained in the general distribution, we can see that such distribution includes both the closed-system centres ('The human body', 'Clothing' and the open ones 'The city', 'The 'Countryside'). Therefore it can be observed that the results obtained by Polish students are similar to

those reported in studies in the Hispanic world; both the more and the less productive centres at the initial bilingual stage in Poland correspond to those which appeared above and below the average in the investigations under the PanHispanic Project of Lexical Availability (Samper Padilla et al. 2003: 57–60). The fact that the results are similar both in Spanish native speakers and Polish students learning Spanish as a foreign language may be attributable to less practise and less frequent use of the vocabulary related to these areas rather than to learning differences.

As for the word types produced by more than 75 % of students, we could say that it was a quite stable vocabulary, given that all words common to both modalities of the initial bilingual stage were present on the list of types produced by more than 75 % of student of the general sample. In the same manner two types from the Middle school (*silla* 'chair' and *calle* 'street') as well as three types of Year Zero (*avión* 'plane', *nariz* 'nose' and *caballo* 'horse') were present on that list.

Type repetitions on the active vocabulary lists did not go unnoticed, given the limited range of the studied sets – 286 types in the Middle school and 297 types in Year Zero – in contrast with the considerable percentage of produced repeated vocabulary, which in both cases exceeded 27.5 %. The three groups of repeated types – 27 common types, 8 types exclusive to Middle school and 15 repeated types exclusive to Year Zero – depict the phenomenon already noted by Samper Padilla et al. (2003) in the case of words such as *silla* 'chair', which was present in 6 centres on the lists of Gran Canaria, or *jardín* 'garden' present in 7 centres on the availability lists of France and Acadia presented by Mackey (1971: 417). All that shows that semantic classes proposed by Gougenheim et al. (1956) show frequent intersection areas, or even the areas of inclusion, as in the case of 'Objects on the table for a meal' in relation to 'The kitchen and its utensils'.

In the analysed sample these intersections are present mainly in the centres with very broad associations, such as 'The 'Countryside' (18 types) and 'The city' (14), as well as in directly related centres, such as 'The 'Countryside' (18) and 'Animals' (10) or 'The kitchen' (13), 'House furniture' (12), 'Parts of the house' (5) and 'Objects on the table' (5).

Finally, the classification of highly available types according to the *Reference levels of the Common European Framework of Reference for Languages* (the Council of Europe 2001) shows that common active vocabulary belongs to fundamental lexis which is necessary for communication. The organization of this lexis on the lists of common available vocabulary shows that the words classified under level A in the reference inventories for teaching Spanish as a foreign language come first on those lists. Only exceptionally the majority of vocabulary of level B appears on any of the first ten positions in two centres of interest with a more specialist vocabulary ('Objects on the table for the meal' and 'Farm and garden work'). The majority of word types of the rest of available vocabulary also belong to level A; however the vocabulary of the next level (B) appears more frequently in 'The human body', 'House furniture' or 'Animals', although it is always preceded by a significant number of level A types.

As for the exclusive word types to each group, it should be noted that Year Zero manages to acquire richer vocabulary with frequent B level types in up to 7 centres of interest when compared with 3 centres of interest in the case of the Middle school.

This brings Year Zero closer to the final objective of the initial bilingual stage, which consists in reaching B1 level. In this way the acquisition of such vocabulary would comply with the criteria of communicative profitability and it will be required in order to perform communicative functions and to develop oral and written texts under the expected level at the end of the initial bilingual stage.

In any case, although Year Zero presents an advantage over the Middle school in attaining this objective, the contribution of level B vocabulary is scarce or even none in as many as 9 centres of interest ('Parts of the house', 'House furniture', 'Objects on the table for the meal', 'Heating and lighting', 'Food and drink', 'School', 'The city', 'Means of transport' and 'Games and entertainment'). As can be inferred from the findings in this study, some of those thematic categories with lexical deficiencies are vital for the student who shall perform a role of a social agent, an inter-cultural speaker and an autonomous learner with a view to level B, which he is expected to have reached when starting the second stage of bilingual education.

9.7 Conclusions: Lexical Availability at the End of the Initial Bilingual Stage

In Poland there are two coexisting modalities of the first educational stage in Spanish bilingual sections which follow the same curriculum and cover the same number of teaching hours: one of them is extensive and includes 3 years in the Middle school, the other one is intensive and is conducted in the form of Year Zero at the lyceum. The objective of this study was to find out which of those modalities equips the student with better lexical competence with a view to reaching B1 lexical competence level, required for the second bilingual stage.

To this end a quantitative and qualitative comparative study of lexical availability was conducted, on the basis of the broad sample of students in both modalities (120 at the Middle school and 90 at Year Zero). The results clearly indicate that the Year Zero group has been better prepared for the second stage, since it has reached superior values in terms of the words average, word types, and the quantity and quality of highly available vocabulary.

However, regardless of Year Zero superiority in terms of overall results, the detailed study based on the centres of interest showed slight advantages on the part of the Middle school sample in some centres, which were quite significant in terms of words average in 'School' and in terms of types in 'School' and 'Means of transport'. This could indicate that in the case of some semantic categories longer time of schooling and contact may be an advantage to their fixation in learners' mental lexicon. Such a situation took place regarding the vocabulary related to the school field, where 3 years of continuous practice worked in favour of the Middle school group.

In any case, the study results point at the covariation of the educative modality factor, taking into account other accompanying social factors, such as age or sex. This is especially visible when considering the interest paid by a teenage boy aged 16 to the vocabulary of 'The human body' or a 16-year old girl to the vocabulary of

'Clothes'. In one case the 'age' factor and in another the 'sex' factor increases the advantages of the Year Zero modality with respect to those thematic stimuli. In the same manner, significant differences in the area of lexical availability between both types of educational schools, regardless of the modality followed, also point to other additional factors, such as the teaching work or the selection of students.

Finally, the present study corroborates the impact of the lexical availability task on the obtained results. This is shown in the high coincidence of Polish results with other investigations in the Hispanic world (Samper Padilla et al. 2003) or in the area of teaching Spanish as a foreign language (SFL) (Carcedo González 1998, 2000) or teaching English as a foreign language (EFL) (Germany and Cartes 2000; Jiménez Catalán and Ojeda Alba 2009, 2010). The coincidence of results is particularly evident regarding the productivity distribution of centres of interest or in the area of repeated vocabulary, which can be explained by means of the relation and superimposition of vocabulary under some thematic stimuli.

The final part of the study has been devoted to the description of the structure and composition of highly available vocabulary (active vocabulary). It has been possible to define a quite stable vocabulary common to both modalities, mainly level A vocabulary (basic one), above all in the first ten positions. As for the vocabulary exclusive to each modality, while in the case of the Middle school the majority of types also belong to level A, in the case of Year Zero, level B types (intermediate ones) appear more frequently, although not in all centres of interest. This tendency seems to give advantage to Year Zero students over Middle school with a view to the second stage of bilingual education. However, both of them present lexical gaps in certain thematic categories, if reaching B1 level (an independent user) of lexical competence is concerned.

The study of the structure of the highly available lexicon in relation to CEFR opens new lines for future studies. One of these lines would be tracking lexical items from the inventories of levels A1, A2 and B1, absent in word lists produced by the bilingual students. It would be also advisable to check the correlation between the CEFR levels and the available lexicon produced by Spanish native speakers, since the ultimate goal of learning a language is to ensure communication with native speakers.

Appendix

C.I.	Top ten common word types with an a.i. of >0.1
01	**A1:** *ojo* 'eye', *nariz* 'nose', *pelo* 'hair', *oreja* 'ear'; **A2:** *mano* 'hand', *pierna* 'leg', *dedo* 'finger', *estómago* 'stomach', *cabeza* 'head'.
02	**A1:** *pantalón* 'trousers', *camisa* 'shirt', *camiseta* 't-shirt', *zapato* 'shoe', *jersey* 'jersey', *falda* 'skirt', *vaquero* 'jeans'; **A2:** *blusa* 'blouse', *vestido* 'dress', *calcetín* 'sock'.
03	**A1:** *cocina* 'kitchen', *habitación* 'room', *salón* 'living room', *dormitorio* 'bedroom', *cuarto de baño* 'bathroom', *ventana* 'window', *jardín* 'garden'.
04	**A1:** *mesa* 'table', *silla* 'chair', *cama* 'bed', *armario* 'wardrobe', *sofá* 'sofa', *televisión* 'telly/television', *sillón* 'armchair'; **A2:** *lámpara* 'lamp'.

(continued)

(continued)

C.I.	Top ten common word types with an a.i. of >0.1
05	**A1:** *agua* 'water', *cerveza* 'beer', *vino* 'wine', *leche* 'milk', *pan* 'bread'; A2: *zumo* 'juice', *patata* 'potato', *tomate* 'tomato'.
06	**A2:** *plato* 'plate', *vaso* 'glass'; **B1:** *cuchara* 'spoon', *cuchillo* 'knife', *tenedor* 'fork'.
07	**A1:** *mesa* 'table', *silla* 'chair'; **A2:** *microondas* 'microwave', *plato* 'plate'; **B1:** *frigorífico* 'fridge/refrigerator'.
08	**A1:** *mesa* 'table', *silla* 'chair', *pizarra* 'blackboard', *libro* 'book', *bolígrafo* 'ballpen'; **A2:** *tiza* 'chalk', *cuaderno* 'notebook'.
09	**A1:** *sol* 'sun', *ventana* 'window', *calor* 'heat'; **A2**: *lámpara* 'lamp', *luz* 'light'; **B1:** *fuego* 'fire'.
10	**A1:** *calle* 'street', *coche* 'car', *casa* 'house', *escuela* 'school', *tienda* 'shop', *parque* 'park', *autobús* 'bus'.
11	**A1:** *casa* 'house', *animal* 'animal', *árbol* 'tree'; **A2**: *perro* 'dog', *cerdo* 'pig', *gato* 'cat', *caballo* 'horse', *vaca* 'cow', *bosque* 'forest'.
12	**A1:** *coche* 'car', *autobús* 'bus', *tren* 'train', *avión* 'plane', *metro* 'underground', *a pie* 'walking'; **A2:** *bicicleta* 'bike/bicycle', *caballo* 'horse'; **B1:** *tranvía* 'tram'.
13	**B1:** *plantar* 'plant', *cortar* 'cut'.
14	**A2:** *perro* 'dog', *gato* 'cat', *caballo* 'horse', *pájaro* 'bird', *vaca* 'cow', *cerdo* 'pig'; **B1:** *jirafa* 'giraffe', *león* 'lion', *elefante* 'elephant'.
15	**A1:** *fútbol* 'football', *baloncesto* 'basketball', *cine* 'cinema', *bailar* 'dance', *teatro* 'theatre', *televisión* 'telly/television', *música* 'music'; **B1:** *voleibol* 'volleyball'.
16	**A1:** *profesor* 'professor', *médico* 'doctor', *bombero* 'firefighter', *enfermero* 'nurse', *policía* 'policeman'; **B1:** *doctor* 'doctor'.

C.I.	Remaining common word types with an a.i. of >0.1
01	A1: *boca* 'mouth', *diente* 'tooth'; **A2:** *pie* 'foot', *brazo* 'arm', *espalda* 'back'**; B1:** *rodilla* 'knee', *corazón* 'heart', *cuello* 'neck', *labio* 'lip', *pecho* 'chest'**.**
02	A1: *gafas* 'glasses'; A2: *bufanda* 'scarf', *bota* 'boot', *traje* 'suit', *chaqueta* 'jacket', *gorra* 'cap', *guante* 'glove'; **B1:** *cinturón* 'belt'.
03	**A1:** *puerta* 'door', *garaje* 'garage', *baño* 'bathroom'; **A2:** *suelo* 'floor', *pared* 'wall', *techo* 'ceiling'.
04	**A2:** l*avadora* 'washing machine'; **B1**: *frigorífico* 'fridge/refrigerator', *alfombra* 'carpet'.
05	**A1:** *café* 'coffee', *té* 'tea', *paella* 'paella', *carne* 'meat', *tortilla* 'omelette', *fruta* 'fruit', *verdura* 'vegetable', *bocadillo* 'sandwich', *sopa* 'soup'; **A2:** *zanahoria* 'carrot', *manzana* 'apple', *naranja* 'orange', *fresa* 'strawberry', *plátano* 'banana', *pollo* 'chicken', *jamón* 'ham', *queso* 'cheese', *mantequilla* 'butter'.
06	
07	**A2**: *lavadora* 'washing machine', *nevera* 'fridge', *fregadero* 'sink'; **B1:** *cuchillo* 'knife', *fregador* 'sink'.
08	**A1:** *lápiz* 'pencil', *ordenador* 'computer', *ventana* 'window', *puerta* 'door', *goma de borrar* 'rubber eraser'**; B2:** *borrador* 'chalk eraser'.
09	
10	**A1**: *cine* 'cinema', *hospital* 'hospital', *árbol* 'tree', *teatro* 'theatre', *supermercado* 'supermarket', *centro comercial* 'shopping centre', *restaurante* 'restaurant'; **A2:** *edificio* 'building'**; B1**: *tranvía* 'tram'; *gente* 'people'.

(continued)

(continued)

C.I.	Remaining common word types with an a.i. of >0.1
11	**A1:** *río* 'river', *flor* 'flower', *iglesia* 'church'; **A2:** *pollo* 'chicken'; **B1:** *gallina* 'hen', *lago* 'lake'.
12	**A1:** *taxi* 'taxi', *barco* 'boat'; **A2:** *motocicleta* 'motorcycle'.
13	
14	**A2:** *pez* 'fish', *pollo* 'chicken'; **B1:** *tigre* 'tiger', *oveja* 'sheep', *gallina* 'hen', *gallo* 'cock', *conejo* 'rabbit', *cocodrilo* 'crocodile', *cebra* 'zebra'; *mariposa* 'butterfly'.
15	**A1:** *discoteca* 'disco', *ordenador* 'computer', *tenis* 'tennis', *cantar* 'sing'.
16	**A1:** *abogado* 'lawyer', *actor* 'actor', *actriz* 'actress', *director* 'director', *cantante* 'singer', *camarero* 'waiter'; **A2:** *peluquero* 'hairdresser', *cocinero* 'cook'; **B1**: *jardinero* 'gardener'.

C.I.	Word types with an a.i. of >0.1 specific to Gimnazjum
01	**A2:** *cara* 'face'; **B1:** *lengua* 'tongue'.
02	**B1:** *ropa interior* **'**underwear**'.**
03	**A2:** *balcón* 'balcony'; **B1**: *servicio* 'toilet'**.**
04	**A1:** *ordenador* 'computer', *radio* 'radio'; **A2:** *fregadero* 'sink'.
05	**A1:** *vodka* 'vodka', *coca-cola* 'coca-cola', *pizza* 'pizza'; **A2:** *pasta* 'pasta', *chocolate* 'chocolate', *helado* 'ice cream', *lechuga* 'lettuce'.
06	**A2:** *botella* 'bottle'; **C1:** *cucharilla* 'teaspoon'.
07	**A1:** *armario* 'cupboard'.
08	**A1:** *clase* 'classroom', *armario* 'cupboard', *mapa* 'map', *televisión* 'telly/television', *flor* 'flower'; **A2**: *mochila* 'backpack'.
09	**A2**: *bufanda* 'scarf';**B2:** *candela* 'candle'.
10	**A1:** *iglesia* 'church', *bar* 'bar', *discoteca* 'disco'; **A2:** *ayuntamiento* 'city hall', *piscina* 'swimming pool', *monumento* 'monument', *bicicleta* 'bike/bicycle', *panadería* 'bakery'; **B1:** *farmacia* 'pharmacy'.
11	**A1:** *tienda* 'shop', *coche* 'car', *gente* 'people'; **A2:** *pájaro* 'bird'; **B1:** *hierba* 'grass'.
12	**A1:** *aeropuerto* 'airport', *parada* 'stop'; **A2:** *motor* 'engine'; **C2:** *aeroplano* 'airplane'.
13	**A2:** *limpiar* 'clean', *dar agua* 'give water'.
14	**A1:** *pescado* 'fish', *ratón* 'mouse'; **A2**: *mosquito* 'mosquito', *araña* 'spider', *burro* 'donkey', *lobo* 'wolf'; **B1:** *oso* 'bear', *tortuga* 'turtle'; *rata* 'rat', *hámster* 'hamster'.
15	**A1:** *nadar* 'swim', *correr* 'run', *comer* 'eat', *fiesta* 'party', *beber* 'drink'; **B1:** *pelota* 'ball', *balonmano* 'handball', *balonvolea* 'volleyball'**.**
16	**A1:** *ingeniero* 'engineer'; **A2**: *periodista* 'journalist', *arquitecto* 'architect', *futbolista* 'footballer', *mecánico* 'mechanic', *piloto* 'pilot'; **B1:** *dentista* 'dentist', *deportista* 'sportman', *economista* 'economist'.

C.I.	Word types with an a.i. of >0.1 specific to Year 0
01	**B1:** *codo* 'elbow', *hombro* 'shoulder'; **B2:** *cerebro* 'brain', *uña* 'nail', *culo* 'ass'.
02	**A2:** *abrigo* 'coat', *gorro* 'cap', *sujetador* 'bra'; **B1:** *algodón* 'cotton', *lana* 'wool'.
03	**A1:** *escalera* 'stairs', *terraza* 'terrace'; **A2:** *pasillo* 'corridor'; **B2:** *comedor* 'dining room'.
04	**A1:** *estantería* 'shelf', *baño* 'bath'; **A2:** *nevera* 'fridge'; **B1:** *mesilla de noche* 'bedside table', *escritorio* 'desk'; **C1:** *televisor* 'tv set'.
05	**A1:** *pescado* 'fish', *arroz* 'rice', *huevo* 'egg', *pepino* 'cucumber'; **A2**: *alcohol* 'alcohol', *sal* 'salt'; **B1:** *limón* 'lemon'.
06	**A2:** *taza* 'cup', *servilleta* 'napkin'; **B2:** *olla* 'pot'.

(continued)

(continued)

C.I.	Word types with an a.i. of >0.1 specific to Year 0
07	**A2:** *lavaplatos* 'dishwasher', *cocina* 'kitchen'; **B1**: *horno* 'oven', *sartén* 'pan', *cuchara* 'spoon', *grifo* 'tap', *tenedor* 'fork', *lavavajillas* 'dishwasher'; **B2:** *olla* 'pot'.
08	**A1:** *papel* 'paper'; **A2:** *lámpara* 'lamp'; **B1:** *escritorio* 'desk'.
09	**A1:** *aire acondicionado* 'air conditioning'; **A2:** *chimenea* 'fireplace'; **B2:** *bombilla* 'light bulb'.
10	**A1:** *metro* 'underground', *tren* 'train', *parada de autobús* 'bus stop', *carretera* 'road', *museo* 'museum', *banco* 'bank'; **B1:** *semáforo* 'traffic lights'.
11	**A1:** *campo* 'field'; **A2:** *jardín* 'garden'; **B1:** *gallo* 'cock', *oveja* 'sheep'; **B2:** *campesino* 'farmer'.
12	**B1:** *camión* 'truck'.
13	**A1:** *agua* 'water', *planta* 'plant'; **A2:** *jardinero* 'gardener'; **B1:** *regar* 'watering', *fregar* 'wash', *tierra* 'soil'.
14	**A2:** *pato* 'duck'; **B1:** *ave* 'bird', *mono* 'monkey', *serpiente* 'snake'; *cisne* 'swan'.
15	**A1:** *libro* 'book', *leer* 'read', *bar* 'bar', *concierto* 'concert', *película* 'film'; **A2:** *pasear* 'walk', *dormir* 'sleep'; **B1:** *ir de copas* 'go out drinking'.
16	**A1:** *estudiante* 'student'; **A2:** *azafata* 'stewardess', *escritor* 'writer', *pintor* 'painter'; **B1**: *secretario* 'secretary'; **B2:** *maestro* 'teacher', *fontanero* 'plumber'.

References

Alba Ovalle, O. 1995. Léxico disponible de la República Dominicana. Santiago de los Caballeros: Pontificia Universidad Católica Madre y Maestra.

Arnau, J. 2001. La enseñanza de lenguas extranjeras a través de contenidos. *Congreso Internacional sobre Adquisición de Lenguas Extranjeras en Edades Tempranas,* Oviedo, 24–26 de Septiembre de 2001. http://web.educastur.princast.es/proyectos/keltic/documentos/cong/C08.pdf. Accessed 9 Feb 2012.

Baralo, M., M. Genís, and M.E. Santana. 2008. *Vocabulario, Elemental A1-A2*. Madrid: Anaya ELE.

Baralo, M., M. Genís, and Mª.E. Santana. 2009. *Vocabulario. Medio B1*. Madrid: Anaya ELE. disponibilidad léxica». Salamanca: Universidad de Salamanca. http://www.dispolex.com

Carcedo González, A. 1998. Desarrollo de la competencia léxica en español LE: análisis de cuatro fases de disponibilidad. *Pragmalingüística* 5–6: 75–94.

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). *Annales Universitatis Turkuensis, Humaniora, Ser. B, Tom*. 238.

Consejería de Educación de España en Polonia. 2005. *Diseño curricular de Año 0.* Varsovia: Consejería de Educación. http://www.mecd.gob.es/polonia/dms/consejerias-exteriores/polonia/centros/centros-curriculos/curriculum0.pdf. Accessed 26 Jul 2013.

Consejo de Europa. 2001. *Marco común europeo de referencia para las lenguas: aprendizaje, enseñanza, evaluación*. 2002. Madrid: Ministerio de Educación, Cultura y Deporte/Instituto Cervantes.

Dimitrijévic, N. 1969. *Lexical availability. A new aspect of the lexical availability of secondary school children*. Heidelberg: Julius Gross Verlag.

Echeverría, M.S. 1991. Crecimiento de la disponibilidad léxica en estudiantes chilenos de nivel básico y medio. In *La enseñanza del español como lengua materna, Actas del II Seminario*

Internacional sobre 'Aportes de la lingüística a la enseñanza del español como lengua materna', ed. Humberto López Morales, 61–78. Río Piedras: Universidad de Puerto Rico.
Ellis, R. 1992. *Second language acquisition and language pedagogy*. Clevedon: Multilingual Matters.
Equipo Club Prisma. 2008/2010. *Club Prisma. Método de español para jóvenes. A1, A2, A2-B1, B1*. Madrid: Edinumen.
Eurydice, Red europea de información sobre educación. 2006. *Aprendizaje Integrado de Contenidos y Lenguas AICLE en el contexto escolar europeo*. Madrid: Ministerio de Educación y Ciencia.
Gallardo del Puerto, F. 2007. On the effectiveness of early foreign language instruction in school contexts. In *Fremdsprachenkompetenzen für ein wachsendes Europa. Das Leitziel "Multiliteralität"*, ed. D. Elsner, L. Küster, and B. Viebrock, 215–229. Frankfurt/Main: Peter Lang.
García Mayo, M.P., and M.L. García Lecumberri (eds.). 2003. *Age and the acquisition of English as a foreign language*. Clevedon: Multilingual Matters.
Germany, P., and N. Cartes. 2000. Léxico disponible en inglés como segunda lengua en instrucción formalizada. In *Estudios pedagógicos*, vol. 26, 39–50. Valdivia: Universidad Austral de Chile.
Gougenheim, G., R. Michéa, P. Rivenc, and A. Sauvageot. 1956. *L'elaboration du français élémentaire. Étude sur l'établissement d'un vocabulaire et d'une grammaire de base*. París: Didier.
Hernández Muñoz, N. 2006. *Hacia una teoría cognitiva integrada de la disponibilidad léxica: El léxico disponible de los estudiantes castellano-manchegos*. Salamanca: Ediciones Universidad de Salamanca.
Higueras García, M. 2008. *Estudio de disponibilidad léxica en 43 estudiantes de ELE*, Memoria de Máster en enseñanza de Español como Lengua Extranjera. Madrid: Ediciones Universidad Antonio de Nebrija.
Instituto Cervantes. 2007. *Niveles de referencia para el español. Plan curricular del Instituto Cervantes. A1-A2; B1-B2; C1-C2,* 3 vols. Madrid: Edelsa.
Jiménez Catalán, R.M., and J. Ojeda Alba. 2009. Girls' and boys' lexical availability in EFL. *ITL International Journal of Applied Linguistics* 158: 57–76.
Jiménez Catalán, R.M., and J. Ojeda Alba. 2010. Disponibilidad léxica en inglés como lengua extranjera. *Lenguaje y Textos* 30: 167–176.
Lightbown, P.M. 2008. Easy as Pie? Children learning languages. *COPAL Concordia Working Papers in Applied Linguistics* 1: 5–29.
López Chávez, J. 1993. *El léxico disponible de escolares mexicanos*. México: Alhambra Mexicana.
López Chávez, J. 1995. El léxico en estudiantes mexicanos de primaria. In *Actas del III Seminario Internacional sobre "Aportes de la lingüística a la enseñanza de la lengua materna"*, ed. Rodríguez-Fonseca y Vázquez. *Revista de Adquisición de la Lengua Española.*
López González, A.M. 2010. La evaluación del desarrollo de la competencia léxica en L2 por medio de la disponibilidad léxica. *Revista RedELE* 18.
López Morales, H. 1973. *Disponibilidad léxica en escolares de San Juan*. Ms.
López Morales, H. 1978. Frecuencia léxica, disponibilidad y programación curricular. In *Aportes de la Lingüística a la Enseñanza del Español como Lengua Materna*, BAPLE, Vol. 6, ed. Humberto López Morales, 73–86.
Mackey, W.F. 1971. *Le vocabulaire disponible du français*. 2 tomos. Paris/Bruxelles/Montréal: Didier.
Mena Osorio, M. 1986. *Disponibilidad léxica infantil en tres niveles de enseñanza básica*. Concepción: Universidad de Concepción.
Michéa, R. 1950. Vocabulaire et culture. *Les langues modernes* 44: 187–192.
Michéa, R. 1953. Mots fréquents et mots disponibles. Un aspect noveau de la statistique du langage. *Les langues modernes* 47: 338–344.
Muñoz, C. (ed.). 2006. *Age and the rate of foreign language learning*. Clevedon: Multilingual Matters.
Muñoz, C. 2008. Symmetries and asymmetries of age effects in naturalistic and instructed L2 learning. *Applied Linguistics* 294: 578–596.
Pottier, B. 1971. *Gramática del español*. Madrid: Alcalá.

Román Morales, B. 1985. *Disponibilidad léxica en escolares de Dorado*. Río Piedras: Universidad de Puerto Rico.
Samper Hernández, M. 2002. *Disponibilidad léxica en alumnos de español como lengua extranjera*, Colección Monografías, vol. 4. Málaga: ASELE.
Samper Padilla, J.A. 1998. Criterios de edición del léxico disponible. *Lingüística* 10: 311–333.
Samper Padilla, J.A. 1999. Léxico disponible y variación dialectal: Datos de Puerto Rico y Gran Canaria. In Estudios de Lingüística hispánica (Homenaje a María Vaquero), ed. A. Morales et al., 551–573. San Juan de Puerto Rico: Universidad de Puerto Rico.
Samper Padilla, J.A., et al. 2003. El proyecto de estudio de la disponibilidad léxica en español. In *Pautas y pistas en el análisis del léxico hispanoamericano*, 27–140. Madrid/Frankfurt: Iberoamericana/Vervuert.
Samper Padilla, J.A., and M. Samper Hernández. 2006. Aportaciones recientes de los estudios de disponibilidad léxica. *LynX: Panorámica de Estudios Lingüísticos.* 5: 5–95.
Tatoj, C., et al. 2008. *Raport ewaluacyjny – Sekcje dwujęzyczne z językiem hiszpańskim w Polsce*. Varsovia: MEN-CODN.

Chapter 10
Cognitive Factors of Lexical Availability in a Second Language

Natividad Hernández Muñoz, Cristina Izura, and Carmela Tomé

10.1 Introduction

The study of the cognitive aspects of lexical availability, understood as the ease with which words are produced as exemplars of a given semantic category (e.g. 'furniture' or 'body parts') has received some attention over the past two decades (Cañizal Arévalo 1991; Ferreira and Echeverría 2010; Hernández Muñoz 2010; Hernández Muñoz et al. 2006; Romero Rubilar 2000; Urrutia 2003). These studies have been based on the assumption that lexical availability in addition to revealing the type, size and richness of the vocabulary used by a community of speakers, also provides evidence of the performance of mental processes.

This assumption is supported by the fact that the same task used to study lexical availability, the semantic or category fluency task, is also extensively used in psycholinguistic research to investigate semantic memory in healthy populations and brain injured patients (Caramazza 1998; Kempler et al. 1998; Monsch et al. 1994; Ostrosky-Solis et al. 2007; Tröster et al. 1989; Warrington and McCarthy 1987). In addition, the task is also commonly used in applied linguistics to assess language learning and recall in the first (Grabowski 2005, 2007) and the second language (Grogan et al. 2009; Luo et al. 2010). This convergence of cognitive, psychological and linguistic studies does not end in the common employment of a particular task but is also apparent in the multidisciplinary use of the normative databases developed from the task and many of the scientific questions that are generated (for the Spanish

N. Hernández Muñoz (✉) • C. Tomé
Departamento de Lengua Española, Universidad de Salamanca,
Plaza de Anaya, 137008 Salamanca, Spain
e-mail: natih@usal.es; carmela_tome@siol.ne

C. Izura
Department of Psychology, University of Swansea, Singleton Park, Swansea SA28PP, UK
e-mail: C.Izura@swansea.ac.uk

R.M. Jiménez Catalán (ed.), *Lexical Availability in English and Spanish as a Second Language*, Educational Linguistics 17, DOI 10.1007/978-94-007-7158-1_10,

language: Álamo et al. 1999; Butman et al. 2000; Izura et al. 2005; Soto et al. 1982). One of such questions whose investigation has been assisted by the results of tasks such as semantic fluency and also word association (i.e. discrete and continuous) is the potential way in which concepts might be organised in a metaphoric cognitive space. Consequently, a number of semantic network models, often used to explain linguistic behaviour, have been developed (Borge-Holthoefer and Arenas 2010; Collins and Loftus 1975; Collins and Quillian 1969; De Deyne and Storms 2008; Steyvers and Tenenbaum 2005).

The present study is also the result of a cross-disciplinary investigation of the lexical availability of Spanish as a second language. The aims of this investigation were twofold. Firstly, to explore the extent to which the availability of a word (e.g., *car*) in response to a given category (e.g. 'transport') is influenced by the same cognitive factors irrespective of whether the word is learned as part of the first (L1) or the second language (L2). Secondly, to investigate whether the lexical availability of words learned in an L2 (e.g. *trompeta*) are influenced by similar words from L1 (e.g., *trumpet*).

The measure of lexical availability combines two factors: one, the frequency at which a word (e.g., *dog*) is produced as a member of a semantic category (e.g. 'animals') and two, the position at which the word occurs in the lists of words provided by a group of individuals (López Chávez and Strassburger Frías 2000). Therefore, lexical availability reflects the rate of occurrence plus the distance from the category label. Based on these measures of lexical availability, Hernández Muñoz et al. (2006) investigated the extent to which lexical availability was influenced by six factors: age of word learning, familiarity of the concept, ease with which the word evokes an image, typicality, word frequency and word length. The study comprised 500 words from five semantic categories (i.e. 'animals', 'body parts', 'clothing', 'furniture' and 'intelligence'). They found that those words that were typical examples of a category, conceptually familiar and early acquired, were significantly more available than atypical examples, conceptually unfamiliar and late acquired words. In contrast, word frequency, word length and imageability did not influence lexical availability. They also looked at whether categories varied in relation to each of the factors under study and showed, among other things, that the category 'animals' comprised words that were learned earlier than the words belonging to any other category.

The lack of word frequency effects on lexical availability observed in Hernández Muñoz et al. (2006) supported Michèa's (1953) original suggestion that the available vocabulary is not necessarily, or exclusively, the most frequent vocabulary. Indeed, research into lexical availability emerged as an attempt to mitigate the influence that printed frequency indexes had on the generation of dictionaries and texts for teaching second languages. Hernández Muñoz et al. (2006) noted that, with reference to the selection of words to be included in dictionaries and teaching material, it is important to consider the fact that lexical availability is not a fixed but a variable property of words. Thus, the availability of a word changes in relation to its semantic category (e.g., *coche* (car) is highly available as a 'transport', intermediately available as part of 'the city' and has a very low availability as part of the countryside'). This has two significant implications: first, the vocabulary to be considered for teaching or for

dictionaries should be framed within a semantic context. Second, words with low lexical availability in a given category are at risk of being excluded from that particular context in dictionaries and teaching material leading to differences between the lexical availability of L1 and L2.

The similarities and differences between lexical availability in L1 and L2 have been recently addressed in a study in which the semantic relations of L1 and L2 words were investigated (Ferreira and Echeverría 2010). Ferreira and Echeverría (2010) used lexical availability lists, obtained from the completion of a category fluency task, to create semantic networks for the words comprised in the lists. The authors looked at the number and strength of the semantic relations based on the frequency with which two words were produced consecutively. They showed that native speakers produced a higher number of strong semantic relations. The words provided by second language learners had more disperse relations. In addition, native speakers had a tendency to produce words in semantic clusters, indicating the existence of highly specific sub-categories. These clusters, however, were not found in the data collected from second language learners.

Another potential source of differences between L1 and L2 lexical availability is order of word learning, often called age of acquisition (AoA). Hernández Muñoz et al. (2006) showed that the age of acquisition was a significant predictor of the lexical availability of Spanish as a first language, implying that those words learned earlier were more available than words learned later. However, Izura and Ellis (2002, 2004) showed that there are important differences between AoA in L1 and AoA in L2. They argued that when L2 is learned after childhood, the order of word learning does not exactly mirror that of the L1. Thus, part of the vocabulary of children consists of words related to toys, stories and daily routines (e.g., *doll, fairy, dragon, dummy, nappy,* etc.). These terms, however, are not relevant to the adult who has started learning an L2 and therefore are acquired relatively late. Similarly, adults learning an L2 master as soon as they can words related to surviving in a foreign country, such as finding places or organising their money (e.g. *address, expensive, cash point*). These terms are naturally excluded from children's vocabularies and are therefore learned later in L1 than in L2. Izura and Ellis (2002, 2004) showed that L2 AoA had an influence in lexical tasks performed in L2 (e.g., picture naming and lexical decision tasks) and that L1 had an influence in the same tasks completed in L1. However, L1 age of acquisition did not affect lexical identification of words in L2 (e.g., speakers of English as an L2 were faster at identifying words such as *address,* learned early in L2 but late in L1, than at responding to words such as *fairy,* learned late in L2 but early in L1).

In the current study, the factors influencing the lexical availability of Spanish as an L2 were explored and compared to the factors that affect lexical availability in L1. AoA was explored in some detail as a potential source of differences between L1 and L2 lexical availability. It was predicted that typicality could have less impact in L2 than in L1 lexical availability as Ferreira and Echeverría (2010) showed that the semantic relations and specificity of the semantic categories in L2 were more scattered than in L1. In relation to AoA, it was predicted that it will affect L2 as well as L1 but only the AoA specific to the corresponding language (i.e. L1 or L2). Finally, following the Revised Hierarchical Model (RHM, Kroll and Stewart 1994),

the possibility of L2 lexical availability being influenced by L1 was explored. The RHM is a theory of L2 development that proposes that at the early stages of L2 learning, L2 words are highly dependent of their translation equivalents in L1. In addition, the model claims that the connections between L2 words and their meanings are weak. According to the RHM, when a learner of Spanish as an L2 has to understand the newly learned word '*cansado*', s/he accesses its translation equivalent '*tired*' first and then via the L1 word activates the meaning. If this is the case, L2 learners would have a tendency to produce a high number of cognates in a category fluency task. This is because cognates are words whose form in L2 and L1 is very similar (e.g., *banana, pear* and *melon* in English are *banana, pera,* and *melón* in Spanish) therefore when the participant has to produce as many words as s/he can that fit in the category *fruit*, the lexical representations in L1 (e.g. *banana, pear, melon*) will activate, with relative ease, its translation equivalent in L2 (e.g. *banana, pera, melón*) via their lexical links. It was predicted, therefore, that L2 lexical availability will be affected by the degree of cognateness between L1 and L2 words.

10.2 Method

10.2.1 Participants

Forty-three (11 males and 32 females) students of Spanish, as an L2, completed the lexical availability task. Participants were volunteer American-English native speakers with a mean age of 20 years (range 17–23 years of age). All were enrolled in intermediate level Spanish courses (B1 and B2) programmed by the International Courses of the University of Salamanca. At the end of the lexical availability task, participants were asked to estimate their level of proficiency in a seven-point scale with 1 meaning a very low proficiency, and 7 as good as a native speaker. Their average score of 4.47 was used to index the participants' proficiency in the B2 level from the six levels established by the European Common Frame for languages: A1, A2, B1, B2, C1 and C2 (Cervantes Institute 2002).

10.2.2 Materials and Procedure

Lexical availability data in Spanish as L2 was gathered from four of the five categories used by Hernández Muñoz et al. (2006). These were the supra-ordinate categories: 'Animals', 'Body parts', 'Clothing' and 'Furniture'.

Categories were presented in booklets where each category name appeared at the top of the page and it was followed by a series of eight columns. Participants were asked to start providing words belonging to the given category in the first column and to wait until they hear a beep sound before they could change to the next column to

write down more words. Thirty seconds were allowed for each column. Therefore, 4 min were dedicated to each category. The order at which categories were presented was counterbalanced across participants. Each session had two parts: first, participants were asked to provide as many words for the given category as possible, and then they were asked to answer a questionnaire of a sociological nature. This latter section was completed in the participants' L1: English.

The lexical availability of each word produced was calculated using the formula of López Chávez and Strassburger Frías (1987, 1991, 2000) which computes an availability value based on the positions of a word in the lists, the number of participants who placed the word at those positions and the lowest position in which the word has been placed in any of the lists.

10.2.3 Predictor Variables

For the purpose of this study, only the words provided in the first 2 min were considered for further study (i.e. a total of 289 words). This was done in order to follow the same procedure as that used by Hernández Muñoz et al. (2006). Values for seven predictor variables were obtained. These were age of acquisition, cognateness, concept familiarity, imageability, typicality, word frequency and word length. A description of each variable is stated below. Variables appear in alphabetical order.

Age of Acquisition

L2 Age of acquisition ratings were obtained from 24 American-English native speakers (8 males, 16 females) with a mean age of 20 years (range 17–25 years). Their average proficiency in Spanish as L2 was 5.1 as self-estimated in a seven-point scale. They were asked to estimate the time at which they considered they learned the 289 words in an eight-point scale where 1 meant learnt during the first year of learning Spanish, 2 learnt during the second year, 3 learnt during the third or fourth years, 4 learnt during the fifth or sixth years, 5 learnt during the seventh or eighth years, 6 learnt during the ninth or tenth years, 7 learnt during the eleventh year or later, and finally 8 was the option to state that the word had never been learnt. Words were presented randomly without reference to the semantic category where they were coming from. Reliability was 0.73.
L1 Age of acquisition measures were obtained from Hernández Muñoz et al. (2006).

Cognateness

Twelve American-English native speakers (3 males, 9 females) with a mean age of 19 years (range 17–29 years) were asked to rate the similarity between the 289 Spanish words and their English translation equivalents in a seven-point scale where

1 meant highly dissimilar and 7 meant highly similar. Participants' average proficiency in Spanish as L2 was 4.96 as self-estimated in a seven-point scale. This technique had been previously employed in a number of studies (De Groot and Nas 1991; Lotto and De Groot 1998; Tokowicz et al. 2002). Word pairs were presented randomly without reference to the semantic category of origin. Reliability was 0.95.

Concept Familiarity

Ratings were obtained from 19 American-English native speakers (8 males, 11 females) with a mean age of 20 years (range 17–27 years). Their average proficiency in Spanish as L2 was 4.86 as self-estimated in a seven-point scale. Participants were asked to rate the list of 289 words according to how often they come in contact with the meaning of the word. The examples of *key* and a *crown* were provided; the former was very familiar and the latter was less familiar. Each word was measured in an eight-point scale where 1 meant less than once a month, 2 once a month and so on up to 7 meaning many times a day. The eighth option (0) allowed participants to indicate that they did not know the word. Words were presented randomly without reference to the semantic category of origin. Reliability was 0.88.

Imageability

Ratings were obtained from 20 American-English native speakers (6 males, 14 females) with a mean age of 20 years (range 17–28 years). Their average proficiency in Spanish as L2 was 4.77 as self-estimated in a seven-point scale. Participants were asked to estimate how easy it was to evoke a mental image for each of the 289 words. The examples of *apple* and *intelligence* were provided; the former was highly imageable but it was difficult to imagine the second. Each word was measured using an eight-point scale where 1 corresponded to very difficult to evoke a mental image and 7 denoted words from which it was very easy to evoke a mental image. The eighth option (0) allowed participants to indicate that they did not know the word. Words were presented randomly without reference to the semantic category of origin. Reliability was 0.84.

Typicality

Ratings were obtained from 25 American-English native speakers (9 males, 16 females) with a mean age of 20 years (range 17–27 years). Their average proficiency in Spanish as L2 was 4.76 as self-estimated in a seven-point scale. Participants were presented with the 289 words separated into four lists. These lists corresponded to the four categories: 'Animals', 'Body parts', 'Clothing' and 'Furniture'. Participants were asked to rate each word as how representative of the given category was. The words of *apple* and *coconut* were given as typical and atypical examples of the category 'Fruit'. Each word was rated using an eight-point scale where 1 meant a very bad example of the category and 7 meant a very good example of the category.

The eighth option (0) allowed participants to indicate that they did not know the word. Reliability was 0.91.

Word Frequency

Four different frequency measures were considered in this study. The printed frequency measure commonly used in Spanish studies along with two new frequency databases recently generated a better reflection of the spoken language (i.e. oral frequency) or a more reliable estimation given the size of the corpus in which the database was based (i.e. frequency from subtitles). A measure of rated frequency was also collected.

(a) ***Frequency from subtitles*** was taken from Cuetos et al. 2011) – a corpus of Spanish frequency based on subtitles taken from present-day movies and television series (broadcasted between 1990 and 2009). The corpus comprises over 94,000 words and is based on a total of 41 million words.
(b) ***Oral frequency*** values were taken from Alonso et al. 2011) – a corpus comprising 67,979 Spanish words based on over three million words extracted from 913 oral recordings from a variety of situations (e.g., radio and television programmes).
(c) ***Printed frequency*** values were taken from Alameda and Cuetos (1995) which is based on a corpus of written Spanish texts comprising two million words from 606 texts varying between novels, essays, newspapers and scientific populations.
(d) ***Rated frequency*** was obtained from 13 American-English native speakers (5 males, 8 females) with a mean age of 21 years (range 17–32 years). Their average proficiency in Spanish as L2 was 5.13 as self-estimated in a seven-point scale. Participants were asked to estimate how often they come into contact with each of the 289 words. Each word was measured using an eight-point scale where 1 meant very infrequently and 7 meant very frequently. The eighth option (0) allowed participants to indicate that they did not know the word. Words were presented randomly without reference to the semantic category of origin. Reliability was 0.79.

Word Length

It was measured as the number of syllables in each word as in Hernández Muñoz et al. (2006).

10.3 Results

A different number of words were produced per category and all of them were considered and submitted to analyses. To ensure that the variables were normally distributed and met the assumptions for subsequent statistical analysis, those with high indices

Table 10.1 Means (*M*), standard deviations (*SD*) and ranges of all the variables considered in the study for each of the categories under investigation

Predictor variables		'Animals' N = 85	'Body parts' N = 63	'Clothing' N = 73	'Furniture' N = 75
Age of acquisition L2)	M	2.38	2.35	2.24	2.08
	SD	0.54	0.66	0.57	0.60
	Range	1.04–3.5	1.24–3.82	1.12–3.44	1.08–3.52
Log. Cognateness	M	0.63	0.56	0.60	0.59
	SD	0.19	0.15	0.17	0.16
	Range	0.39–0.89	0.39–0.89	0.39–0.89	0.39–0.89
Concept familiarity	M	2.50	3.87	3.84	4.39
	SD	0.96	0.94	0.97	0.94
	Range	1.25–6.17	1.90–5.46	1.46–6	1.89–6.08
Imageability	M	6.28	5.91	6.18	6.15
	SD	0.52	0.72	0.60	0.62
	Range	4.19–7	3.80–6.84	4.25–7	4–7
Typicality	M	6.16	6.22	5.39	4.21
	SD	1.01	0.72	1.28	1.45
	Range	2.46–6.84	2.52–6.79	2.21–6.68	2.16–6.84
Log. Frequency-subtitles	M	1.01	1.34	1.26	1.39
	SD	0.51	0.63	0.55	0.67
	Range	0.05–2.48	0.09–2.56	0.07–2.23	0.09–3.14
Log. Frequency-oral	M	0.84	1.11	1	1.21
	SD	0.54	0.61	0.47	0.68
	Range	0.12–2.67	0.12–2.27	0.12–1.81	0.12–3.09
Log. Frequency-printed	M	1.17	1.73	1.34	1.57
	SD	0.60	0.61	0.64	0.68
	Range	0–2.68	0.40–2.96	0–2.40	0–3.04
Rated frequency	M	3.74	4.29	4.40	4.78
	SD	1.05	1.21	0.99	1.97
	Range	1.6–6.68	1.73–6.42	1.89–6.60	1.92–6.56
Word length	M	0.56	0.53	0.61	0.59
	SD	0.09	0.09	0.15	0.12
	Range	0.30–0.78	0.30–0.70	0.30–1	0.30–0.95
Sqrt-lexical availability*100	M	21.20	28.32	23.03	17.04
	SD	17.89	22.92	20.44	16.74
	Range	0–92.2	0–80.62	0–87.75	0–86.6

Note: *Log* logarithm, *Sqrt*. Square root

of skewness and kurtosis (i.e. absolute values above 3.29) were transformed. Thus, a logarithm transformation was applied to word frequency, syllable length and cognateness. Age of acquisition, concept familiarity and rated frequency were normally distributed. Imageability and typicality showed high indices of skew and kurtosis; however, square root or logarithm transformations did not improve their distribution. Therefore, the raw measures were maintained. Lexical availability was transformed using the square root formula. The resulting lexical availability values were then multiplied by 100 for ease of presentation. Table 10.1 shows the number

of words per category along with the means, standard deviations and ranges for each variable in all four categories.

Three different analyses were carried out. First, a series of ANOVAs were calculated to investigate differences across categories in relation to the variables of interest. Second, four multilevel regression analyses were computed in order to assess the influence of the predictor variables in lexical availability of Spanish as an L2. Third, the potential influence of the age of acquisition of the first language was assessed.

10.3.1 Analysis 1

As in Hernández Muñoz et al. (2006), a series of one-way analyses of variance were conducted to assess whether the scores of the words per variable differed across categories. There were significant differences among the categories on all of the predictor variables apart from cognateness. These were analysed further by using post-hoc tests (Tukey's HSD) to compare the categories pairwise on each of the factors (see Fig. 10.1).

Age of acquisition: Categories differed in the age of acquisition ratings given to the words produced within them (F (3, 294) = 3.94, MSE = 1.38, $p < 0.01$). Post-hoc tests showed that age of acquisition ratings were significantly lower for the category 'Furniture' (M = 2.08, SD = 0.6) than for the 'Body parts' category (M = 2.35, SD = 0.66, $p < 0.05$) and the 'Animals' category (M = 2.38, SD = 0.54, $p < 0.01$). This means that overall words that are related to the category 'Furniture' are learned earlier than those corresponding to the 'Body parts' and 'Animals' categories.

Concept familiarity: Categories differed in the familiarity ratings given to the words within them (F (3, 294), MSE = 52.72, $p < 0.001$). Post-hoc tests showed that familiarity ratings were significantly higher for words in the category 'Furniture' (M = 4.39, SD = 0.94) than for 'Clothing' (M = 3.84, SD = 0.95, $p < 0.01$), 'Body parts' (M = 3.87, SD = 0.94, $p < 0.01$) and 'Animals' (M = 2.50, SD = 0.96, $p < 0.001$), which were significantly less familiar than 'Clothing' ($p < 0.001$) and 'Body parts' $p < 0.001$). This implies that 'Furniture' is the category that produced most familiar terms.

Imageability: Categories differed significantly on imageability F (3, 294) = 4.58, MSE = 0.007, $p < 0.01$). Post-hoc tests showed that items in the category "Body parts" M = 5.91, SD = 0.72) were rated as significantly less imageable than 'Clothing' M = 6.18, SD = 0.60, p = 0.05) and 'Animals' M = 6.28, SD = 0.62, p < 0.01).

Typicality: Categories differed significantly on typicality (F(3, 294) = 43.71, MSE = 0.377, $p < 0.001$). Post-hoc tests showed that typicality ratings were significantly lower for 'Furniture' (M = 4.21, SD = 1.45), than 'Body parts' (M = 6.22, SD = 0.72, $p < 0.001$), 'Animals' (M = 6.16, SD = 1.01, $p < 0.001$) and 'Clothing' (M = 5.39, SD 1.28, $p < 0.001$), which in turn had significantly lower ratings than

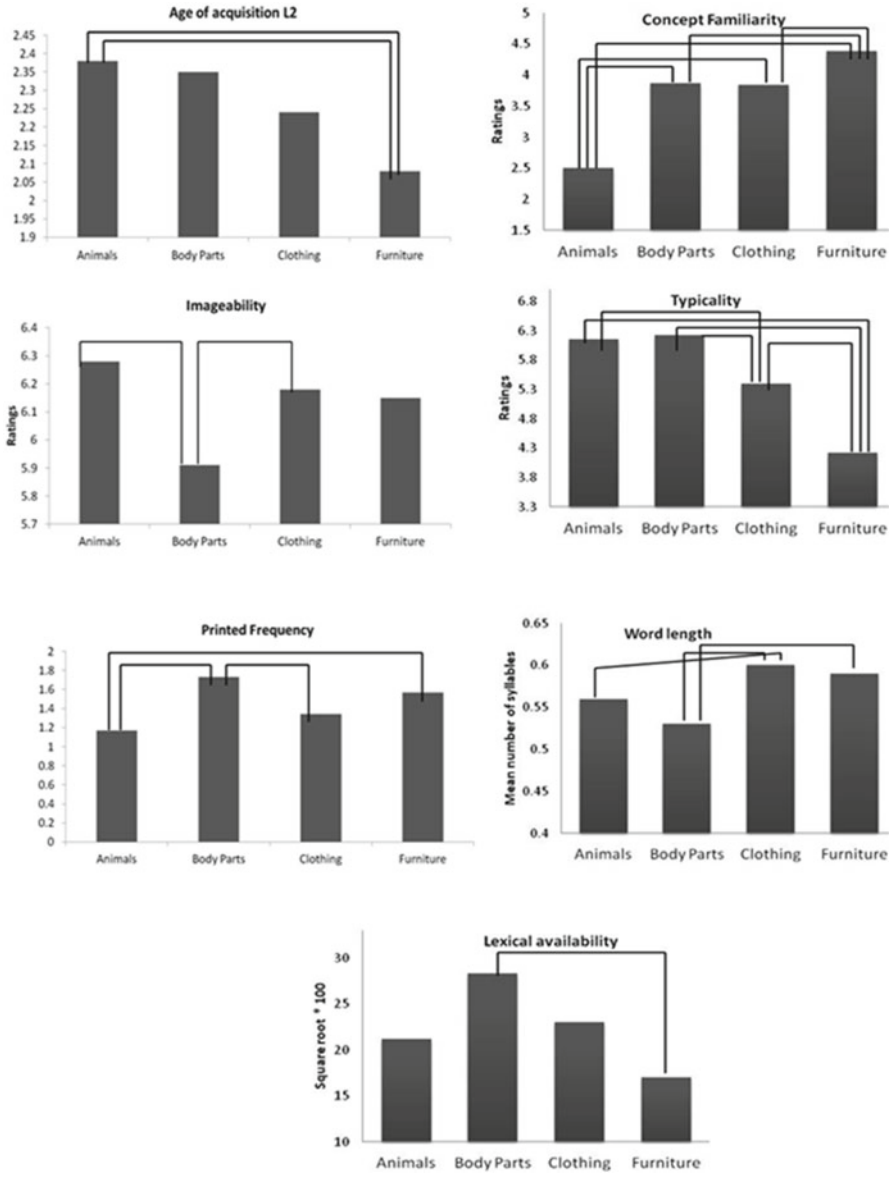


Fig. 10.1 Bar graphs for each variable showing significant differences across categories (Note: only one graph for word frequency is shown i.e. printed frequency)

'Body parts' ($p < 0.001$) and 'Animals' ($p < 0.001$). This is similar to what was found in Spanish as L1 Hernández Muñoz et al. (2006) and implies that 'Furrniture' and 'Clothing' are more widely dispersed categories in which items tend to stay further from the 'core' of the category than is the case for 'Animals' and 'Body parts'.

Word frequencies from existing databases: Categories differed significantly on the three measures of word frequency obtained from three different databases. Oral frequency (F (3, 273) = 5.30, MSE = 1.78, p < 0.01); frequencies from subtitles F (3, 286) = 6.30, MSE = 2.18, p < 0.001); and frequency from printed material (F (3, 294) = 11.24, MSE = 4.51, p < 0.001). Post-hoc tests showed that words in the category 'Animals' M (oral) = 0.84; M (subtitles) = 1.01; M (printed) = 1.17) had significantly lower frequencies than 'Body parts' M (oral) = 1.11, p < 0.05; M (subtitles) = 1.34, p < 0.01; M (printed) = 1.73, p < 0.001), and 'Furniture' M (oral) = 1.21, p < 0.05; M (subtitles) = 1.39, p < 0.01; M (printed) = 1.57, p < 0.001) across all frequency measures. The category 'Animals' was also significantly lower than the category 'Clothing' (M = 1.34, p < 0.05), for the frequency from subtitles, and the category clothing (M = 1.34) had significantly lower written frequencies than the category 'Body parts' p < 0.01).

Rated frequency: Categories differed significantly on rated frequency (F (3, 294) = 11.74, MSE = 14.70, p < 0.001). Post-hoc tests showed that rated frequency was significantly lower for the category 'Animals' (M = 3.74, SD = 1.05) than for the category 'Body parts' (M = 4.29, SD = 1.21, p < 0.01), 'Clothing' (M = 4.40, SD = 0.99, p < 0.05), and 'Furniture' (M = 4.78, SD = 1.97, p < 0.01).

Word length: Categories differed significantly on word length (F (3, 294) = 5.36, MSE = 0.72, p < 0.01). Post-hoc tests showed that 'Clothing' (M = 0.61, SD = 0.15) had significantly longer names than 'Animals' (M = 0.56, SD = 0.09, p < 0.05), and 'Body parts' (M = 0.53, SD = 0.09, p < 0.01), which in turn had shorter names than the category 'Furniture' (M = 0.59, SD = 0.12, p < 0.05).

Lexical availability: The categories also differed overall on lexical availability (F (3, 294) = 3.94, MSE = 1484.96, p < 0.01). Post-hoc tests showed that words from the 'Body parts' category (M = 28.31, SD = 22.92) had significantly higher lexical availabilities than words from the 'Furniture' category (M = 17.04, SD = 16.74, p < 0.01). This implies that the words generated for the category 'Furniture' were much more variable across participants than the words generated for the category 'Body parts' (Fig. 10.1).

10.3.2 Analysis 2

The correlations of the predictor variables with each other and with lexical availability across the combined categories for all the words used in the analysis are shown in Table 10.2. Only those correlations that were significant at the 0.05 level or below are shown. Among the predictors, *age of acquisition in L2* showed a negative correlation with concept familiarity, typicality, word frequency and lexical availability and a positive correlation with the number of syllables. This implies that in Spanish as L2 early acquired words are those that are familiar, imageable, frequent, lexically available and shorter. *Concept familiarity* correlated positively, in addition to age of acquisition in L2, with imageability, frequency and lexical availability and

Table 10.2 Significant correlations between predictor variables and lexical availability

	1	2	3	4	5	6	7	8	9	10
1. AoA L2										
2. C. Fam.	−.66*									
3. Imag.	−.59*	.24*								
4. Cogn.	ns	ns	.24*							
5. W. Length	.17*	−.14**	ns	.17*						
6. Freq. Oral	−.51*	.55*	.18*	ns	ns					
7. Freq. Subt.	−.61*	.62*	.31*	ns	ns	.84*				
8. Freq. Print	−.49*	.55*	.13**	−.22*	−.18*	.83*	.82*			
9. Freq. Rated	−.71*	.82*	.32*	−.14**	−.29*	.57*	.64*	.57*		
10. Typicality	ns	−.34*	.17*	ns	ns	−.13**	−.13**	−.12**	−.32*	
11. Lex. Ava.	−51*	.30*	.36*	ns	−.16*	.28*	.37*	.37*	.43*	.41*

Note: *Correlation is significant at the 0.01 level 2-tailed, **Correlation is significant at the 0.05 level 2-tailed, ns. not significant

AoA L1 age of acquisition of the second language, *C. Fam*. concept familiarity, *Imag.* imageability, *Cogn*. cognateness, *W. Length* word length, *Freq. Oral* frequency oral, *Freq. Subt.* frequency from subtitles, *Freq. Print* printed frequency, *Freq. Rated* rated frequency, *Lex. Ava.* lexical availability

negatively with the number of syllables and typicality. This indicates that second language learners rate words as familiar when they are early acquired, imageable, frequent, lexically available, shorter and not very typical examples of the category. *Imageability* showed significant and positive correlations with familiarity cognateness frequency, typicality, lexical availability and negatively with age of acquisition in L2. This suggests that participants rated words as imageable if they also were familiar, cognates, frequent, typical of their category, lexically available and acquired early. *Word frequency measures* were positively correlated with familiarity, imageability, lexical availability and negatively with age of acquisition in L2, typicality and cognateness. It is interesting to note that only printed and rated frequencies were inversely correlated with number of syllables with the longer words being rated and appearing in print less frequently. *Typicality* was positively correlated with imageability and lexical availability and negatively correlated with familiarity and word frequency, implying that those items judged more typical are those less familiar and less frequent but imageable and lexically available. Finally, *lexical availability* showed significant correlations with all the variables under study apart from cognateness. Lexical availability showed the highest correlation with age of acquisition in L2, followed by rated frequency and typicality. The correlations of lexical availability with concept familiarity, imageability, and word length were progressively lower but still significant.

Predictor variables were submitted to a series of multilevel regression analyses. When predictor variables are intercorrelated among themselves, results from regression analyses can be misleading due to problems of multicolinearity. This means that the shared variance between two given variables (e.g., rated frequency and age of acquisition in L2 in the present study $r = -0.71$) is regressed against the dependent variable at once without noting that the predicted variance comes from two different

Table 10.3 Standard error and t values for one of the analyses carried out on lexical availability

Step 1	Std error	T
'Body parts'	2.79	2.69*
'Clothing'	2.71	0.124
'Animals'	2.59	0.418
Step 2		
Age of acquisition L2	2.27	−6.02*
Concept familiarity	1.23	0.45
Imageability	1.62	0.47
Cognateness	0.44	−2.20**
Log frequency-subtitles	1.65	0.35
Log number of syllables	6.92	−1.04
Typicality	0.66	7.64*
	R^2	0.46

Note: *$p < 0.01$, **$p < 0.05$

sources (the regression equation assumes is the same source). Any significance would be randomly allocated to one of the predictors and not the other one.

Highly correlated variables were entered into independent regression analyses and thus four different multilevel regression analyses were conducted as the result of the combination of the four frequency measures. It is important to note that when the rated frequency was entered into the analysis, neither age of acquisition in L2 nor concept familiarity were included in the analysis.

Multilevel regression is a statistical analysis technique particularly appropriated for data that comes grouped in clusters like the semantic categories of the present study. It can be thought of as a hierarchical regression with two steps. For the present study in the first step, variance in lexical availability due to differences between the four categories on the predictor variables was extracted. For that, three dummy variables were created: 'Body parts', 'Clothing' and 'Animals' with 'Furniture' working as the reference category. In the second step, the predictor variables were entered simultaneously to determine the ability of each predictor to account for differences between words in their lexical availability after overall differences between the categories have been accounted for. The analysis explaining the greatest percentage of the variance (46 %) is shown in Table 10.3. Consistent main effects of age of acquisition in L2 such as typicality and cognateness were found across all the analyses.

10.3.3 Analysis 3

In order to examine whether the age of acquisition of L1 had an impact in the lexical availability in L2, the results from the lexical availability task from Hernández Muñoz et al. (2006) and the present study were compared.

All the words produced in L1 and L2 in response to each of the four categories considered in both studies (i.e. 'Animals', "Body parts", 'Clothing' and 'Furniture')

were extracted and submitted to a regression analysis. That was a total of 162 words: 53 for the category 'Animals', 46 words for the category 'Body parts', 35 for 'Clothing' and 28 for 'Furniture'. A multilevel regression analysis was carried out with the same predictor variables as in the previous analyses and three steps. Age of acquisition in L1 was entered in step three and it did not show a significant influence in the lexical availability of Spanish as L2 ($t = -0.637$, $p > 0.1$).

10.4 Discussion

In this study, 43 American-English native speakers learning Spanish as an L2 were asked to provide as many words as possible from four categories (i.e. 'Animals', 'Body parts', 'Clothing' and 'Furniture'). All the words elicited in the first 2 min of this category fluency task were considered for analysis. The number of words provided differed across categories with the category 'Animals' generating the largest amount of words (n = 86) and the category 'Body parts' the lowest number of words (n = 63). Estimations for age of acquisition in L2, concept familiarity, typicality, imageability, cognateness and rated frequency were gathered for all the words produced. Three different types analyses were carried out. First, the four semantic categories were compared in order to explore any differences among them in relation to the variables of interest. Thus, categories differed in relation to all the variables considered in the present study apart from cognateness. Here, we present a summary of the findings and a comparison to those reported by Hernández Muñoz et al. (2006). The category 'Furniture' comprised words acquired in L2 significantly earlier than the words related to 'Body parts' and 'Animals'. In contrast, the category 'Animals' had the earliest acquired items in L1 (Hernández Muñoz et al. 2006). As a matter of interest, at the first level (A1) in the teaching materials of Spanish as L2 (based on the Common European Framework for languages Cervantes Institute 2002), there is a relative abundance of words related to the category 'Furniture'. The four categories also differed in the frequency of the words that describe them with 'Body parts' showing the highest frequency.

This is unsurprisingly similar to what was found by Hernández Muñoz et al. (2006) since both studies are based in similar frequency databases (e.g., Alameda and Cuetos 1995). Interestingly, however, rated frequency, a measure collected in the present study only, followed the same pattern as the word frequency values measured by the databases. The length of the words varied across categories with the category 'Clothing' generating longer words in L2. However, in L1, the longest words belonged to the category 'Furniture'. This could be due to differences in the number of participants across studies. The larger the number of participants the larger the amount of words produced. Compounds such as '*short-sleeve T-shirt'* tend to be produced towards the end of the list and by few participants. This means that fewer of these compound terms would have been among the most available terms in Hernández Muñoz et al. (2006) than in the present study. The four categories also differed in terms of typicality, imageability and concept familiarity with 'Furniture' generating the less typical but more familiar items and 'Body parts' the less imageable. Similar results were found by Hernández Muñoz et al. (2006).

In the second set of analyses, the authors looked at the extent to which the selected variables contributed towards the availability of words in L2. Four multilevel regression analyses showed that age of acquisition in L2, cognateness and typicality were significant predictors of lexical availability in Spanish as L2. Two of these factors, age of acquisition and typicality, also predicted the lexical availability of terms in Spanish as L1 Hernández Muñoz et al. (2006). However, and interestingly, the influence of concept familiarity reported in L1 seems to have been substituted by cognateness in L2. This difference can be explained following the Revised Hierarchical Model RHM, (Kroll and Stewart 1994). The model proposes that L2 semantic connections are weak but that L2 words are strongly connected to their L1 translation equivalents. Thus, when L2 speakers are asked to provide as many elements of a given category as possible, they do not rely solely on their conceptual system since they might not have learned the L2 words for all the concepts they know or they might have learned them but the relationship between them might be weak. Instead, they rely more on the links concepts have with their first language. Once the first language words have been accessed, if they are in addition similar in form (i.e. cognates), are more likely to activate their L2 translation equivalents and therefore appear sooner and more often in the category fluency task.

In the third analysis, the influence that L1 age of acquisition has on L2 lexical availability was examined along with the other variables of interest. L1 age of acquisition did not have an effect over and above the other variables. This supports previous findings (Izura and Ellis 2002, 2004; Izura et al. 2011) showing that the age of acquisition effect is not due to the age at which we learned words (or objects or faces) but the order at which these items are learned. Thus, when the L2 is learned after childhood, the L1 and L2 age of acquisitions overlap to a certain extent because some words are learned early in both cases (e.g., *hand* and *water*) but they do not match. This is because while part of the vocabulary is relevant for children but not for adults (e.g., *nappy, balloon, dolly*), other words are relevant for adults but not so much for children (e.g., *bill, accommodation, menu*). This implies that the closer the L1 and L2 are learned the more similar the L1 and L2 age of acquisition measures, and as a consequence L1 and L2 lexical availability measures will be. Therefore, using L1 lexical availability measures to select the teaching vocabulary for L2 is particularly relevant if the L2 is learned during childhood. The criteria (i.e. L1 lexical availability indexes) can be applied when L2 is learned during the adolescence or adulthood but taking into account that in this latter case learners have slightly different needs.

The assumptions the authors have been making about age of acquisition are based on the idea that this effect resides in the connections between representations such as those between concepts and words (Ellis and Lambon Ralph 2000). However, interesting questions arise if we consider that the age of acquisition might be a property of the conceptual system itself. In this case, cultural differences might play a key role. This is because although there are a large number of concepts that are learned at similar orders in two different cultures (e.g., *bed* in British English and *cama* in Spanish) other concepts are not (e.g., *kettle* is learned early in British English but its translation *tetera* is a relatively late acquired word in Spanish). Therefore, differences between L1 and L2 age of acquisition can emerge because

L1 and L2 speakers have different needs but also because they have different cultures. This question is out of the scope of the present study but it might be interesting to address it in future research.

10.5 Conclusions

There is a close relationship between lexical availability, applied linguistics and cognitive research. This is due to the fact that lexical availability is a product of cognitive processes and also a reliable source, useful for teachers and learners, of the vocabulary used by native speakers.

The authors have shown that the available words in L2, as in L1, are those that are learned first and that they also correspond to the more typical examples of the category. In addition, the availability of words in L2 also depends on their similarity to L1 words. This supports models of language development such as the Revised Hierarchical Model RHM, Kroll and Stewart (1994) that proposes strong lexical links between L1 and L2 translation equivalents and weaker links between the concept and L2 words, particularly at the initial stages of L2 learning.

The fact that cognate words (e.g., trumpet vs. *trompeta*) are words more available in L2 raises interesting questions. For example, is there a relationship between cognateness, ease of learning and order of learning? However, the observation that cognateness and no concept familiarity had an influence in the lexical availability of L2 seems to indicate that in L2 the lexical component has great impact in the generation of words. This is supported by the fact that the age of acquisition of L2 reflects the order of acquisition of the lexical form but not the concept (already learned in childhood). Could it be the case then that lexical availability in L2 is mainly lexical? The observed influence of typicality indicates that concepts are involved to a certain extent. However, further research is needed.

Overall, the authors have observed that L1 and L2 lexical availability indexes show commonalities but they also reflect different cognitive realities.

Acknowledgments We express our gratitude for the speed and competence with which Ian Patterson proofread the manuscript and to all the individuals who kindly participated in the data collection.

References

Alameda, J.R., and F. Cuetos. 1995. *Diccionario de frecuencias de las unidas de castellano*. Oviedo: Servicio de Publicaciones de la Universidad de Oviedo.

Álamo, C.D., M.L Mir, T. Olivares, J. Barroso, and A. Nieto. 1999. Efecto de la edad, nivel educativo y estado cognitivo general sobre la fluidez verbal en hispanohablantes. In *Datos normativos preliminaires. First international congress of neuropsychology in the internet: Uniting horizons in neurology*. http://www.uninet.edu/union99/congress/libs/val/v01.html

Alonso, M.A., A. Fernández, and E. Diez. 2011. Oral frequency norms for 67,979 Spanish words. *Behavior Research Methods* 43: 449–458.
Borge-Holthoefer, J., and A. Arenas. 2010. Semantic networks: Structure and dynamics. *Entropy* 12: 1264–1302.
Butman, J., R.F. Allegri, P. Harris, and M. Drake. 2000. Fluencia verbal en español: Datos normativos en Argentina. *Medicina* 60: 561–564.
Cañizal Arévalo, A. 1991. Redes semánticas y disponibilidad léxica en el español de escolares mexicanos. In *Actas del Congreso El Español de América*, ed. C. Hernández et al., 631–641. Valladolid: Junta de Castilla y León.
Caramazza, A. 1998. The interpretation of semantic category-specific deficits: What do they reveal about the organization of conceptual knowledge in the brain? *Neurocase* 4: 265–272.
Cervantes Institute. 2002. Marco común eurpeo de referencia para las lenguas: aprendizaje, enseñanza, evaluación. Unión Europea. http://cvc.cervantes.es/obref/marco
Collins, A.M., and E.F. Loftus. 1975. A spreading-activation theory of semantic processing. *Psychological Review* 82: 407–728.
Collins, A.M., and M.R. Quillian. 1969. Retrieval time from semantic memory. *Journal of Verbal Learning and Verbal Behavior* 8: 240–247.
Cuetos, F., M. Glez-Nosti, A. Barbón, and M. Brysbaert. 2011. SUBTLEX-ESP: Spanish word frequencies based on film subtitles. *Psicológica* 32: 133–143.
De Deyne, S., and G. Storms. 2008. Word associations: Network and semantic properties. *Behavior Research Methods* 40: 213–231.
De Groot, A.M.B., and G.L.J. Nas. 1991. Lexical representation of cognates and noncognates in compound bilinguals. *Journal of Memory and Language* 30: 90–132.
Ellis, A.W., and M.A. Lambon Ralph. 2000. Age of acquisition effects in adult lexical processing reflect loss of plasticity in maturing systems: Insights from connectionist networks. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 26: 1103–1123.
Ferreira, R., and M.S. Echeverría. 2010. Redes semánticas en el léxico disponible de inglés L1 e inglés LE. *Onomázein* 21: 133–153.
Grabowski, J. 2005. Speaking, writing and memory span performance: Replicating the Bourdin and Fayol results on cognitive load in German children and adults. In *Proceedings writing 2004*, ed. L. Allal and J. Dolz. Geneva: Adcom Productions CD-ROM.
Grabowski, J. 2007. The writing superiority effect in the verbal recall of knowledge: Sources and determinants. In *Writing and cognition: Research and applications – Studies in writing*, ed. M. Torrance, L. van Waes, and D.W. Galbraith, 165–179. Amsterdam: Elsevier.
Grogan, A., D.W. Green, N. Ali, J.T. Crinion, and C.J. Price. 2009. Structural correlates of semantic and phonemic fluency ability in first and second languages. *Cerebral Cortex* 19: 2690–2698.
Hernández Muñoz, N., C. Izura, and A.W. Ellis. 2006. Cognitive aspects of lexical availability. *European Journal of Cognitive Psychology* 18: 734–755.
Hernández Muñoz, N. 2010. Social aspects of oral and lexical written production in Spanish. *SKY Journal of Linguistic*s 23: 101–123.
Izura, C., and A.W. Ellis. 2002. Age of acquisition effects in word recognition and production in first and second languages. *Psicológica* 23: 245–281.
Izura, C., and A.W. Ellis. 2004. Age of acquisition effects in translation judgement tasks. *Journal of Memory and Language* 50: 165–181.
Izura, C., N. Hernández Muñoz, and A.W. Ellis. 2005. Category norms for 500 words in five semantic categories. *Behavior Research Methods, Instruments, & Computers* 13: 385–397.
Izura, C., M.A. Pérez, E. Agallou, V.C. Wright, J. Marín, H. Stadthagen González, and A.W. Ellis. 2011. Age/order of acquisition effects and the cumulative learning of foreign words: A word training study. *Journal of Memory and Language* 64: 32–58.
Kempler, D., E.L. Teng, M. Dick, I.M. Taussig, and D.S. Davis. 1998. The effects of age, education and ethnicity on verbal fluency. *Journal of the International Neuropsychological Society* 4: 531–538.

Kroll, J.F., and E. Stewart. 1994. Category interference in translation and picture naming: Evidence for asymmetric connections between bilingual memory representations. *Journal of Memory and Language* 33: 149–174.

López Chávez, J., and C. Strassburger Frías. 1987. Otro cálculo del índice de disponibilidad léxica: Presente y perspectiva de la investigación computacional en México. In *Actas del IV simposio de la Asociación Mexicana de lingüística aplicada*, ed. H. López Morales, 187–207. Mexico City: UNAM.

López Chávez, J., and C. Strassburger Frías. 1991. Un modelo para el cálculo del índice de disponibilidad léxica individual. In *La enseñanza del español como lengua materna*, ed. López Morales, 91–112. Río Piedras: Universidad de Puerto Rico.

López Chávez, J., and C. Strassburger Frías. 2000. El diseño de una fórmula matemática para obtener un índice de disponibilidad léxica confiable. *Anuario de letras* 38: 227–251.

Lotto, L., and A.M.B. De Groot. 1998. Effects of learning method and word type on acquiring vocabulary in a new language. *Language Learning* 48: 31–69.

Luo, L., G. Luk, and E. Bialystok. 2010. Effect of language proficiency and executive control on verbal fluency performance in bilinguals. *Cognition* 114: 29–41.

Michèa, R. 1953. Mots fréquents et mots disponibles. Un aspect nouveau de la statistique de langage. *Les Langues Modernes* 47: 338–344.

Monsch, A.U., M.W. Bondi, N. Butters, J.S. Paulsen, D.P. Salmon, P. Brugger, and M.R. Swenson. 1994. A comparison of category and letter fluency in Alzheimer's disease and Huntington's disease. *Neuropsychology* 8: 25–30.

Ostrosky-Solis, F., A.L. Gutierrez, M.R. Flores, and A. Ardila. 2007. Same or different? Semantic verbal fluency across Spanish-speakers from different countries. *Archives of Clinical Neuropsychology* 22: 367–377.

Romero Rubilar, C. 2000. Variabilidad y prototipicidad en vocabulario disponible. *RLA* 38: 139–148.

Soto, P., M.V. Sebastián, E. García, and T. del Amo. 1982. *Categorización y datos normativos en España*, Colección Monografías del ICE. Madrid: Universidad Autónoma de Madrid.

Steyvers, M., and J.B. Tenenbaum. 2005. The large-scale structure of semantic networks: Statistical analyses and a model of semantic growth. *Cognitive Science* 29: 41–78.

Tokowicz, N., J.F. Kroll, A.M.B. De Groot, and J.G. Van Hell. 2002. Number of translation norms for Dutch-English translation pairs: A new tool for examining language production. *Behavior Research Methods, Instruments, & Computers* 34: 435–451.

Tröster, A.I., D.P. Salmon, D. McCullough, and N. Butters. 1989. A comparison of the category fluency deficits associated with Alzheimer's and Huntington's disease. *Brain and Language* 37: 500–513.

Urrutia, M.A. 2003. Redes semánticas en línea: una tarea de acceso léxico a partir de un estudio experimental. *RLA* 41: 119–141.

Warrington, E.K., and R.A. McCarthy. 1987. Categories of knowledge: Further fractionations and an attempted integration. *Brain: A Journal of Neurology* 110: 1273–1296.

Part III
Conclusion

Chapter 11
Researching Lexical Availability in L2: Some Methodological Issues

Marta Samper Hernández and Rosa María Jiménez Catalán

11.1 Introduction

Throughout the pages of this book, lexical availability and vocabulary researchers show the potential of lexical availability research to uncover lexical aspects of learners of English and Spanish as second or foreign languages (L2).[1] In this chapter, we address some methodological issues that are necessary to consider in undertaking research on lexical availability. To this end, we apply an analysis to the studies included in the book paying attention to the following aspects: Type of study, Population, Sample, Lexical availability Task, Data processing and analysis. The purposes of our chapter are twofold; on the one hand, we set out to unfold the characteristics shared by the studies included in this volume, on the other, to clarify basic terms and concepts in lexical availability research.

[1] Strictly speaking, the terms 'second language' and 'foreign language' refer to different contexts: natural versus the formal context of the classroom. In practice, it is difficult to draw a sharp line between them as there may be situations in which both are mixed; as happens when acquiring French in France in interactions with native speakers at the same time than attending classes to learn the language. Therefore, either the terms are used interchangeably or one of them is used to stand for the other. In this book, we use 'second language', 'foreign language' and 'L2' interchangeably to refer to a language different from the mother tongue; we use them regardless of the language learning context or situation: be it English or Spanish as school subject, as vehicular language or as immersion language. Likewise, unless we need to do such distinction, we use the acronym L2 to refer to a language other than the mother tongue (L1).

M. Samper Hernández (✉)
Departamento de Filología Española. Clásica y Arabe, Universidad de Las Palmas de Gran Canaria, Las Palmas de Gran Canaria, Spain
e-mail: msamper@dfe.ulpgc.es

R.M. Jiménez Catalán
Departamento de Filologías Modernas. Facultad de Letras y Educación, Universidad de La Rioja, C/San José de Calasanz s/n, 26004 Logroño, La Rioja, Spain
e-mail: rosa.jimenez@unirioja.es

R.M. Jiménez Catalán (ed.), *Lexical Availability in English and Spanish as a Second Language*, Educational Linguistics 17, DOI 10.1007/978-94-007-7158-1_11,

11.2 Characteristics

11.2.1 Type of Study

The book contains 11 chapters, this included. As mentioned in the preface, the first chapter is a brief account of the history of lexical availability studies narrated by Professor Humberto López Morales, the founder of this tradition of studies in Spanish language. His chapter serves to introduce nine original studies on L2 learners' lexical availability. The approach adopted is mainly quantitative but also qualitative in some chapters. These are purposely designed as to achieve thematic and structure unity by following an identical pattern: introduction, literature review, objectives, methodology, results, discussion and conclusion. They are mainly observational studies, and within this category we find descriptive as well as comparative and correlation studies in which we look at the relation between lexical availability and age, level, learning contexts and educational stages.

Allwright and Bailey (1991) make a distinction between two traditions in research with regard to language learning in classroom contexts: experimental and ethnographic. In the former, researchers try to corroborate or refute hypotheses derived from theory and they do so by means of the careful selection of samples, data gathering instruments, and observation schedules. In the latter, data are drawn out of observations of the language learning context, without any manipulation of the variables, from a stance closer to the ethnographic research. The studies reported in the book were conducted in real learning contexts without any modification of their naturalistic conditions; that is to say, they are not experimental studies created in the laboratory specifically aimed to prove or disprove a hypothesis.

11.2.2 Population

English and Spanish are within the world three top spoken languages together with Chinese. Both are also found in the ranking of the most learnt languages. There are no concluding figures but estimates point to millions of English learners as well as Spanish learners spread throughout the five continents (see Graddol 2006; Instituto Cervantes 2012). Obviously, it would be impossible to conduct research with all the learners of English or Spanish in the world, even on all the learners of one country.

Availability studies deal with well-defined target populations. With caution, a study on the lexical availability of 6th Primary school EFL learners in La Rioja might be generalized to other populations of similar characteristics. In Spain, English is a compulsory subject throughout primary and secondary education. For each course and educational stage, the linguistic content, the number of hours of instruction, the teaching methodologies, teachers' training and teachers' selection are regulated by law. The 17 self-governing Spanish communities, among them, La Rioja, share similar conditions as regards age range, number of hours of instruction, schedule, and teaching approach.

Table 11.1 A sample of studies on lexical availability in Spanish and English as L2

Language learning context and country	Author/s	Year
Spanish L2		
Secondary education, Turku, Finland	Carcedo-González	2000
International courses, Salamanca, Spain	Samper Hernández	2002
International studies abroad, Sevilla, Spain	Sánchez-Gómez	2005
University, China	Jing Lin	2006
Language schools, Madrid, Spain	López-Rivero	2008
Alcalingua, Alcalá, Spain	Frey-Pereira	2008
Language schools, Madrid, Spain	Pérez-Serrano	2009
University language centre, Cádiz, Spain	Sánchez-Saus	2009
Free University of Berlin and Cervantes Institute, Germany	Medina-Arejita	2009
Secondary schools, Romania	Sandu	2009
Secondary schools, Poland	López-González	2010
University, Eslovenia	Sifrar Kalan	2011
Inmigrants, Valladolid, Spain	Fernández-Merino	2011
Secondary schools and University, Morocco	Serfati and Abidi	2013
Inmigrants in secondary schools, Pamplona, Spain	Jiménez-Berrio	2013
University, Las Palmas de Gran Canaria, Spain	Del Pino	In progress
English L2		
Secondary education, Chile	Germany and Cartes	2000
University, Chile	Ferreira	2006
Primary schools, La Rioja, Spain	Jiménez Catalán and Ojeda Alba	2009a
Primary schools, La Rioja, Spain	Jiménez Catalán and Ojeda Alba	2009b
Secondary schools, La Rioja, Spain	Fernández-Fontecha	2010
University, Chile	Ferreira and Echeverría	2010
Primary schools, La Rioja, Spain	Jiménez Catalán	2010

Thus, in this volume, the Target Populations are determined by the objectives arising from each study in particular. In the case of Spanish as L1, the PanHispanic project brings a solid, validated framework through the great number of studies accomplished by means of the same test and the same methodological guidelines in different parts of Spain and the world. As far as Spanish L2 is concerned, we find the pioneering studies conducted by Carcedo-González (1999, 2000) on Finnish students, learners of Spanish as a foreign language at secondary education and university. His research is considered a benchmark for other studies that have appeared over the last decade. A sample of studies conducted on Spanish L2 and English up to the moment is displayed in Table 11.1.

As Table 11.1 reveals, in the case of English L2, research is still scarce; contrary to what occurs in Spanish, where there exists the big PanHispanic project on lexical availability, we do not have an English as L1 reference, although the pioneering research of Dimitrijevic (1969) with Scottish students at a secondary school, as well as those carried out by Bailey (1971) with monolingual and bilingual in English and

Spanish, can serve as a starting point. Likewise, the studies conducted in Spanish as L2 in the last decades, as well as the studies included in this book, can serve to draw comparisons among groups of learners of similar characteristics. Both the Spanish and English L2 studies in this book are exploratory in nature; its purpose is to observe trends and check if the methodology designed for lexical availability research in L1 may be suitable in L2.

11.2.3 Sample

The most remarkable traits of the samples analyzed in this book along with the populations from which they were drawn are summarized in Table 11.2. As we can see, the size of the groups ranges between 210 and 18 informants. Three studies are based on a sample of 100 or more informants, four comprise a sample of between 40 and 50 informants, and two are based on samples of less than 30 informants.

Although exploratory in nature, the nine studies in this book have a high degree of external validity if one takes into account the size of the sample in relation to the population, as well as the similarities between populations. For example, the 18 students that make up the sample in the study by Gallardo and Martinez actually correspond to a population of 60 English learners over 55 in the Third Experience Classrooms (*Aulas de la experiencia*) in the University of the Basque country. These classrooms make up for the total Population and are scattered in three campuses: Alava, Gipuzkoa, and Vizcaya; it is in the latter where the sample was collected (see Chap. 4). Likewise, the external validity of the samples of learners of Spanish in the International Courses run by University of Salamanca (Chaps. 7 and 10) must be appreciated bearing in mind the total size of the population (550 informants), which gives us a high degree of representativeness, particularly taking into account that Labov (1966) established a 0.025 % as a good ratio in sociolinguistic studies. In addition, it is necessary to consider that Target populations look at the specific context or situation in order to understand it. As noted in Allwright and Bailey (1991: 51), in these situations it does not matter so much the external validity but the understanding of reality: “Instead of claiming that whatever has been discovered must be true of people in general, a naturalistic enquirer will claim that whatever understanding has been gained by an in-depth study of a real-life classroom may illuminate issues for other people”.

11.2.4 Data Collection Instrument: The Lexical Availability Task

All the chapters in this book are based on a similar data collection instrument, an associative task, used in lexical availability studies, this gives the book a great deal of consistency. The processes followed are almost identical: by means of a

Table 11.2 Main characteristics of the samples in the book

Study	L1	L2	Level	Hours	Target population	Sample	Age
Chapter 2	English	English	B2-C1	1,000	Total: 100	50 NS	15–16
Ferreira and Echeverría					The Royal school, Haslemere, Surrey, United Kingdom.	50 NNS	+18
					Total: 100		
					University of Concepción, Chile		
Chapter 3	Spanish	English	A1+		Total: 3,005	13 out of 300	12
Jiménez Catalán, Agustín, Fernández and Canga			Samples for comparison selected on the basis of same scores on VLT		6th Primary Rioja, Spain		
					Total: 150	13 out of 150	18–19
					1st year of Education Degree		
					University of La Rioja, Spain		
Chapter 4	Spanish	English	Beginners	90–180	Total: 60	18	56–69
Gallardo del Puerto and Martínez Adrián			False beginners	Mean = 115	Senior EFL students at University of the Basque Country, Spain		
Chapter 5	Spanish	English	A1	629	Total = 2,613	190	12
Agustín Llach and Fernández Fontecha			A1+		6th Primary Education		
					Total: 2,643		
				944	9th Grade Secondary Education		15
					Rioja, Spain		
Chapter 6	Spanish	English	A1	629	6th Primary Education: Total = 2,613	25	12
Jiménez Catalán and Fitzpatrick			A1+	839	8th Grade Secondary Education = 2,792	25	14
					La Rioja, Spain		

(continued)

Table 11.2 (continued)

Study	L1	L2	Level	Hours	Target population	Sample	Age
Chapter 7	English	Spanish	Basic: A1-B1		Total: 550	45	+18
Samper Hernández	Italian Japanese		Spanish advanced: B2-C1		International students at University of Salamanca, Spain		
Chapter 8	Slovene	English		English = 9 years	Total: 50	40 females	20–24
Šifrar Kalan		Spanish		Spanish = 8 years	Total: 30	20 English	22–30
					Faculty of Arts of the University of Ljubljana, Slovenia	20 Spanish	
Chapter 9	Polish	Spanish	B1	630	Total: 270	90	13–15
López González			B2		Year Zero at High School intensive	120	16–19
					Total: 265		
					3rd Year at Middle School extensive		
Chapter 10	Spanish	Spanish	B1, B2		Total: 800	43	17–23
Hernández, Izura and Tomé	English				American students at University of Salamanca, Spain		

paper-and pencil questionnaire, students are presented with cue words[2]; and asked to write down all the words that come to their minds in response. The time given for cue word is 2 min. Each category is displayed on a separate page made up of numbered lines. Participants are not allowed to move to the following page until the 2 min period is over. Numeration of word responses is an important factor in lexical availability research in order to establish the occurrence of each word response, and the lexical availability index (see Sect. 11.3 in this chapter); in other words, the position that each word holds in the list together with frequency is used to calculate its index of availability. Similarly, controlling time serves for comparing the words retrieved by groups in response to a given prompt in an equal time span.

There is also a great degree of consistency in relation to the cue words included in most lexical availability studies. The most frequently used are the 16 centers of interest proposed by the pioneering work of French researchers in their attempt to cover the totality of the speakers' semantic fields Gougenheim et al. (1967: 152–153). In this book, we find variation in the number of prompts depending on the objectives of each study. The number ranges from the 16 traditional prompts (Chaps. 7 and 9) up to two prompts (Chap. 3). Regardless the number, the studies in this book are based on prompts already used in the PanHispanic project. Again, this allows future comparisons among chapters as well as comparisons with other studies of this kind. For instance, the results reported for 'Town' and 'Countryside' in Chap. 3 can be contrasted to the results reported in Chaps. 4, 5, 6, 7, 8 and 9, as these cue words are included in them all. Moreover, they can be compared to the ones obtained in Spanish as L1 and L2. Thanks to using an identical task and prompts, results could certainly be compared to the ones achieved by most research availability studies.

We should also note that the present studies do not set a limit to the number of possible word responses but employ an open lists system, as it is frequently done in lexical availability studies. Therefore, time rather than number of responses is the imposed ceiling.

Although research based on closed lists is also used in accordance with the research purpose, in fact they were used in the early work by Gougenheim et al. (1967) and Mackey (1971), open lists have the advantage of avoiding the limit of a preset number of responses, thus allowing researchers the possibility of conducting many and varied analyses, particularly of sociolinguistic nature. Likewise, all the studies in the book make use of a written rather than oral task, as it happens in the PanHispanic Project as well as in the French early work. Nevertheless, care should be taken with populations different to the ones targeted in the book, when perhaps oral rather than the written task would be advisable. This is the case, for instance, of preschool children Sánchez-Corrales and Murillo Rojas (1993) or old-age pensioners with low level of education Borrego-Nieto (2008). Table 11.3 groups the studies in this volume according to prompts, time and other characteristics.

[2] It is important to note that 'cue word', 'prompt', 'center of interest', or 'semantic category' are all used interchangeably by researchers as shown in Table 11.3.

Table 11.3 Time, prompts, formulas and terminology

Study	Time	Prompts	Terms	Additional tasks	Formula
Chapter 2 Ferreira and Echeverría	2′	'Body parts', 'Food and drink', 'Terrorism and crime', 'Health and medicine' Total: 4	Semantic category	Background questionnaire	López-Chávez and Strassburguer-Frías (1991)
Chapter 3 Jiménez, Agustín, Fernández and Canga	2′	'Town' and 'Countryside' Total: 2	Semantic domain	Background questionnaire	Means, test for significance Corpus analysis
Chapter 4 del Puerto and Martínez Adrián	2′	'Parts of the body', 'Clothes', 'House', 'Furniture', 'Food and drink', 'Table', 'Kitchen', 'School', 'Town', 'Countryside', 'Means of transport', 'Animals', 'Hobbies', 'Professions', and 'Colours' Total: 15		Background questionnaire	Means, test for significance
Chapter 5 Agustín and Fernández	2′	'Parts of the body', 'Food and drink', 'School' 'Town', 'Countryside', 'Transport', 'Animals', 'Sports', and 'Professions' Total: 9	Prompts, cue words, Semantic fields	Background questionnaire	Means, test for significance
Chapter 6 Jiménez and Fitzpatrick	2′	'Animals', 'Parts of the body', 'Countryside', 'Food and drink', 'Professions', 'School', 'Hobbies' 'Town', 'Transport' Total: 9	Prompts Cue words	Background questionnaire	Means, test for significance, (external frequencies VocabProfile)

Chapter 7 Samper Hernández	2′	‘Parts of the body’; ‘Clothes’; ‘Parts of the house’ without (furniture); ‘Furniture’ (house); ‘Food and drink’; ‘Objects Placed on the Table at Meals’; ‘The Kitchen and its Utensils’; ‘The School: Furniture and Materials’; ‘Lighting, Heating and Means of Airing Places’; ‘The City’; ‘The Countryside’; ‘Means of Transport’; ‘Gardening and Farming’; ‘Animals’; ‘Games and Entertainment’; ‘Professions and Jobs’. Total. 16	Lexical domains Cue words Topic Lexical fields Prompts	Background questionnaire	López-Chávez and Strassburger-Frías (1991)
Chapter 8 Šifrar Kalan	2′	‘Parts of the body’ *Partes del* (*cuerpo*), ‘House’ (*La casa*), ‘Food and drink’ (*Comida y bebida*), ‘School’ (*La escuela*), ‘City’ (*La ciudad*) ‘Countryside’ (*El campo*), ‘Animals’ (Animales), ‘Games and Entertainment’ (*Juegos y distracciones*) Total: 8 English & Spanish)	Lexical Semantic category	Background questionnaire	López-Chávez and Strassburger-Frías (1991)
Chapter 9 López González	2′	(1) ‘Parts of the human body’, (2) ‘Clothing’, (3) ‘Parts of the house’, (4) ‘House furniture’, (5) ‘Food and drink’, (6) ‘Objects on the table for the meal’, (7) ‘The kitchen and its utensils’, (8) ‘School (furniture and materials)’, (9) ‘Heating and lighting’, (10) ‘The city’, (11) ‘The countryside’, (12) ‘Means of transport’, (13) ‘Farm and Garden Work’, (14) ‘Animals’, (15) ‘Games and entertainment’, (16) ‘Jobs and professions’ Total: 16	Center of interest	Background questionnaire	López-Chávez and Strassburger-Frías (1991)
Chapter 10 Hernández, Izura and Tomé	4′ 2′	‘Animals’, ‘Body parts’, ‘Clothing’ and ‘Furniture’ Note. For this study only the words produced in the first 2′ were considered Total: 4	Semantic category	Background questionnaire	López-Chávez and Strassburger-Frías (1987, 1991, 2000)

11.2.5 Data Editing

The careful editing of the data is a must in lexical availability studies, since a careless editing may affect the results as well as possible comparisons with other studies. Therefore, it is absolutely necessary to follow established guidelines concerning the lemmatization of word responses. These guidelines follow the criteria established by Samper Padilla (1998) for L1 Spanish populations, later applied to L2 Spanish by Carcedo-González (2000), and revised by Samper Hernández (2001, 2002), taking into account specific problems concerning Spanish L2. Carcedo-González (2000) was adapted to English as L2 by Jiménez Catalán and Ojeda Alba (2009a). The reader may find explicit references to data editing criteria in the contributions by Ferreira and Echeverria (Chap. 2) and López González (Chap. 9) for Spanish L2. As to English L2, explicit references are provided in Chap. 3, where the excerpt below was drawn:

> (i) correcting spelling mistakes, (ii) counting repeated words only once per prompt, (iii) discarding unintelligible words and Spanish words, (iii) inserting a hyphen in lexical units containing more than one word (e.g., post-office), (iv) deleting proper names that have the same spelling in English and Spanish as for instance, Paris, Portugal, but keeping those that are written in a different way in these languages (e.g., New York, London). Jiménez, Agustín, Fernández and Canga, (Chap. 3).

11.3 Data Processing and Analysis: Terminology and Basic Concepts

Data processing and analysis is another relevant issue in lexical availability studies. As explained in depth by López Morales in Chap. 1, the adoption of a mathematical formula capable of yielding the availability of the different lexical units was the result of scholars' reflections over years of research. In this section we provide a list of those terms and concepts needed to understand the conclusions achieved by lexical availability studies. Together with the term we provide a brief description, in the belief that they may be useful for students and researchers who are not familiarized with this line of research. Firstly, we list some of the key concepts that are treated in depth by López Morales' chapter. Following, we describe three computer tools that work with the formula created by López Chávez and Strassburger Frías, the one most frequently used in the PanHispanic project. The last part of this section comprises some of the values provided by the programs; among them, the most important one is, logically, the index of lexical availability, although others such as lexical cohesion index also need consideration.

11.3.1 Key Concepts

In Chap. 1, López Morales provides precise definitions of the following concepts: (i) thematic words versus non thematic words, (ii) frequent vocabulary versus available

lexicon, and (iii) basic vocabulary versus fundamental lexicon. They are essential concepts for lexical availability researchers. We will not describe them again in this section rather we remit the reader to Chap. 1, where these concepts are contextualized and contrasted. Likewise, the author gives a detailed review of the steps adopted to arrive at the formula that calculates the index of availability. In this case, we consider useful to provide the reader with further information not contained in the chapters.

11.3.2 Computer Tools

Lexidisp

Sponsored by the association of linguistics and philology of Latin America (ALFAL), Cervantes Institute, and University of Alcalá, Lexidisp is a computer tool that allows researchers to calculate the lexical availability index of each word retrieved by informants. The program also calculates word occurrences, relative and cumulative frequencies for each of the lexical units and makes possible the comparison of sets of data. Lexidisp is one of the programs used in lexical availability studies. This program was used, for instance, in Chap. 7. Further information can be found in Moreno et al. (1995).[3]

Dispolex

Dispolex is a Web site[4] created by researchers at the University of Salamanca, which aimed at collecting information on researchers and projects of lexical availability within the Hispanic world as well as providing the necessary tools for the development of this line of research. It includes a virtual data bank with a tool based on the same mathematical formula as Lexidisp for the calculation of the rate of lexical availability López-Chávez and Strassburguer-Frías (2000). Dispolex also offers the possibility to compare and store data from different investigations. Further information regarding this Web site can be found in Bartol-Hernández and Hernández-Muñoz (2003) and also in Hernández-Muñoz (2010).

Dispogen II

Software created in MATLAB by Echeverria et al. (2005). It allows the calculation of lexical availability values for each word in response to each cue word. As Lexidisp and Dispolex, Dispogen II uses the formula created by López-Chávez and

[3] Information available at http://www.linguas.net/Proyectos/LexiDisp/tabid/73/language/es-ES/Default.aspx

[4] Information available at http://www.dispolex.com

Strassburguer-Frías (1991). As Ferreira and Echeverria note in Chap. 2, this formula: "…computes lexical availability values according to the position that a word takes in a list**,** the number of participants who elicit the word at those positions, and the lowest position the word is observed in any of the lists". For further information see Echeverría et al. (2005) and Hernández-Muñoz et al. (2006).

11.3.3 Measures

Index of Availability Words

This index is the essential concept in lexical availability studies. It stands for the value given to words according to the easiness they come to our minds when conversation deals with a specific theme. According to López Morales (Chap. 1) the order of appearance of a word in the individual data and in the group data provides evidence on the degree of availability of a word: high available words are more likely to appear first in the list of responses. The calculation of this index is performed out of the formula created by López-Chávez and Strassburguer-Frías (1991) which provides an index for each lexical unit based both on its frequency and on the order of appearance in the list of responses (see Chap. 1 for further information). The following factors are taken into account in this formula: (a) the word absolute frequency, (b) the absolute or raw frequency of the word in each position, (c) the number of participants in the task, (d) the number of positions achieved by a given center of interest, (e) the positions in which a given word is found. The formula works in this way:

$$D(P_j) = \sum_{i=1}^{n} e^{-2.3\left(\frac{i-1}{n-1}\right)} \bullet \frac{f_{ji}}{I_1}$$

It should be read as follows: *D (Pj)* stands for the lexical availability value of the word *j* within a semantic category; *n* stands for the highest position achieved in the center of interest in the whole task; *i* stands for the position of the word on a given list; *e* is the natural number (2,718181818459045); *fji* refers to the absolute frequency of word *j* in position *i*; I_1 stands for the number of informants who performed the task (López-Chávez and Strassburguer-Frías 1987, 1991, 2000; Hernández-Muñoz et al. 2006). (See also Chap. 2, in this volume).

Cohesion Index

This concept was proposed by Max Echeverría (1991) to refer to the degree of coincidence in the words retrieved by informants for each prompt. A clear definition of this concept is provided by Hernández-Muñoz (2010: 105) in her analysis of oral and written lexical production in Spanish L1:

> The index of cohesion gives us information about how compact or open a semantic category is and therefore, about the degree of coincidence in the subjects' responses. It is obtained by dividing the average number of responses given by the informant by the number of different words. The maximum value is 1 and it means a complete coincidence in responses. The minimum value is = 0 and it means that there are no common responses

Compact as well as disperse prompts usually coincide in most studies. Two compact prompts are 'Parts of the body' and 'Clothes' for both we find a high percentage of shared words in most informants. In comparison, disperse prompts are 'Gardening' or 'Games and hobbies', where we can observe a wide variety of word types in the informants' responses see for instance, Table 7.3 (in Chap. 7).

11.4 Conclusions

This chapter has attempted to provide evidence of the main characteristics of the studies included in the book. It has also aimed to clarify the specific methodological traits of lexical availability studies as shown in the chapters in the book. All of them share focus and methodology by means of using the same task to look at lexical availability in learners of Spanish and English. Perhaps one of the main contributions of this book is the common tendencies revealed in studies designed and carried out independently the one to the other. In this respect, it is important to highlight here some of the common tendencies observed. Firstly, results tell us, not surprisingly, that native speakers outperform non-native speakers in the number of words retrieved in response to prompts. Similarly, advanced learners outperform lower level learners in their lexical availability retrieval. However, regarding lexical richness, we observe that both native and non-native speakers, as well as advanced and low level learners retrieve a greater number of words from basic most frequent) categories than from advanced less frequent) categories. Secondly, as far as word responses are concerned, a tendency is common to our studies: some prompts are more productive than others and this occurs regardless of the L1, L2, language level, age or sex. Thus, we see how 'Food and drink', 'Animals' and 'Town' are among the most productive prompts, and 'Kitchen', 'Furniture' and 'Table', among the least productive. Likewise, some prompts are more compact and others more disperse. This is the case for instance of 'Countryside' much more disperse than 'Town'. We find other relevant findings that have to do with the specificities of each chapter, and that we will not deal with in this conclusion. However, even if we only pay attention to the common tendencies observed in our studies (and, in turn, their similarity to the results obtained in the framework of lexical availability research in Spanish and English as L1 and L2), it is possible to confirm the validity of the lexical task as a research instrument. As already mentioned, the studies collected in the book were undertaken independently the ones to the others. In other words, they were not designed or planned as a series of studies to prove or disprove their validity but as exploratory studies arising from different research projects in different learning contexts.

Our findings have implications both for language education and for lexical availability and vocabulary researchers. Each study provides insights into the words that learners have in their minds and are capable of retrieving when prompted to do so. In this regard, the book contains invaluable data from learners of different mother tongues, target languages, ages, and learning contexts that can serve as a common framework for future comparative studies.

However, there is still a long way to go in lexical availability research, fortunately. There are certain methodological aspects that could be improved. Although, as we have explained, the samples in our studies are valid with regard to the populations they were drawn from, our research is exploratory in nature and, therefore, the findings should not be taken as conclusive. It is necessary to conduct further research with a larger number of informants and different target populations in order to corroborate or refute the conclusions arrived at by our studies.

An aspect worth further investigation has to do with the type of prompts susceptible to be included in the lexical availability task. Except for 'Terrorism and crime' or 'Health and medicine', the prompts employed to elicit words have usually been the most frequently used in lexical availability studies. In future research, it would be interesting to include new prompts as well as word classes different to nouns in order to elicit other kind of word responses as well as patterns of associations. In this respect, presenting learners with contextualized prompts such as 'at the restaurant', 'a gathering at a restaurant' (rather than 'Food and drink') or 'a day at the countryside' (rather than 'Countryside'), would be particularly advisable as to ascertain whether contextualized prompts yield a wider range of word classes and more infrequent words than the traditional prompt.

Likewise, many studies in lexical availability are rather sociolinguistic accounts of the available lexicons of speakers of different communities. It seems as if studies in L2 have followed this thread, in their attempt to give a quantitative or qualitative description of the available words or the words learners produce by means of prompts in the lexical availability task. It is perhaps time to explore other lines of research closer to psycholinguistics and to vocabulary testing. Regarding the former, Chap. 10 stands for the kind of studies that could be addressed in the future; but certainly, more research of this kind is needed, particularly studies aimed to explore networks and chains of associations in the words retrieved by learners of different characteristics. As to vocabulary testing, although many of the studies have been related to educational contexts, most have focused on the description of the words learners retrieve according to centers of interest. This is, certainly, a first step, as to know the words learners are capable of retrieving in response to prompts related to specific situations or semantic fields and the words that they are not capable of retrieving is important for language teachers. However, the lexical availability task is not a vocabulary test in the fashion of Lex30 or Vocabulary Levels Test to cite some well-known examples in L2 vocabulary research. It does not tell us how many words learners know. To be fair, vocabulary size was neither the purpose of the lexical availability studies nor the purpose of the task. However, exploring the potential of the lexical availability task, alone or in combination with

vocabulary tests, might be also another step forward in lexical availability studies and in vocabulary research.

References

Allwright, D., and K.M. Bailey. 1991. *Focus on the language classroom. An introduction to classroom research for language teachers*. Cambridge: Cambridge University Press.

Bailey V.J. 1971. *A study of lexical availability among monolingual-bilingual speakers of Spanish and English.* Unpublished thesis. Houston: Rice University.

Bartol-Hernández, J.A., and N. Hernández-Muñoz. 2003. *Dispolex: Base de datos de la disponibilidad léxica*. Salamanca: Universidad de Salamanca. http://www.dispolex.com

Borrego-Nieto, J.B. 2008. Edad y cultura léxicas. In *Estudios sobre disponibilidad léxica en los jóvenes aragoneses*, ed. M.L. Arnal Purroy, 227–244. Zaragoza: Institución "Fernando el Católico" C.S.I.C.

Carcedo-González, A. 1999. Desarrollo de la competencia léxica en español LE: análisis de cuatro fases de disponibilidad. *Pragmalingüística* 5–6: 75–94.

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

Dimitrijévic, N. 1969. *Lexical availability*. Heidelberg: Julius Gross Verlag.

Echeverría, M.S. 1991. Crecimiento de la disponibilidad léxica en estudiantes chilenos de nivel básico y medio. In *La enseñanza del español como lengua materna, Actas del II Seminario Internacional sobre 'Aportes de la lingüística a la enseñanza del español como lengua materna*, ed. Humberto López Morales, 61–78. Río Piedras: Universidad de Puerto Rico.

Echeverría, M.S., P. Urzúa, and I. Figueroa. 2005. *Dispogen II. Programa computacional para el análisis de la disponibilidad léxica*. Concepción: Universidad de Concepción.

Fernández-Fontecha, A. 2010. Gender and motivation in EFL vocabulary production. In *Gender perspectives on vocabulary in foreign and second languages*, ed. R.M. Jiménez-Catalán, 93–117. New York: Palgrave Macmillan.

Fernández-Merino, P.V. 2011. Presencia del léxico disponible de inmigrantes en glosarios específicos de vocabulario. *Cuadernos Comillas* 2: 1–18.

Ferreira, R.A. 2006. *Disponibilidad léxica en inglés como lengua materna e inglés como lengua extranjera: Estudio del léxico disponible desde un enfoque psicolingüístico*. Unpublished dissertation, University of Concepción.

Ferreira, R.A., and M.S. Echeverría. 2010. Redes semánticas en el léxico disponible de inglés L1 e inglés LE. *Onomazein* 21: 133–153.

Frey-Pereira, M.L. 2008. *El léxico disponible en los escritos de alumnos de español como lengua extranjera. Estudio comparativo de dos córpora.* Unpublished thesis. Alcalá de Henares: Universidad de Alcalá.

Germany, P., and N. Cartes. 2000. Léxico disponible en inglés como segunda lengua en instrucción personalizada. *Estudios Pedagógicos* 26: 39–50.

Gougenheim, G. 1967. La statistique de vocabulaire et son application dans l'enseignement des langues. *Revue de l'Enseignement Supérieur* 3: 137–144.

Gougenheim, G., R. Michéa, P. Rivenc, and A. Sauvageot. 1967. *L'élaboration du français fondamental 1er degré. Étude sur l'élaboration d'un vocabulaire et d'un grammaire de base.* Nouvelle édition refondue et augmentée. Paris: Didier.

Graddol, D. 2006. *English next*. London: British Council.

Hernández-Muñoz, N. 2010. Social aspects of oral and written lexical production in Spanish. *SKY Journal of Linguistics* 23: 101–123.

Hernández-Muñoz, N., C. Izura, and A.W. Ellis. 2006. Cognitive aspects of lexical availability. *European Journal of Cognitive Psychology* 18: 730–755.

Instituto Cervantes. 2012. *El Español, una lengua viva*. Informe 2012, Instituto Cervantes.

Jiménez-Berrio, F. 2013. *Léxico disponible de inmigrantes escolares no hispanohablantes*. Pamplona: Servicio de Publicaciones de la Universidad de Navarra.

Jiménez-Catalán, R.M. 2010. Gender tendencies in EFL across vocabulary tests. In *Gender perspectives on vocabulary in foreign and second languages*, ed. R.M. Jiménez-Catalán, 117–139. New York: Palgrave Macmillan.

Jiménez-Catalán, R.M., and J. Ojeda Alba. 2009a. Girls' and boys' lexical availability in EFL. *ITL International Journal of Applied Linguistics* 158: 57–76.

Jiménez-Catalán, R.M., and J. Ojeda Alba. 2009b. Disponibilidad léxica en inglés como lengua extranjera en dos tipos de instrucción. *Lenguaje y Textos* 30: 166–176.

Labov, W. 1966. *The social stratification of English in New York City*. Cambridge: Cambridge University Press.

Lin, J. 2006. *Disponibilidad léxica de alumnos chinos estudiantes de español como lengua extranjera*. Final master's report. Alcalá de Henares: Universidad de Alcalá.

López-Chávez, J., and C. Strassburguer-Frías. 1987. Otro cálculo del índice de disponibilidad léxica. In *Presente y perspectiva de la investigación computacional en México. Actas del IV Simposio de la Asociación Mexicana de Lingüística Aplicada*. México: UNAM.

López-Chávez, J., and C. Strassburguer-Frías. 1991. Un modelo para el cálculo del índice de disponibilidad léxica individual. In *La enseñanza del español como lengua materna*, ed. H. López-Morales, 91–112. Río Piedras: Universidad de Puerto Rico.

López-Chávez, J., and C. Strassburguer-Frías. 2000. El diseño de una fórmula matemática para obtener un índice de disponibilidad léxica confiable. *Anuario de letras* 38: 227–251.

López-González, A.M. 2010. La evaluación del desarrollo de la competencia léxica en L2 por medio de la disponibilidad léxica. *redELE Red Electrónica de Didáctica del Español como Lengua Extranjera* 18. http://www.educacion.gob.es/dctm/redele/MaterialRedEle/Revista/2010_18/2010_redELE_18_01Lopez.pdf?documentId=0901e72b80dd3177/. Accessed 28 Nov 2011.

López-Rivero, E. 2008. Estudio de disponibilidad léxica en 43 estudiantes de ELE. Final master's report. Madrid: Universidad Antonio de Nebrija.

Mackey, W.F. 1971. *Le vocabulaire disponible du français*. Paris/Bruxelles/Montréal: Didier.

Medina-Arejita, E. 2009. Las nociones específicas del PCIC y la disponibilidad léxica como instrumento de selección del vocabulario. Final master's report, Instituto Cervantes and Universidad Internacional Menéndez Pelayo.

Moreno, Francisco, J. Enrique Moreno, and A.J. García de las Heras. 1995. Cálculo de la disponibilidad léxica. El programa LexiDisp. *Lingüística* 7: 243–249.

Pérez-Serrano, M. 2009. Estudio de disponibilidad léxica en estudiantes de E/LE en los centros de interés "Medios de transporte" y "Profesiones y oficios". Final master's report, Instituto Cervantes and Universidad Internacional Menéndez Pelayo.

Samper-Hernández, M. 2001. Dificultades de los estudios de disponibilidad léxica en ELE: los criterios de edición de los materiales. In *Nuevas aportaciones al estudio de la lengua española. Investigaciones filológicas*, ed. J.A. Bartol et al., 277–285. Salamanca: Luso-Española de Ediciones.

Samper-Hernández, M. 2002. *Disponibilidad léxica en alumnos de español como lengua extranjera*, Colección Monografías nº 4. Málaga: ASELE.

Samper-Padilla, J.A. 1998. Criterios de edición del léxico disponible. *Lingüística* 10: 311–333.

Sánchez-Corrales, V., and M. Murillo Rojas. 1993. Campos semánticos y disponibilidad léxica en preescolares. *Revista Educación* 17(2): 15–25.

Sánchez-Gómez, C. 2005. Naturaleza gramatical del léxico disponible en informantes de español como lengua extranjera. *Interlingüística* 16(2): 977–986.

Sánchez-Saus, M. 2009. La variable 'nivel de español' en el léxico disponible de los estudiantes de español como lengua extranjera. *Pragmalingüística* 17: 140–153.

Sandu, B. 2009. *Disponibilidad léxica en español como lengua extranjera de los alumnos rumanos* (*Estudio exploratorio*). Diploma of Advanced Studies' report. Las Palmas de Gran Canaria: Universidad de Las Palmas de Gran Canaria.

Serfati, M., and L. Aabidi. 2013. *Disponibilidad léxica de ELE en Marruecos niveles de Secundaria y Enseñanza Superior en la región Souss Massa Drâa*. Agadir: Universidad Ibn Zohr.

Šifrar Kalan, M. 2011. Leksikalna Razpoložljivost V Španščini Kot Tujem Jeziku. Unpublished thesis. Ljubljana: Univerza v Ljubljani.