Content list available at www.urmia.ac.ir/ijltr

Iranian Journal of Language Teaching Research

Urmia University

The available lexicon: A tool for selecting appropriate vocabulary to teach a foreign language

Antonio Manuel Ávila Muñoz [a, *]

[a] *University of Málaga, Spain*

ABSTRACT

This study aims to provide foreign language professionals with a sound methodology for selecting a suitable lexicon apropos of their students' level in the language. The justification of said selection is, herein, rooted in a cognitive argument: If we are able to observe the manner in which words are organized within the mind, we will be better able to select the words needed for the natural process of communication. After analyzing over lists of lexical availability compiled by previous analyses, this study puts forth a glossary filtered by way of an objective procedure based on the mathematical concept known as *Fuzzy Expected Value*. I begin first by rigorously defining the concept of lexical availability and then thoroughly examining and explaining the manner in which I have obtained the results. Next, I employ the cognitive theory of prototypes to expound upon the organizational apparatus which arranges words within speakers' minds. Subsequently, and in accordance with objective criteria, a lexical selection is proposed. To end, I contemplate and muse upon the significance of a program that would enable us to identify the most appropriate vocabulary according to the students' level of linguistic competence. In order to further substantiate this study, it will be juxtaposed with the specific notions outlined by the curriculum of The Cervantes Institute. Moreover, it will relate to the teaching levels proposed by the *American Council on the Teaching of Foreign Languages* (ACTFL) and *Common European Framework of Reference* for Languages (CEFR).

Keywords: lexical availability; theory of prototypes; lexical selection; fuzzy expected value

ARTICLE HISTORY

Received: 27 Jun. 2016 **Revised version received:** 30 Oct. 2016

Accepted: 10 Dec. 2016 **Available online:** 1 Jan. 2017

* Corresponding author: Faculty of Letters & Philosophy, University of Málaga, Spain
Email address: amavila@uma.es

Introduction

Studies carried out on lexical availability offer a number of possibilities for its practical application. Other disciplines closely linked to linguistics have made interesting use of available lexicon material. All have found that their records provide a source of easily studied reliable data. By means of these studies different experiments have been applied to different fields (sociolinguistics, psycholinguistics, dialectology, ethnolinguistics among them). One field which has developed this option extensively has been language teaching, whether as a mother tongue or as a foreign language. In fact, availability studies first appeared in the mid-20th century in response to concerns expressed in this area of applied linguistics. Nevertheless, rarely were the lists of available lexicon applied to teaching effectively.(Carcedo-González, 2000; Samper-Hernández, 2002; Bombarelli, 2005, among others).

In this work we propose a selection process of relevant lexicon based on the Fuzzy Set Theory applied to lexical availability (Ávila-Muñoz, 2016; Ávila-Muñoz and Sánchez-Sáez, 2010 and 2011). In general, this proposal allows us to differentiate between various levels of lexical compatibility relating to a center of interest or notional scope. Each level includes a number of lemmas that increase as the degree of compatibility decreases. Since the lists of lexical availability are usually obtained from native speakers, these mathematical parameters permit identification of lemmas that are readily accessible to these speakers and thus facilitate the selection of those essential to the design of foreign language word-learning programs. Obtaining a *mathematical compatibility index* of each word in the notional scope to which it is associated allows an objective selection of adequate lemmas at each learning level. The efficacy of the theoretical model proposed in this work is illustrated with material obtained from the analysis of lexical availability of Spanish in Malaga, Spain during the period 2005-2010.

Origins and Basic Principles

The idea of lexical availability came into being in 1951 when the French Ministry of Education created a committee to draw up a lexicon for teaching purposes (Gougenheim, Michéa, Rivenc & Sauvageot, 1954), the primary objective being the creation of a list of the most suitable words for students of French as a foreign language.

At the start of the research on lexical availability there was criticism of the lexico-statistical process of selection based solely on criteria of appearance frequency. In order to obtain the lexical frequency in a specific language, complex methodological processes are used to facilitate observation of the lexicon that have specific statistical stability (*see below* 5.1). In other words, the lists of frequency contain the lexicon to which speakers resort to build their messages, irrespective of the theme of the discourse. We know that despite the richness of a language, we tend to use a small number of lexical and structural units that are, obviously, the most frequent. Earlier research has shown that irrespective of the language studied, an educated person is considered to manage between four and five thousand words and that this figure decreases to two thousand in the case of the lesser (but by no means illiterate) educated individual (López-Morales, 1986, p. 59). The *frequency of use* variable, in any case, has shown itself to be one of the most important in language processing, be it from the perspective of production or of decoding. At the present time

dictionaries of lexical frequency in the major Western languages are available and the new technologies are aiding the appearance of new studies in this field that are applied in different ways, one of which is planning the lexicon for teaching native and second languages (Davies, 2005; Davies & Gardner, 2009; Lonsdale & Le Bras, 2009).

However, lists of lexical frequency contain almost exclusively "grammatical" lemmas and fail to include others of lexical content, thematic lemmas very common to any speaker as long as the requisite conditions for communication are satisfied to actualize them. In actual fact, the "non-grammatical" terms that are linked to a theme or specific matter designate beings and objects. In the majority of cases it is a question of words that appear almost always in groups and thus, the use of one encourages the use of others that will be repeated often when certain themes are broached.

Available lexicons may prove to be the adequate complement for compensating the irregularities observed in frequency repertoires because in them we find the reflection of the lexical flow released in a specific communcation situation. This type of list uses the concept of situation frequency that contrasts with that used by the lists of lexical frequence, focused on the objective of "frequent lexicon". In fact, the available lexicon makes sense in the maxim that certain often-used words in a given language are closely related to the appearance or not of specific topics.

To compile the vocabulary available —though not necessarily frequent— we commence by drawing up word association tests centered on stimuli or centers of interest. Specific vocabulary emerges around these notional nuclei that is presumed to be the potential lexicon belonging to the active vocabulary of the subjects and which they will use if a conversation follows the requisite channels. These lists generally comprise a limited total number of units, either because of speakers' delayed reactions or because the lists are closed once a predetermined number of words is reached. Words that appear at the top of lists are assumed to be those that come to mind first and are most readily available. Consequently, the available lexicon forms part of speakers' mental lexicon but is not generally updated in daily linguistic exchanges unless there is thematic specialization.

A simple and frequently revised mathematical formula based on the availability index is applied to the lists obtained. Broadly speaking, this index is a numerical parameter that endeavours to relate frequency and order criteria. The mathematical calculation attempts to equate the frequency with which a word is updated in a center of interest with its position in the different lists under study. The most available terms —that is to say, those which obtain an availability index closer to 1 in an interval between 0 and 1— are those which appear more often at the top of the lists submitted for evaluation.

Methodology

Our proposal is based on the results of lists of lexical availability compiled in Malaga city, Spain between the years 2005 and 2010, using the following methodological principles appropriate in research on lexical availability:

1. The sample was designed on the basis of representation criteria in a population stratified by age, sex and educational level (n=72) out of a total of 557.770 inhabitants in 2005, representing a sample/population ratio of 1/7746. The informants were selected according to the criteria laid down in the framework of the Study Project of Social Conditioning of Lexical Availability. (*Proyecto de Estudio del Condicionamiento Social del Léxico Disponible* CONSOLEX Project. Ávila-Muñoz & Villena-Ponsoda, 2010). We consider that the number of informants in this study is sufficiently representative to meet its principal objective: To show the efficacy of the theoretical model that serves to carry out an adequate selection of vocabulary to teach in foreign language classes. Nevertheless, we are aware of the advantages that a larger sample, made up of different diatopic and geographic varieties, would provide:

> A. Enlarging the sample —as demonstrated in earlier reseach— means that the formula used to calculate the availability is more stable and reliable.
>
> B. Enlarging the corpus with a more varied population sample is a requisite condition for establishing standards that demonstrate the benefits of the model cited above and its universal application to foreign language teaching as we discuss in section 6 of this article.

2. Many variables of post-stratification were taken into account as well. The sociological information was collected from a specific questionnaire completed by the participants after the text. In this, the subjects provided information referring to social, psycho-social and other variables on the frequency of their relationships in basic interactive networks.

3. Linguistic data, on the other hand, was collected according to normal processes used in the field of lexical availability studies: The informants were given two minutes for each center of interest in which to write, in columns on a blank sheet, all the words that came to mind related with the stimuli proposed by the surveyor.

4. The following centers of interest or notional nuclei were used: 01. Parts of the body; 02. Clothing; 03. Parts of the house (not including furniture); 04. House furniture; 05. Food; 06. Objects needed to lay a table; 07. The kitchen and kitchen utensils; 08. School: furniture and materials; 09. Heating, lighting and ventilation of a precinct; 10. The city; 11. The countryside; 12. Means of transport; 13. Agricultural and garden tasks; 14. Animals; 15. Games and amusements; 16. Professions and trades.

5. The completed lists were edited according to consensual criteria of lemmatization and the data bases underwent a process of statistical analysis that finally produced the lists of lexical availability. To do this free software distributed by Alcalá University and approved by the Cervantes Institute was used: *LexiDisp* is an application for Windows© that can calculate lexical availability. It is based on the from the mathematical formula developed by López-Chávez and Strassburger in 1987. Researchers also had access to an electronic portal through which to communicate new information to anyone interested in the subject: http://www.dispolex.com/ [last consulted: 6 October, 2016]. In actual fact, Dispolex is more than just a simple means of contact because the duly registered user may incorporate their material in a general data bank. According to its creators, this contributes to the shaping of a large Pan-Hispanic storehouse, shaped in such a way as to adapt to the characteristics of the different types of research. Furthermore it is possible to access, free of

charge, the tools needed to carry out the most commonly encountered calculations of lexical availability: availability index, frequency, appearance percentage, number of words, number of lemmas, informant averages, cohesion index and project comparisons.

Lexical Availablility and Prototypicality

Establishing a precise definition of the concept "available vocabulary" entails adopting an interdisciplinary perspective that integrates the language into the complex framework of human cognition. When an individual updates lists of available vocabulary he or she puts an individual cognitive task into motion in which various mechanisms of information processing are activated. Disciplines such as psycholinguistics, ethnolinguistics or cognitive linguistics become essential in capturing the true nature of the lists of available vocabulary. Of all these, psycholinguistics is perhaps the discipline that best aids understanding the nature of the available vocabulary; once again it is difficult to find a clear boundary between linguistics and psychology, at least when one considers the linguistic facts from the perspective of the individual speaker.

The availability index is an excellent indicator of the degree of prototypicality that words possess within each of the centers of interest. One should not associate the availability index to the nature of the lemmas because, to a great extent, this index is linked, as we shall show, to the collective conceptual classification. The availability of a word in a center of interest responds, essentially, to the concept of "accessibility". We assume that each center of interest revolves around a prototype created from the concept that determines this center of interest. This proposition stems from the development of the well-known prototype theory (*Prototypes theory*: Wittgenstein, 1953; Rosch, 1978; Lakoff, 1987). When an individual consents to take part in the lexical availability experiment, they access their lexical network using the prototype of the notional nucleus as the entrance doorway. Then they will make their way around the lexical network, an action that will lead them gradually away from the centre of the prototype (Example 1):

Example 1

Access and development of an individual lexical network. Subject 012. center of interest 05: Food

bread, meat, fish, cold cuts, cheese, milk, potatoes, fruit, melon, watermelon, peach, banana, pea, lettuce, yoghurt, tomato, apple, mandarin, pear, kiwi, cookie, coffee

As Example 1 shows, availability corresponds to the concept of "accessibility" from the center of a lexical nextwork and each individual offers his or her own path through the network. In our example the informant accesses the lexical network through the word *bread*, triggering the lexical framework. In the manner of a mesh, each word prompts another with which it is associated and so on until the speaker has exhausted his or her lexical network. According to our hypothesis, one might interpret that, *grosso modo,* the lists produced for each informant are a portrayal of the access to their particular vocabulary. Every individual, through their own psycho-social and educational evolution, possesses a different lexicon that, due to cultural and environmental factors however, is

likely to be similar to the rest of the members of the community. Obviously, obtaining the structure of this lexical network for each subject is an impossible task since it is presumed to be determined by a multitude of uncontrollable biographical factors. Furthermore, the dynamic, continually changing structure is kindled by the interaction of the speaker with his or her surroundings. Nevertheless, from the individual performances, we are able to estimate quantitatively the structure of accessibility to the lexicon for a population in a given center of interest. The quantification of this accessibility is what measures the concept of availability of each term in that center of interest, once all the information contributed by each and every individual member of the population is included.

The association between the words and the accessibility each one presents establishes how to represent the structure of the lexicon of a center of interest. In this representation the words closest to the notional nucleus prototype will show a greater accessibility value (a greater availability index).

Table 1 shows the lemmas found in the 02. *Clothing* center of interest with an availability index (AI/) > 0,1 and used by 30% of the sample.

Table 1
Words with Availability Index (AI) > 0,1 and Used by 30 % of the Sample 02. Clothing

LEMMA	Availability	% Occurence
1. *trouser*	0.69386	86.111
2. *shirt*	0.56530	70.833
3. *T-shirt*	0.46897	66.667
4. *sock*	0.46854	73.611
5. *skirt*	0.45956	58.333
6. *sweater*	0.36373	55.556
7. *coat*	0.31732	56.944
8. *shoe*	0.31189	55.556
9. *jacket*	0.27031	44.444
10. *panty*	0.26481	47.222
11. *bra*	0.23406	41.667
12. *underpants*	0.22593	37.500
13. *blouse*	0.21939	39.167
14. *dress*	0.20018	33.611
15. *scarf*	0.18220	34.722
16. *stocking*	0.16753	36.111
17. *tie*	0.15548	27.778
18. *swimsuit*	0.12347	23.611
19. *slipper*	0.12060	27.778
20. *jean*	0.11069	19.444
21. *tracksuit*	0.10879	22.222
22. *hat*	0.10651	19.444

From this it can be seen that the sixteen lemmas used by more than 30% of the population under study (% Occurrence) are those which attain a higher availability index. This leads us to confirm that the lemmas with the highest availability index in each center of interest would constitute the collective conceptual categorization of the stimulus. As the availability index of the lemmas decreases, and is thus used by less speakers, the lexical elements move away from the prototypical nucleus of the center of interest. Following this approach, and as a function of the values reached

by each lemma in its corresponding center of interest, it is possible to establish different lines of conceptual categorization certainty. These would thus comprise, at the first stage, the words closest to the prototypical nucleus of the center of interest under study. As the index of availability decreases and the lemmas move away from the center of the stimulus, the lexical elements begin to move closer to what we could call the "line of uncertainty", the limit at which the subject might begin to doubt whether the information furnished corresponds exactly to the proposed stimulus. The limit of uncertainty is, in fact, the threshold at which the speaker doubts if the word noted is sufficiently compatible or not with the prototype. The existence of this line of uncertainty becomes obvious if we consider that re-entry patterns exist in the lists of availability when informants perceive that they are moving too far away from the common prototypical nucleus. As we have seen in Example 1, the path through the lexical networks associated to specific notions (centers of interest) takes the subjects progressively away from what is considered to be the shared prototypical centre as the most accessible elements are used up. The subject usually regains access to the network when they consider they have moved too far away —line of uncertainty— and searches for a new point of entry that will bring them to a more central position with regard to the common prototype (Ávila-Muñoz and Sánchez-Sáez, 2010, pp 60-70).

The Fuzzy Set Theory

One of the main obstacles faced by a teacher in a foreign language class is, precisely, choosing the most relevant and suitable vocabulary according to the level of the students (Schmitt, 2010; Barcroft, 2015). The search for contrasted tools to choose a suitable lexical corpus appears to be necessary. Here the mathematical concept of "fuzzy set" (Zadeh, 1965, pp. 338-353; Zimmerman, 2001) can be of use to us. Broadly speaking, fuzzy sets are a generalization of the set theory in which the compatibility of the elements with the concept represented by the set is considered rather than the belonging of the elements. In this way it is possible to establish different levels of compatibility between an element and the set measuring this compatibility instead of the belonging-not belonging dichotomy peculiar to the classic set theory. In our case, this concept of "compatibility" corresponds to the concept of "accessibility". The interesting part of this approach does not lie in the determination of the availability process itself which amounts to a reformulation of previous models and is compatible with them. The essential is to transfer the problem to an established and contrasted mathematical framework that permits us to use the tools available. This, finally, will provide solutions to other problems arising from that already posed: the simple sorting of the words. Without doubt, one of these is to resolve finally and objectively the inconvenience involved in the precise selection of lexical elements suitable to different learning levels of foreign languages.

One of the tools that puts the fuzzy set theory within our reach is the assertion of the compatibility value characteristic of the fuzzy set, FEV (Fuzzy Expected Value) or its WFEV (Weighted Fuzzy Expected Value). By using this we can establish, on the one hand, a characterization limit of belonging values and, on the other, parameters to identify "very characteristic" or "hardly characteristic". It is a matter of proposing an objective cut off mark in the higher and lower levels of the fuzzy set that does not depend directly on the subjective perception of the researcher; such a mark would directly link the degree of compatibility of each element with a valuation of the set

of selected elements. This factor allows us to parametrize the differentiating process. The procedure for constructing the theoretical model capable of representing the fuzzy set used to distinguish the revelvant vocabulary at each stage of the process is explained step by step in detail in Ávila-Muñoz and Sánchez-Sáez (2010, pp. 59-63). This model has served to build a computer tool to calculate automatically the degree of belonging of a set of elements —words contained in the availability lists— to a specific set —notional centers.

Consequently, one of the most obvious applications is, as we have mentioned, to ascertain the relevant vocabulary for a center of interest. Up until now debate has been centered around different options that only offer justification of subjective approaches. Such approaches, while relevant in themselves are questionable for the scant formalization on which they are founded. Thanks to the *Fuzzy Expected Value* (FEV) we can establish the characteristic value of belonging in a fuzzy set with regard to a specific measure, which in our case will be the relative size of the sets of terms that exceed the compatibility level (cut-off sets). FEV is capable of proposing a value of belonging that establishes a balance between the number of terms that exceed it and the value itself. Furthermore, if we weight the relevance of size of the set cut off mark to a different degree, we can establish degrees of restriction inside the set that will allow us to construct "very representative" or "hardly representative" concepts.

If we apply this approach to a real example we can specifiy more accurately what the composition of the areas close to the prototypic nucleus will be and thus select lemmas suitable for each stage in the process of foreign language learning. The lemmas chosen would be those showing the highest degree of compatibility because, in our case, they would be readily available for all speakers and therefore be compatible with the set. In our example, the model constructed considers that of the 299 lemmas comprising the thematic nucleus 14. Animals, only forty-five (15% of the total) can be considered "very compatible" with the set they represent. The cut-off mark was fixed automatically in a compatibility index of 0.16 by the *Fuzzy Expected Value* (FEV) theoretical model. Table 2 shows these forty-five lemmas classified in six compatibility levels with only four forming the maximum prototypical nucleus: *dog, cat, lion, tiger* or in other words 1.33% of the set with a compatibility level of 0.89. From this point the remaining cut-off levels correspond to different areas that, while still very compatible with the referent stimulus, are moving progressively away from the central nucleus. Logically the inventory could be extended and encompass the different security limits until it reaches the uncertainty zone limit of the center of interest. In any event, the words classified in Table 2 prove to be the most compatible with the subject of the center of interest, or in other words, they imply the conceptual collective categorization of the proposed stimulus (referent set). Thus, they are only ones that, objectively, should be selected in the notional nucleus if we are to consider the inclusion of the lemmas most accessible to the native speaker population in our class material. By this means of lexical selection we eschew working with the particular vagaries of individual speech. In this case the use of objective and reliable mathematical tools allows us to focus solely on shared standard facts (Samper-Padilla, 1999, p. 554).

Table 2
Representation of the Collective Conceptual Categorization of the Prototype Center of Interest 01. The Human Body

LEVEL 1
dog, cat, lion, tiger
LEVEL 2
horse, elephant, cow, bull, bird, pig
LEVEL 3
giraffe, snake
LEVEL 4
mouse, hen, rabbit, leopard, monkey
LEVEL 5
canary, donkey, rat, zebra, duck, bear, whale,goat, cockroach, pigeon, eagle
LEVEL 6
Boar, ant, sparrow, parrot, dolphin, rhinoceros, fish, squirrel, bee, wolf, tortoise, chicken, budgerigar camel, shark, panther

Core lexicon and level according to the Common European Framework of Reference for Languages and the American Council for the Teaching of Foreign Languages

The European Year of Languages was celebrated in 2001 and the Council of Europe published the Common European Framework of Reference for Languages (CEFR), thus stressing the importance it gave to language learning in the sphere of the European Union.

From then on all member states focused on developing communication and interaction among their citizens in order to encourage mobility, cooperation and mutual understanding. The need to learn languages was stressed as the means to achieve these goals.

The objective of the Common European Framework of Reference for Languages is to serve as a guide for professionals in the field of language teaching to help design study programmes, draw up curricular material and establish universal guidelines to certify and evaluate linguistic competence in Europe.

In addition, the *American Council for the Teaching of Foreign Languages* presents speaking proficiency guidelines that describe the functions speakers are able to carry out at each of the established levels of language skills. These guidelines can be consulted for the contents, context, accuracy and type of discourse associated with the functions of each level.

If we focus on the task developed by the European Council we can see how this organism proposed a series of spheres of communicative relevance, which to a great extent, match the centers of interest used in research on lexical availability. Table 3 shows the breadth of this match (Bartol-Hernández, 2010, pp. 94-95).

Table 3
Comparison: Centers of Interest-European Council Topics and Sub-topics Source: (Bartol-Hernández, 2010, p. 95)

Centers of interest- Lexical Availability	Council of Europe Topics and Sub-topics
01. Parts of the body	7.1. Parts of the body
02. Clothing	9.3. Clothing and fashion
03. Parts of the house	2. Home
04. Furniture in the house	2. Home
05. Food and drink	10. Food and drink
06. Objects needed to lay a table	10. Food and drink
07. Kitchen and kitchen utensils	2. Home
08. School	8. Education
09. Lighting, heating and ventilation of a precint	
10. The city	9. Shopping 11. Public services 12. Places
11. The countryside	12. Places 2.8. Flora and fauna
12. Means of transport	5.1. Public transport 5.2. Private transport
13. Agricultural and garden tasks	
14. Animals	2.8. Flora and fauna
15. Games and amusements	4. Leisure, amusements
16. Professions and trades	1.10. Profession, work

With the aim of demonstrating the validity of our lexical selection model for teaching foreign languages we will use the *Plan Curricular (PCIC*) drawn up by the Spanish Cervantes Institute in 2006 based on the recommendations proposed in the European framework of reference. This plan was intended to specify and develop the levels of teaching Spanish as a foreign language. The work is divided into three volumes that correspond to basic user level (A1 and A2 according to *CEFR, Novice and Lower Intermediate* according to ACTFL), independent user (B1 and B2, *Intermediate Mid and Intermediate High*) and competent user (C1 and C2, *Advanced, Mid and Advanced High*). The *PCIC* also includes two notional inventories that bring together the specifications of contents related to the concept based on the notional-functional focus of the Seventies. These inventories include words as well as pluri-verbal lexical units, collocations and idiomatic expressions.

The first inventory, *General Notions* links the predominantly abstract lexical units that a speaker may use in any communication situation. The second, *Specific Notion*s is more relevant to our interests. It refers to specific situations related to a particular topic. By considering the centers of interest as notional spheres it is possible to compare them with the specific notions inventory. However, we have to admit that a comparison of both inventories shows that the centers of interest proposed in the classic studies of lexical availability cover less than half the spheres proposed in the *PCIC*. This suggests a need to renew and adjust the list of centers of interest that are used in the field of lexical availability (Bombarelli, 2005; Bartol-Hernández, 2010; Sánchez-Saus Laserna, 2011; Paredes-Garcia, 2014; González-Fernández, 2013. Despite the foregoing, the lack of adjustment of the centers of interest to the basic notions referring to daily life present in the *CEFR* do not invalidate the present study. Once the appropriate adjustments have been made, the selection of the lexicon can benefit from the proposal we are making as we will demonstrate in the following

section where we take as an example one of the centers of interest used in studies of lexical availability that matches one of the *PCIC* proposals: *Food*.

Core lexicon in the Cervantes Institute Curricular Plan

The *PCIC* specific notions divide the topic of *Food* into seven sections: diet and nutrition; beverages; food; recipes; dishes; kitchen and table utensils; restaurant.

Table 4 compiles the lexicon considered the most prototypic of the *Food* center of interest, organized according to the six compatibility levels resulting from the application of the model based on *Fuzzy Expected Value*. The third column indicates the level according to the *CEFR* and the fourth column the equivalent level in the ACTFL. The last column shows the specific topic assigned by the *PCIC*.

Table 4

Reference Level of Core Lexicon according to the PCIC. Center of Interest 0.5. Food

Lemma	Compatibility	*CEFR* Level	ACTFL	Specific *PCIC* Notion
		LEVEL 1		
bread	0.958877	A1	*Novice*	5.3. Food
		LEVEL 2		
milk	0.803341	A1	*Novice*	5.2.Beverages
tomato	0.803224	A2	*Lower Intermediate*	5.3. Food
meat	0.799878	A1	*Novice*	5.3. Food
		LEVEL 3		
lentil	0.692349	B1	*Intermediate Mid*	5.3. Food
potato	0.656898	A2	*Lower Intermediate*	5.3. Food
fish	0.634333	A1	*Novice*	5.3. Food
rice	0.610909	A2	*Lower Intermediate*	5.3. Food
chickpea	0.606576	B1	*Intermediate Mid*	5.3. Food
egg	0.599870	A1	*Novice*	5.3. Food
lettuce	0.588998	A2	*Lower Intermediate*	5.3. Food
		LEVEL 4		
pepper	0.554343			5.3. Food
vegetable	0.499876	A1	*Novice*	5.3. Food
onion	0.434353	A2	*Lower Intermediate*	5.3. Food
apple	0.423343	A2	*Lower Intermediate*	5.3. Food
orange	0.423230	A2	*Lower Intermediate*	5.3. Food
fruit	0.422985	A1	*Novice*	5.3. Food
oil	0.422873	B1	*Intermediate Mid*	5.3. Food
banana	0.419893	A2	*Lower Intermediate*	5.3. Food
cheese	0.419211	A2	*Lower Intermediate*	5.3. Foo
butter	0.408971	A2	*Lower Intermediate*	5.3. Food
pear	0.400021			5.3. Food
		LEVEL 5		
ham	0.399873	A2	*Lower Intermediate*	5.3. Food
melon	0.396565			5.3. Food
steak	0.395443	A2	*Lower Intermediate*	5.3. Food
watermelon	0.389854			5.3. Food
sugar	0.287899	A2	*Lower Intermediate*	5.3. Food
yogurt	0.278987	A2	*Lower Intermediate*	5.3. Food
omelette	0.277652	A1	*Novice*	5.3. Food
water	0.268987	A1	*Novice*	5.2. Beverage
pasta	0.265654	A2	*Lower Intermediate*	5.3. Food
chicken	0.259056	A2	*Lower Intermediate*	5.3. Food
carrot	0.258789	A2	*Lower Intermediate*	5.3. Food

salt	0.249091	A2	*Lower Intermediate*	5.3. Food
salad	0.248112	A1	*Novice*	5.3. Food
pulse	0.234904	B2	*Intermediate High*	5.3. Food
squash	0.233211			5.3. Food
		LEVEL 6		
chorizo	0.210091	B1	*Intermediate Mid*	5.3. Food
chocolate	0.199813	A2	*Lower Intermediate*	5.3. Food
coffee	0.198721	A1	*Novice*	5.3. Beverages
strawberry	0.198109	A2	*Lower Intermediate*	5.3. Food
spaghetti	0.187909			5.3. Food
macaroni	0.186491			5.3. Food
sausage	0.186555			5.3. Food
biscuit	0.175621	A2	*Lower Intermediate*	5.3. Food
salami	0.156741			5.3. Food
cereal	0.155542	A2	*Lower Intermediate*	5.3. Food
cake	0.149087	A2	*Lower Intermediate*	5.3. Food

Observe the relation of compatible lexicon with the elementary levels proposed by the *CEFR* and the ACTFL. Only *lentil, chickpea, oil* and *chorizo* correspond to level B1 or Intermediate Mid and *pulse* to B2 or Intermediate High.

Some lemmas do not appear as classified in the *PCIC*: *pepper, pear, melon, watermelon, squash*, *spaghetti, macaroni, sausage, salami* and *mandarin. Pepper* and *pear* at compatibility level 4 should have the same pedagogical consideration as others in the same cut-off level: *vegetable*, *fruit* (A1, *Novice*), *onion, apple, orange, banana, cheese, butter* (A2 *Lower Intermediate*) or *oil* (B1, *Intermediate Mid*) and be included in one of the levels where the latter are found. It is interesting, in this sense, to compare the compatibility indexes of *pepper* (0.20800) and *pear* (0.19019) with those of *vegetable* (0.20438) or *fruit* (0.19218) which, as we have mentioned are situated in level A1 and *Novice*. In the light of these results, our hypothesis is further corroborated when we see that even the term *pepper* is more compatible than the two included in the *PCIC*.

Similarly*, melon* (0.15528)*, watermelon* (0.13798) or *squash* (0.12458), situated at Level 5 cut-off mark should have the same consideration as others situated on the same level: for example, *salad* only reached a compatibility of 0.09245 but is, nonetheless, considered to be at Level A1 by the *PCIC*. The same can be said, obviously, of the lemmas found in Level 6 but not included in any of the levels proposed by the *PCIC*: *spaghetti, sausage, macaroni* and *mandarin.*

To substantiate these first results, we carried out a search in frequency indexes of the terms not included in the *PCIC*, comparing them with those that are contained in this *Curricular Plan*. Although few nouns normally appear in frequency lists, the inclusion of some of them indicates, evidently that their use is very frequent in the language being studied, regardless of the topic broached at any given moment (see § 2 of this work).

The most developed frequency dictionaries, those which include statistical weighted formulas such as typical deviation, do not merely present simple collections of words but rather a selection of those words that have shown, through the statistical methods used, to be the most stable and basic of the sample in question. These dictionaries usually accompany the frequency index with two other quantitative parameters: one measures the distribution frequency in the corpus under

consideration (*scattering index* with values oscillating between 0-1. Values closer to 1 indicate that the lemma is distributed homogeneously through the different genera that make up the corpus); the other weights the equitable share of frequencies upward, in detriment to those lemmas that show greater frequency but unequal distribution (l*exical use factor*). Table 5 shows these three statistical factors linked to the frequency of use of each lemma listed in Table 4.

The indexes used to create Table 5 have been extracted from a study carried out by Ávila-Muñoz (1999) from a speech corpus composed of over one thousand five hundred words. This data base was divided into four types of discourse with the aim of obtaining the most realistic image possible of daily speech activity (bi-directional, face to face, unregulated distance bi-directional, regulated distance bi-directional and unidirectional. Ávila-Muñoz, 1999, pp. 17-48).

Table 5
Frequency Indexes of Core Lexicon according to PCIC

Lemma	Use	Frequency	Dispersion
	LEVEL 1		
bread	130.2756	276.8543	0.4705
	LEVEL 2		
milk	47.2410	70.7747	0.6674
tomato	1.78951	8.3264	0.2149
meat	22.3540	47.8770	0.4669
	LEVEL 3		
lentil	1.60789	6.24483	0.2574
potato	5.92849	24.9793	0.2373
fish	47.2707	66.6115	0.7096
rice	15.8270	27.0609	0.5848
chickpea	7.06941	16.6528	0.4245
egg	13.51216	37.4690	0.3606
lettuce			
	LEVEL 4		
pepper	0.00000	14.5712	0.0000
vegetable	0.98225	14.5712	0.0674
onion			
apple			
orange	6.08994	16.6528	0.3656
fruit	5.63484	14.5712	0.3867
oil	5.87888	49.9586	0.1176
banana	1.34250	8.3264	0.1612
cheese	0.00000	6.2448	0.0000
butter	8.69532	22.8977	0.3797
pear			
	LEVEL 5		
ham			
melon			
steak			
watermelon			
sugar	5.57903	16.6528	0.2149
yogurt	3.20000	6.24482	0.0320
omelette	7.43685	12.4896	0.5954
water	161.02391	266.4461	0.6043
pasta	0.00000	6.24483	0.0000
chicken	4.66473	8.3264	0.3200
carrot			
salt	0.00000	14.5712	0.0000

salad	4.36461	4.1632	0.0875
pulse			
squash			
	LEVEL 6		
chorizo	0.00000	6.2448	0.0000
chocolate	7.52485	16.6528	0.4518
coffee	64.12512	139.4679	0.4597
strawberry	0.00000	4.1632	0.0000
spaghetti			
macaroni			
sausage			
biscuit	6.10103	12.4896	0.4884
salami			
cereal			
cake	0.00000	4.1632	0.0000
corn	0.00000	8.3264	0.0000
mandarin	0.57616	12.4896	0.0461
tuna	3.80479	39.5506	0.0962

The first observation that springs to mind from Table 5 is that the lemmas *water* and *coffee* (usage factor 161.02391 and 64.12512 respectively) are nevertheless situated in compatibility levels 5 and 6. This is logical if we bear in mind that the notional sphere where these terms appear is that of food and not that of beverages. Despite this fact, their presence in lists of lexical availability indicates that they are lemmas widely used in spoken Spanish.

Apart from these exceptions, the frequency indexes shown support our first results:

Level 1: *bread* is the term with highest use indicators in the table (with the exception of *water*). Its inclusion in this cut-off level appears to be justified.

Level 2: *milk* and *meat* are lemmas with high markers of use, frequency and dispersion. In fact, only fish, situated at the following compatibility level presents higher values. *Tomato* also appears at this level with lower frequency levels than in earlier indexes.

Level 3: *lettuce* does not appear in the frequency list under consideration. With the exception of the aforementioned *fish* and also *lentil* —the latter with low frequency indexes— the average use of terms contained in the list is 10.58427.

Level 4: The use average at this level is considerable lower (4.7706). *Onion, apple* and *pear* are not included in the frequency list consulted. *Pepper* and *cheese*, however do appear: the former with a frequency of 14.5712 and the latter with 6.2448. As these were collected from only one of the subcorpora under consideration, dispersion is 0 and therefore, index of use is also 0.

Level 5: *ham, melon, steak, watermelon, carrot, pulse* and *squash* are not present in the glossary consulted. With the exception of the aforementioned *water*, the average use of the lemmas that do appear is 4.2075. The terms *pasta* and *salt* appear with a 0 use index since once again their dispersion is also 0.

Level 6: *Spaghetti, macaroni, sausage, salami* and *cereal* are missing from the frequency corpus consulted. *Chorizo, strawberry, cake* and *corn* appear with a index use 0 for the same reason. With the exception of *coffee* referred to, the average use of the rest of the lemmas is 4.

Consequently, and in view of the results presented in Table 5, we observe that, in a general way, the frequency indexes support the lexical selection method proposed: the lemmas with the highest values of use, frequency and scattering are included in the first cut-off levels. The average use at these levels decreases as they move away from the prototypic core, as shown in Graph 1. It must be remembered that dictionaries of lexical frequency do not favour the appearance of lemmas related to specific notional fields. This reason justifies the fact that some of the terms contained in the different cut-off levels established by the model used in our investigation are not present in the frequency list consulted. Nevertheless, the appearance of some, although with low indicators, alert us to the need to take them into account in the planning of lexical teaching. In this sense, it must be pointed out that our model permits the inclusion of some frequent lemmas that are not included in the *PCIC*: *pepper, mandarin*.

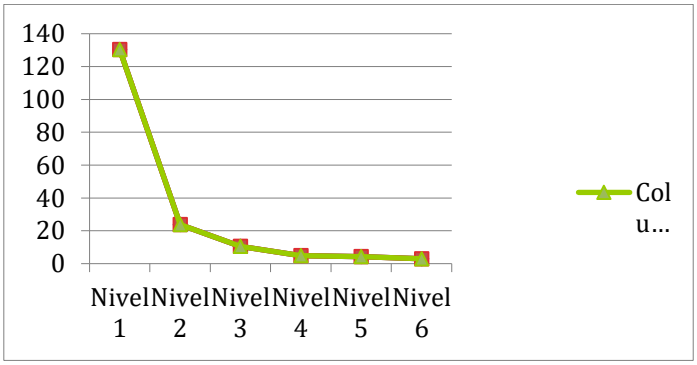

Figure 1. Average use of Lemmas Contained in Different Cut-off Levels
Source: Ávila-Muñoz, 1999

Discussion

Data base proposal

Our proposal is based solely on data obtained from the research carried out in the city of Malaga. Therefore, since it deals with a geographically marked language variety —Spanish spoken in the south of Spain— it must be remembered that the general lexicon material for foreign language teaching must be chosen with varied geolectal criteria. Doubtless, a more diverse and robust corpus that includes material from different geographical areas is necessary. Thus we will be sure of working with a shared standards lexicon. Possibly then, lemmas such as l*entil* or *chickpea* will enjoy a different consideration to that mentioned in previous paragraphs.

At all events, in the case of Spanish, building an international data base designed to facilitate the problem of lexicon selection for teaching general Spanish is possible and, in fact, is already under way. The goal of the Pan-Hispanic Research Project on Lexical Availability (PPHDL Proyecto Panhispánico de estudio sobre la Disponibilidad Léxica) is the creation of a corpus of these characteristics. Sponsored by Humberto López-Morales, el PPHDL intends to obtain a dictionary of available lexicon of Spanish, the methodological basis of which is the preparation of lists of lexical availability in different Spanish speaking areas around the world. At the present time this project is at a very advanced stage of execution. The general data base of this macro project contains samples of available lexicon from the majority of countries and regions where Spanish is spoken either as a mother tongue or as an emerging linguistic variety with a great social impact (such as, for example the situation of Spanish in the United States of America). Thanks to a solid proposal to unify methodological design criteria, the PPHDL encourages the exchange of data and the development of comparative studies between the results of the different local projects. As we have previously mentioned, a reader interested in the subject may consult the present state of the matter on http://www.dispolex.com/. Among other things, at this virtual meeting point, exhaustive information on the epistemic fundamentals of the investigation is gathered, together with the precise method of data collection in those areas where lexical samples have been or are being collected.

Once the general data base is completed it will be possible to apply the theoretical model defended in this work with sufficiently broad data to cover the wide spectrum of what we know as common Spanish.

Teaching Programing

The espistemic fundamentals seen so far infer the need to create proposals that allow teachers of foreign languages to put into effect coherent teaching programs that include a vocabulary suited to the level of the linguistic competence they wish to teach. This programing should contemplate the following phases or stages:

1. Accurate selection of centers of interest. As we have seen during our developing work, in the case of Spanish as a Foreign Language (SFL) one possibility would be to work with the 20 specific notions of the *PCIC* which are used for the materials and subsequent evaluations for the Cervantes Institute Diplomas (DELE) for Spanish as a Foreign Language: 1. Individual: physical dimension; 2. Individual: perceptive dimension and mood; 3. Personal identity; 4. Personal relationships; 5. Food; 6. Education: 7. Work; 8. Leisure; 9. Information and the media; 10. Home; 11. Services; 12. Shops and shopping; 13. Health and hygiene; 14. Travel, accommodation and transport; 15. Economy and industry; 16. Science and technology; 17. Government, politics and society; 18. Artistic activities; 19. Religion and philosophy; 20. Geography and nature. Obviously, to achieve an adequate presentation of these notions in the form of stimuli for the informants, some notions will have to be reformulated. To do this in the *PCIC* each one has specific sub-sections to permit more stimuli and avoid ambiguities. For example the specific notion 6. Education includes the following specific sub-sections: 6.1. Educational institutions and centers; 6.2. Teachers and students; 6.3. Educational system: 6.4. Learning and teaching; 6.5. Examinations and grades; 6.6. Studies and grades; 6.7. Classroom language; 6.8. School materials and classroom furniture. To

cover each of these sub-sections one proposed stimulus for informers participating in the process of obtaining lists of lexical availability could be: *school, furniture, materials, personal, institutions, teaching-learning process*.

2. Calculating lexical availability. This index indicates if the word is accessible to the informants by means of a formula that includes the word position in the list and the number of informants who have written it down. Thus, we should obtain a overall picture of the collective prototype studied and the terms most directly related with the notional sphere in which they are gathered.

3. Calculating the compatibility index. This marker shows the distance in relation to the prototype and, therefore, its accessibility. This allows us to obtain a picture of the lemmas most compatible with the notional field.

4. Compatibility level. This detail allows us to emphasize the basic words and divide them into six levels that would be adjusted to the levels A1 and A2 of the CEFR and Novice and Lower Intermediate of the ACTFL.

5. Assigning of corresponding level both in the CEFR and ACTFL. In order to assign each lemma to a teaching level the teacher of foreign languages must be aware at which level of the CEFR or ACTFL each of the lemmas included in the lists of availability corresponds.

6. Revising the classification of some lemmas included in the curricular plans. In our example, once again we could complete the level according to the CEFR and the ACTFL with words that do not appear in the *PCIC* and reassign another level to some lemmas, such as *tomato* which, although very compatible, is considered to be at level A2 or Lower Intermediate.

7. Adjusting to learning objectives. The most compatible lemmas should be those chosen for teaching at basic levels. Then, in turn, new lemmas should be introduced at higher levels as their level of compatibility decreases. By the same token, teachers should be aware that the availability lists contain regionalisms, dialectalisms and technicisms. These lemmas should be marked in some way, especially if we wish to re-use our lists in different localities. For example if *ajoblanco* is marked as a typical dish in certain regions of southern Spain, a Mexican teacher would probably not include it for their students at basic level.

Conclusions

Our research has used lexical availability studies as a starting point. From these we have observed how community prototypes are created and how the lemmas can become more or less compatible with the collective prototype associated with specific notions. Our final purpose was to present an instructive proposal capable of making a suitable lexical selection adjusted to the teaching levels at which we work. Our pedagogic proposal, therefore, employs solid, theoretical fundamentals and analysis that endorse it.

To justify the said proposal, and given the importance of the publication by the European Council in 2001 of the Common European Framework of Reference for Languages, we compared the specific notions proposal in the Cervantes Institute Curricular Plan with the centers of interest. We saw how the centers of interest with which lexical availability traditionally works only cover certain sections of the notions essential to language learning. Despite this, we observed the similarities and selected the center of interest 02 Food as an example. Using that data we determined the level according to both CEFR and ACTFL at which each of the lemmas appearing in the first six compatibility levels are situated according to the results offered objectively by the procedure based on the Fuzzy Expected Value. As expected, the majority of the lemmas belong to basic user level (A1-A2; *Novice-Lower Intermediate*) confirming the hypothesis that core lexicon determines the basic vocabulary of a language. By contrast, there are some divergencies between the *PCIC* proposal and the lemmas included in the compatibility lists. Some of these indicate that there are very compatible elements that do not appear even in the lower levels proposed by the *PCIC*. Others refer to a possibly inappropriate inclusion of lemmas at levels which, according to the compatibility calculation, should be situated on different scales of learning.

To forestall these inconveniences, we propose the creation of a specific data base for teaching purposes for each of the foreign languages to be taught. This data base should help foreign language teachers to select the lexicon according to the students' level and learning objective under consideration. The selection will be based on the compatibility index of each term within its notional sphere —that is to say its degree of accessiblity for native speakers.

The final result of this procedure should be the creation of an accessible data base that contains each of the lemmas included in the different centers of interest or notions studied. This list of lexical units must also include the name of the center of interest or notion in which it appears, the lexical availability index, the lexical compatibility index and, as a function of these parameters, the corresponding CEFR and ACTFL levels. In this way the foreign language teacher will have at hand all the parameters necessary for selecting, at any given time, the lemmas best suited to the needs of their students. And, even more important, the model we propose favours functional learning adapted to the real needs of students.

References

American Council on the Teaching of Foreign Languages (2012). *Performance Descriptors for Language Learners.* Alexandria, VA: ACTFL.
https://www.actfl.org/sites/default/files/pdfs/PerformanceDescriptorsLanguageLearners.pdf.

Ávila-Muñoz, A. M. (1999). *Léxico de frecuencia del español hablado en la ciudad de Málaga*. Málaga: Universidad de Málaga.

Ávila-Muñoz, A. M. (2016). El léxico disponible y la enseñanza del español. Propuesta de selección léxica basada en la teoría de los conjuntos difusos. *Journal of Spanish Language Teaching*, 3, 31-43. DOI: 10.1080/23247797.2016.1163038

Ávila-Muñoz, A. M. & J. M. Sánchez-Sáez (2010). La disponibilidad léxica. Antecedentes y fundamentos. In A. M. Ávila-Muñoz & J. A. Villena-Ponsoda (Eds.), *Variación social del léxico disponible en la ciudad de Málaga* (pp. 35-82). Málaga: Sarriá.

Ávila-Muñoz, A. M. & J. M. Sánchez-Sáez (2011). La posición de los vocablos en el cálculo del índice de disponibilidad léxica: procesos de reentrada en las listas del léxico disponible de la ciudad de Málaga. *Estudios de Lingüística de la Universidad de Alicante (ELUA),* 12(1), 133-59.

Ávila.Muñoz A. M. y J. A. Villena-Ponsoda (eds.) (2010). *Variación social del léxico disponible en la ciudad de Málaga.* Málaga: Sarriá.

Barcroft, J. (2015). El método IBI en la enseñanza del léxico: teoría, investigación y nuevas perspectivas. *Journal of Spanish Language Teaching,* 2 (2), 112–125. DOI: 10.1080/23247797.2015.1105512.

Bartol-Hernández, J. A. (2010). Disponibilidad lexica y selección del vocabulario". In R. M. Castañer-Martín & V. Lagüéns-Gracia (Coords.) *De moneda nunca usada. Estudios dedicados a José María Enguita Utrilla* (pp. 85-107). Zaragoza: Instituto Fernando el Católico.

Bombarelli, Á. (2005). *La disponibilidad léxica como herramienta didáctica: una propuesta de selección del vocabulario para un nivel umbral de ELE* (Memoria de Máster inédita). Salamanca: Universidad de Salamanca.

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

Consejo de Europa. (2002). *Marco común europeo de referencia para las lenguas: aprendizaje, enseñanza, evaluación*. Madrid: MECD. http://cvc.cervantes.es/ensenanza/biblioteca_ele/marco/. [consulta 19 de noviembre de 2015].

Davies, M. A. (2005). *Frequency Dictionary of Spanish Core Vocabulary for Learners*. Abingdon: Routledge.

Davies, M. A. & D. Gardner (2009). *Frequency Dictionary of Contemporary American English: Word Sketches, Collocates and Thematic Lists*. Abingdon: Routledge.

González-Fernández, J. (2013). *La disponibilidad léxica de los estudiantes turcos de español como lengua extranjera* [en línea]. MarcoELE 16. Disponible en http://marcoele.com/disponibilidad-lexica-de-estudiantes-turcos/ [consulta 19 de noviembre de 2015].

Gougenheim, G., R. Michéa, P. Rivenc & A. Sauvageot. (1954). *L'élaboration du français élémentaire. Étude sur l'établissement d'un vocabulaire et d'une grammaire de base.* Paris: Didier.

Instituto Cervantes. (2006). *Plan Curricular del Instituto Cervantes*. Madrid: Biblioteca Nueva. http://cvc.cervantes.es/Ensenanza/Biblioteca_Ele/plan_curricular/default.htm. [consulta 19 de noviembre de 2015].

Lakoff, G. (1987). *Women, Fire and Dangerous Things: What categories reveal about the mind*. Chicago-London: The University of Chicago Press.

Lonsdale, D. & Y. Le Bras (2009). *A Frequency Dictionary of French: Core Vocabulary for Learners*. Abingdon: Routledge.

López-Chaves, J. & C. Strassburger 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* (pp.20-41). México: UNAM.

López-Morales, H. (1986). Sociolingüística. Madrid: Gredos.

Paredes-García, F. (2014). A vueltas con la selección de 'centros de interés' en los estudios de disponibilidad léxica: para una propuesta renovadora a propósito de la disponibilidad léxica en ELE [en línea]. *Revista Nebrija de Lingüística Aplicada a la Enseñanza de Lenguas,* 16. Disponible en http://www.nebrija.com/revistalinguistica/a-vueltas-con-la-seleccion-de-centros-de-interes-en-los estudios-dedisponibilidad-lexica-para-una-propuesta-renovadora-a-proposito-de-ladisponibilidad-lexica-en-ele [consulta 19 de noviembre de 2015].

Rosch, E. (1978). Principles of categorization. In E. Rosch & B. Lloyd (Eds.) *Cognition and Categorization* (pp. 27-48). Hillsdale: Lawrence Erlbaum.

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

Samper-Padilla, J. A. (1999). Léxico disponible y variación dialectal: datos de Puerto Rico y Gran Canaria. In A. Morales, E. Forastieri, J. Cardona y H. López Morales (Eds.) *Estudios de lingüística hispánica. Homenaje a María Vaquero* (pp. 550-73). San Juan de Puerto Rico: Universidad de San Juan de Puerto Rico.

Sánchez-Saus Laserna, M. (2011). *Bases semánticas para el estudio de los centros de interés del léxico disponible. Disponibilidad léxica de informantes extranjeros en las universidades andaluzas*. (Tesis doctoral). Cádiz: Universidad de Cádiz. Disponible en http://rodin.uca.es/xmlui/handle/10498/15862 [consulta 19 de noviembre de 2015].

Schmitt, N. (2010). *Researching Vocabulary: A Vocabulary Research Manual*. Basingstoke: Palgrave Macmillan.

Wittgenstein, L. (1953). *Philosophical Investigations*. New York: McMillan.

Zadeh, L. A. (1965). Fuzzy sets. *Information and Control,* 8, 338-53.

Zimmermann, H. J. (2001). *Fuzzy Set Theory and its Applications*. Boston: Kluwer Academic Publishers.

Antonio Manuel Ávila Muñoz has PhD in Spanish Studies and Linguistics and he is a Senior Teacher and Lecturer of General Linguistics at the University of Malaga, Spain. His interest in Sociolinguistic studies has mainly been focused on works and publications about several sub-disciplines such as social dialectology, linguistic variation, corpus linguistics, cognitive sociolinguistics and lexical statistics. He has published several books on sociolinguistics and lexical analysis, as well as specific works in periodical publications and magazines (such as *Review of Cognitive Lingüistics*, *International Journal of Spanish Language, Spanish in Context, ELUA*, *Oralia*, among others) and chapters in different books (John Benjamins; Honoré Champion).