Journal of Research in Applied Linguistics
ISSN: 2345-3303 – E-ISSN: 2588-3887 – https://rals.scu.ac.ir

Please cite this paper as follows:

Agustín-Llach, M. P. (2023). Foreign language semantic categorization: Evidence from the semantic network and word connections. *Journal of Research in Applied Linguistics*, *14*(1), 205-222. https://doi.org/10.22055/RALS.2023.18077

Research Paper

Foreign Language Semantic Categorization: Evidence From the Semantic Network and Word Connections

Maria Pilar Agustín-Llach

Department of Modern Philologies, Faculty of Letters and Education, University of La Rioja, Logroño, Spain;
maria-del-pilar.agustin@unirioja.es

Received: 20/01/2022 Accepted: 28/11/2022

Abstract

Categorization is an essential cognitive activity to make sense of the world. Category-generating and word association studies help us elucidate how learners organize words in the mind. Categorization can indicate how learners acquire new vocabulary and how they use it to interpret reality. The present study analyses the word retrieval mechanisms of two groups of learners by looking into the word-pair associations produced as a response to a category-generation task. These associations in response to two semantic categories: *countryside* and *animals* were classified and scrutinized for shared patterns. Differences and similarities between semantic categories, and proficiency level of learners were explored. Results revealed learners favour coordination as the preferred mechanism for word attachment and retrieval, followed by event-based associations (metaphorical/ contextual and experiential knowledge). Although frequencies vary significantly depending on L2 proficiency and semantic category, associational behaviours are comparable, with a preference for semantic associations over formal and presence of both taxonomic and thematic associations. The implications of these findings can reach the EFL classroom by informing which words to teach together and through which strategies.

Keywords: Word Associations; Semantic Network; Semantic Categories; Lexico-semantic Attachment; Lexical Learning.

1. Introduction

Semantic categories have been the base of different semantic fluency tests as they are used as cues to generate participants' responses and associations (cf. Tomé Cornejo 2015). The categorization process is acknowledged to be a cognitive mechanism based on the organization and classification of the information obtained from the outside world and experience (cf. Sass et al. 2009, Coni et al. 2019). To make sense of the world outside human beings classify and organize information based on equivalence assumptions (cf. Tomé Cornejo 2015: 245).

In this sense, research has posed the network metaphor to account for how words are connected and stored in the human mind (see e.g. Fitzpatrick and Thwaites 2020). Namely, words establish relationships among themselves based on semantic similarity forming a network; when recalled, the most accessible words, e.g. prototypes, are activated and in turn they activate other related words by calling into their shared semantic features, their collocational proximity, their functional/ metonymic relations, and so on (cf. e.g. Fitzpatrick 2013). The type of connections that are established among words is relevant in order to determine the degree of accessibility of the words and to look into the learner variables that determine those connections, such as proficiency (e.g. Zareva 2007).

Network analysis through multiple response semantic fluency and categorization tasks provides us with valuable qualitative information concerning a) the type of relations that structure the extended semantic neighbourhoods (i.e. within semantic categories), related to that b) how the growth of the lexicon happens through different types of attachment, and c) how the words are recalled to production. Guided by the assumption that [r]*etrieving one category member produces a spread of activation to other category members facilitating their later retrieval* (Collins and Loftus 1975: 419), the present study carries out an analysis of the links that attach one response to the next with the objective of finding out the nature of the connections in the learners' mental lexicon; how proficiency affects the nature of these connections, and

also to explore the impact of different semantic categories in the type of word connections established. This is the main significance of the paper, and to our knowledge, the first to try to answer such concerns with L2 learners.

In spite of the above, the next literature review section will show lack of definite findings concerning L2 categorization and lexico-semantic connections in the mental lexicon of Foreign Language (FL) learners, as well as how semantic network theory and spreading activation theory have been rarely used to explore semantic categorization and word connections in the FL despite their big potential.

Within the framework of the spreading activation theory, network theory and categorization theory, we focus on exploring the nature of the associations made by EFL learners at two different proficiency levels in order to ascertain the role that both proficiency level and stimulus category have on the process of categorization. Specifically, we want to answer to a) What is the nature of the connections in the learners' mental lexicon?, b) How does proficiency affect the nature of these connections?, and c) What is the impact of different semantic categories in the type of word connections established? These questions address the problem of how L2 learners at different proficiency levels form their semantic categories and how these are internally structured. Discerning the aspects that learners' focus on in order to establish connections in their minds has further significance as insight into lexical learning processes. Specifically, the significance of this study lies on the novelty of looking into semantic categorization and the internal structure of semantic categories of L2 learners at different proficiency levels.

2. Literature Review

Words are stored in the mental lexicon which is located within semantic memory. But words are not stored randomly, they follow a certain structure. Collins and Quillian (e.g. 1969) argued that words are connected in a tree-like hierarchical way. These words, represented as nodes, are connected via links that lead, first, to essential nodes, and then, spread to other less essential or less relevant nodes or information, which in turn also activate their own related nodes. Links represent semantic or lexical relationships. In elaborating on this previous model, Collins and Loftus (1975) proposed a spreading-activation model based on semantic networks. Specifically, they elaborated on several assumptions. First, they believed that activation spreads in a decreasing gradient, which is related to the accessibility and strength of the connections. The firstly activated nodes would be most accessible and as their accessibility decreases so does their activation. Second, they argued that activation is serial, activating one node at a time spreading over a series of nodes. Third, activation decreases over time and intervening activity; fourth, activation from different sources can summate. They also believed that the organization of the mental lexicon operates on semantic similarity, with semantically similar concepts being stored closely together with strong links. The names of the concepts, on their part, form a lexical network where formal similarity also plays a role.

Although Collins and Loftus and their associates conducted several empirical research studies to probe their theory (e.g. 1969, 1975), we have no notice of previous empirical research that has dealt with this model and its application within the field of second language acquisition. However, grasping how learners organize the words of the FL in the mind is a major undertaking to gain further insight into how lexical acquisition happens in the foreign language. Through word associations with semantic categories, we intend to address that issue. May this stand as the statement of the problem we intend to deal with in the present study.

The multiple responses fluency task, such as the one used here, seems an adequate instrument to capture semantic structure network as proposed by Collins and Loftus (1975) and their assumptions on spreading activation[1] (cf. Ferreira and Echeverría, 2010; Sánchez-Saus Laserna, 2016; Šifrar Kalan, 2014). In a semantic categorization task a participant is asked to produce in a certain amount of time the words that come to their mind after receiving the prompt, generally representing a semantic category, e.g. *animals, the countryside, food and drink*, just to name but a few (e.g. Bartol Hernández, 2006, Jiménez Catalán 2014, Borodkin et al 2016). This procedure presents several advantages. First, as it is a free association test with no restrictions, it stands as a powerful instrument to look into the structure of the semantic representations and the associations of the words within a semantic category. It provides an accurate access to the organized knowledge stored in the mental lexicon, which is crucial for successful interaction and communication (cf. Karuza et al. 2016, Gudmundson 2020). Second, the presentation of a stimulus word activates other words, which are retrieved from the semantic memory in a chain-like way. Responses might be triggered by the prompt but also by previous responses. The production of these words allows us to observe the relationship and the connection between them and the

prompt, and interestingly also among the responses themselves, i.e. we can get an insight not only into the stronger associations but the weaker ones too (De Deyne, Kenett, Anaki, Faust and Navarro, 2017).

Word association (WA) and verbal fluency tasks have been traditionally used, indeed, in the literature to examine the complex network of the bilingual mental lexicon. Specifically, this type of task has been used in order to compare the organization of the mental lexicon in NL and in FL and to examine emergent lexical organization and development in FL (e.g. Aitchison 1994; Fitzpatrick, 2007; Hernández Muñoz and López García 2014, Peppard, 2007; Pranoto and Afrilita, 2018; Roux, 2013; Séguin, 2015; Zareva, 2007), with the aid of mathematical tools, theories and models to make sense of and systematize data and results (cf. Steyvers and Tennenbaum 2005, Vitevich 2008, Borge and Arenas 2010, Morais et al. 2013). Identification of the associative connections can reveal not only the structure of the networks, but it can also help elucidate the occurrence of the different types of semantic information in linguistic processing (cf. Hernández Muñoz and López García 2014). Networks with high clustering coefficient and short global distance or path length, where nodes are densely connected to each other and it takes few steps to travel the whole network, are easier to navigate, i.e., words are accessed more easily and quickly, the speaker can go from one node/lexical item to another using few steps. Similarly, network organization can be described in terms of community structure. A network with communities with nodes densely linked within the same community is easier to navigate than when nodes of one community are linked to nodes outside the community. These types of networks have been observed in monolingual adults (Borge-Holthoefer and Arenas, 2010, Steyvers and Tenenbaum, 2005).

The present study is framed within the key notion of the network metaphor identified by Fiztpatrick and Thwaites (2020) concerning WA studies (the other two being the influence of specific lexical variables and models pf psycholinguistic processing). It intends to provide a better understanding of second language acquisition and teaching by using sophisticated mathematical and technological developments.

WA research has become increasingly popular with applied linguists dealing with second language acquisition (cf. Fitzpatrick and Thwaites 2020 for a very thorough review of WA research studies). In order to examine how learners store words in their mental lexicon, different formats of word association tests (WAT) have been used. The most frequent one asks learners to react to a stimulus or prompt with a single response. However, we also come across tests where learners are required to respond to the cue with several answers (e.g. Sheng et al. 2007). This last type of test is believed to better capture the nature of the variation of responses regarding their prototypicality (e.g. Schmitt 1998, Hernández Muñoz 2014), to allow measurement of both storage and accessibility of relations (Sheng et al 2007), and to provide faithful access to semantic knowledge and semantic network organization (Hernández Muñoz and López García 2014) (a more detailed account of methodological approaches within WATs is beyond the boundaries of this study, but see Fitzpatrick et al. (2013) and Fitzpatrick and Thwaites (2020) for a precise and fine-grained review of WA methodologies). The data yielded by WATs lend themselves to easy quantification and can inform about different aspects of the lexicon.

Taxonomies of WAs have been going on for a long time in the literature (e.g. Aitchinson 1994, Murphy 2001, Murphy and Lin 2001, Mirman et al. 2017, Fitzpatrick and Thwaites 2020). They generally distinguish among two main types of relations a) formal or clang associations based on formal (written or oral) similarity and b) semantic associations based on meaning relations. Semantic associations have proved to include myriad distinctions: shared semantic features, categorial relations (paradigmatic, base word/ coordinate, superordinate, subordinate relations), collocational proximity or position-based connections, functional/ metonymic relation, metaphorical connections (spatial, causative), paradigmatic connections based on synonymy and antonymy, connections based on word knowledge, cultural and emotional experiences, or encyclopaedic knowledge (Cf. Hernández Muñoz and López García 2014, Fitzpatrick 2006, 2007, Zareva 2007, Precosky 2011).

The assumption behind the data provided by WATs is that they reveal how speakers organize the words in the mind, and how they structure the mental representations and concepts (cf. Zareva 2007). Quantitative and qualitative analyses have followed to examine a) number of responses (also frequency of response) (e.g. Zareva 2007), b) types of associations between cue and response (e.g. formal, paradigmatic, syntagmatic, encyclopaedic, etc.) (e.g. Meara 1982, Wolter 2001), and c) prototypicality of responses (e.g. Schmitt 1998, Higginbotham 2010, Shivabasappa et al. 2017). Hence, learners are characterized by producing fewer and more heterogeneous responses (Zareva 2007), more syntagmatic responses, more clang or formal associations (Meara 1982, Wolter 2001, Precosky 2011), and responses with looser conceptual or contextual links (Fitzpatrick 2006). More advanced learners produce similar numbers of responses

to the stimuli but they are more heterogeneous than natives (Zareva 2007), they produce more paradigmatic responses than intermediate learners (Wolter 2001), more non-equivalent meaning connections (Fitzpatrick and Izura 2011), and more collocational answers (Precosky 2011). Familiarity with the target words also affects the type of association established, with novel words giving rise to formal associations (Namei 2004, Chaffin 1997), or to definitional, syntagmatic, (Chaffin 1997); and known or more familiar words eliciting more paradigmatic (Namei 2004) and event-based responses (Chaffin 1997). Chronological age, age at acquisition, L2 proficiency, culture, and cognitive development are the factors alluded to explain the differences in association behaviours (Ji et al. 2001, Peña et al. 2002, Namei 2004, Zareva 2007, Comesaña et a. 2009, Fitzpatrick and Izura 2011, Fitzpatrick et al. 2013).

In the specific realm of WA data generated from semantic categorization tasks, similar to ours, Hernández Muñoz and López García (2014) found that semantic similarity was the most frequent type of association in responses to several category-like cue words. Superordinate and subordinate relations followed in frequency with collocational proximity being the least frequent type of association selected by their adult native speakers. In a more specific study with younger native speakers at two school grades, Henríquez Guarín et al. (2016) observed that in responding to the category "human body" older speakers preferred to relate their responses according to general semantic similarity, followed by spatial proximity. Younger learners inverted the types of associations with spatial proximity being the first option and general semantic similarity the second. Both groups used synonyms and collocations least frequently for their associations or word attachments. We believe the specificities of the prompting category are indeed having a big impact on the selection of the attachment mechanisms (Rosch 1978, Cuenca and Hilferty 1999, Mahecha and Mateus 2017). Semantic category or theme has also been found to be of great relevance in the types of associations produced (e.g. Aitchinson 1994, Murphy 2001, Lin and Murphy 2001, Tomé Cornejo 2015, Mirman et al. 2017).

Previous research has tried to establish the type of categorization which predominates in human cognition. Although category-generation studies are difficult to systematize due to their different methodologies and to the overriding individual differences of participants, most studies agree that taxonomic organization is prevalent in adult, educated participants with specific domain knowledge (Lucariello et al. 1992, Mirman et al. 2017), but thematic organization is never abandoned (e.g. Sass et al. 2009) and suits well in the early stages of knowledge development, e.g. when encountering novel words (Chaffin 1997), novel situations or events (Murphy 2001, Lin and Murphy 2001), or with specific tasks or stimuli (cf. Mirman et al. 2017). Being able to switch between taxonomic and thematic organization depending on the purpose is a sign of cognitive flexibility, and therefore implies a better adaptation and learning capacity (Coni et al 2019).

An important issue within categorization is how categories are internally structured and the relationships established within its members. Accordingly, in the literature (Hernández Muñoz 2005, Tomé Cornejo 2015, Shivabasappa et al. 2017) category types are distinguished according to their internal structure into well-defined categories, natural or taxonomic and thematic categories. The membership level of the instances to the category and the number of shared conceptual features are key in this distinction. Thematic categories are those whose membership banks on experience or knowledge of the world, rather than on the properties of the members themselves (e.g. Lucariello et al. 1992, Lin and Murphy 2001, Sass et al. 2009). For example, in the present study, we have selected a natural taxonomic and a thematic category for the analysis, *animals* and *countryside*, respectively. *Animals* is an inclusive category, one in which elements establish a belonging relation of the type: X is/belongs to Y. These relations base on semantic or conceptual similarity and are characterised by class inclusion (Lin and Murphy 2001). *Countryside*, on its part, is a more relational category where members cannot always relate to the category label, so X is related to Y depending on the experience (and culture) of the respondent (e.g. Hernández Muñoz 2005, Tomé Cornejo 2015). Thematic associations appear when members of the category are grouped together because they occur in the same event, scheme, setting, or situation (Murphy 2001, Lin and Murphy 2001, Sass et al. 2009). The members of these different semantic categories will establish different types of associations among themselves, since their membership bases on different relations as well. The selection of type of association or other might depend on the meaningfulness and saliency of the relation (cf. Lin and Murphy 2001). Thus, Jiménez Catalán (forthcoming) identified more metonymic associations in *the house*, more position-based associations in *hobbies*, more paradigmatic in *professions*, and more meaning-based or relational in *school*. Accordingly, taxonomic categories such as *animals* can be expected to generate more categorial, coordinate and paradigmatic associations related to class inclusion. Thematic categories such as *countryside* might include a higher proportion of spatial, causal, functional or temporal relations (cf Lin and Murphy 2001, Sass et al. 2009).

As can be observed from the above literature review, there is a scarcity of research dealing with categorization in EFL learners at different proficiency levels, especially in the high school years. Additionally, spreading activation theory and network analysis have not, to our knowledge, been previously used to examine learners' connections in categorization. Two main aspects can be critically highlighted when approaching these previous studies and their findings. First, the fact that most of them deal with primary school learners either monolingual, or bilingual, but not with adolescent EFL learners. And second, previous studies use a similar methodology for data gathering, but a different one for data analyses. Specifically, previous studies tend to use statistical tests of the ANOVA or regression type. However, the present study uses network theory, which views the phenomenon from a different perspective allowing thus for complementary conclusions and findings which offer a new and fresh approach to the data. The present research study intends to cover that gap in an exploratory way and thus, open up avenues for further research with the use of network analysis and EFL categorization.

3. Statement of the Problem and Research Questions

With the previous considerations in mind, the present study looks into the associative component from data derived from a free word association task. Basing on previous taxonomies, we want to find out the nature and type of the associations learners use to attach one word to another. This study differs from previous ones, in that it does not examine first responses associated to stimuli, but rather it looks into connections of bigrams (pair of words) arisen in a chain-like manner from an initial prompt. In other words, here we are exploring the associations among responses themselves to establish their nature depending of learners' proficiency and semantic category analysed. The present study wants to be exploratory and descriptive with no claims of exhaustiveness. The Research Questions that guide the study are the following:

1. What is the nature of the connections in the learners' mental lexicon?
2. How does proficiency affect the nature of these connections?
3. What is the impact of different semantic categories on the type of word connections established?

4. Method

This paper offers a cross-sectional study of the pseudolongitudinal type where the responses of two groups of students with different age and L2 proficiency levels are compared. As will be explained in more detail below, data concerning the total number of responses were gathered and responses were classified on a repeated and type-like way. Additionally, some graph metrics were calculated with *Gephi* to delve deeper into the structure of the semantic categories in the learners' mental lexicon.

4.1 Participants

A total of 198 learners participated in this study. They were divided into two groups. Group 1 is made up of learners in their last year of Compulsory Secondary Education (grade 10), who are aged between 15 and 16. The other group, Group 2, is conformed by learners in their last year of Baccalaureate, the year previous to University entrance (grade 12). They are aged between 17-18. All of them were learners of English as a foreign language, they were at the A2 level (Group 1) and B1 (Group 2) (as per official guidelines of the educational authorities, and as revealed by the results of the *Oxford Placement Test* which accounts for learners general proficiency level). Group 1 will provide data to answer R.Q.1 and R.Q. 3 and together with group 2 will conform the sample for R.Q. 2.

Since previous research revealed differences in the ways speakers of different proficiency levels refer to the several semantic categories and in the types of connections they establish among words (see section word associations in the literature review, e.g. Wolter 2001, Zareva 2007, Fitzpatrick and Izura 2011, Precosky 2011), we deemed it pertinent to compare two groups of learners at different EFL proficiency level for the lexical connections and their categorization process.

The following table, Table 1, sums up the informants' main characteristics.

Table 1. *Informants' Characteristics*

	Group 1	Group 2
N (196)	98	98
Age	15-16	17-18
Grade	4 ESO/10	2 Bach/12
L2 level	A2	B1

4.2 Data Gathering Instrument

A semantic categorization task, specifically a semantic fluency task type, was used to elicit production of vocabulary data from informants. In particular, learners had to write, in two minutes, as many words, or phrases came to their mind (cf. e.g. Hernández Muñoz 2014, Jiménez Catalán 2014, Borodkin et al. 2016) in relation to the prompts: *animals* and *countryside*. In total, learners spent four minutes on this task. These two prompts were selected on three grounds: they feature a) different productivity, b) different response diversity or response spread and c) different cohesion index. *Animals* is an inclusive or closed category which gives rise to many responses but very homogeneous responses. Thus, it is more prone to taxonomic categorization (see also Borodkin et al. 2016). The semantic category *Countryside* is a less productive prompt, but where a broader amount of types are to be found, it is more open and gives rise to more heterogeneous responses (cf. Hernández Muñoz 2005, Tomé Cornejo 2015). Thematic categorization is believed to be higher in this category.

Participants were instructed in Spanish and each prompt and the corresponding responses occupied an independent sheet of paper. The semantic categorization task used is of the semantic fluency type and collects multiple responses from learners (cf. Schmitt 1998, Hernández Muñoz 2014) and gives thus a more complete picture of learners' lexicons (Sheng et al 2006, Precosky 2011, Borodkin et al. 2016). Multiple-response association tests tend to prompt chain responses that associate one another rather than to the stimulus word (cf. Precosky 2011). That is, the word produced will facilitate a recall of other related concepts or word forms.

4.3 Procedures and Analyses

I had learners complete the semantic categorization task in class. I then typed the responses into computer-readable form for each of the prompts. Data were submitted to *Gephi* software package (Cherven 2015). This program *Gephi* offers the total of the bigrams or two-word associations produced by learners, and, among other measures, their frequency in the data. For instance, if an informant has produced the 5-word chain: *pear-apple-banana-apricot-fresh*, the bigrams we obtain are these 4: *pear-apple*; *apple-banana*; *banana-apricot*; *apricot-fresh*. And it is the nature of the link within the bigram, the one we look into in the present study.

Responses were then examined for their patterns of association. Specifically, we classified bigram links depending on whether they were connected via a formal, a collocational, a semantic association or an encyclopaedic association. We could, thus, distinguish among the following categories collected in Table 2. The assumption, as Fitzpatrick and Izura (2011: 375) state is that "the number of responses made in a particular category can provide information about the organization, availability, and salient features of words and their concepts in the mental lexicon".

Table 2. *Taxonomy of Associations for Responses*

Type of Associations	Categories of Associations	Example
Meaning-Based	Categorial membership/ paradigmatic	Cat-dog
	Hypernym/ superordinate	Fish-shark
	Metonymic, functional or metaphorical	Tree-forest, dog-paw
	Shared semantic features	River-lake
	Synonyms and antonyms	Cow-bull
Formal Associations	Derivation	Farm-farmer
	Formal similarity	Farm-forest
Position-Based	Position-based (collocation)	Fresh-fruit
Encyclopaedic Associations	Experiential (influenced by experience, culture, Emotions, word knowledge)	Rhino-Endangered_animal

The taxonomy we used in the present study consists of form associations based on a) similarity [farm-forest] and b) derivations [farm-farmer], position-based of the collocation type [pet-feed], meaning-based associations according to a) semantic similarity based on categorial membership (base-words, coordinates, paradigmatic) [lion-tiger, cat-dog, tree-flower], b) hypernorym/ superordinate, [fish-shark, rose-flower] c) metonymic, functional or metaphorical associations [tree-forest, tail-dog], d) shared semantic features [river-lake, hamster-mouse] e) synonyms and antonyms [cow-bull, big-small], position-based associations of the collocational type [fresh-air], and experiential associations (influenced by experience, culture, emotions, word knowledge) [rhino-endangeredanimal, village-party]. We have adapted this taxonomy from Fitzpatrick (2006, 2007) to fit the objectives of the present study.

Together with the *Gephi* tool which provided graph data, SPSS 26.0 was also used to perform the different statistical analyses necessary. Specifically, we conducted chi-square tests to answer R.Q.1-3 and a chi-square goodness of fit to account for differences in the types of associations (R.Q.1). Percentages and raw numbers were also presented. In the present study, an additional approach was taken to explore the structural characteristics of the L2 lexicon, through the use of network analysis tools. The assumption underlying the use of these tools is that the mental lexicon is a complex system with a web-like structure or network, where *nodes* represent words (or lexical items) and *edges* represent the links between the words (cf. Borodkin et al 2016). Analyses conducted with Gephi allowed for calculating different measures such as the clustering coefficient, average shortest path length, and average degree or connected components.

5. Results

Our first research question asked about the types of associations produced by learners as a response to a category-generation fluency task. To answer it, we first, present some descriptive measures for all groups analysed. We obtained data concerning the total number of responses and types of responses and the number of word-pair associations, again on a repeated and type-like way. Additionally, we calculated some graph metrics with *Gephi*. Table 3 offers the results.

As can be observed, *animals* is a more productive category than *countryside*, since it generates more total responses. However, it gives rise to fewer types or different responses indicating that it is a more closed field with a higher cohesive index among responses. In other words, learners give many responses, which are shared by many participants. Similarly, and a result linked to the length of the chain, *animals* also generates more bigrams or word-pair associations. Nevertheless, the difference between "bigram types" and "bigram tokens" is higher for *animals*, indicating that learners repeat the exact associations made within this category; whereas exact associations are not so frequent in the semantic category *countryside*. Here participants display a wider array of different connections. This concurs with previous data on tokens and types explained above and with graph metrics as we will see below.

Turning to graph metrics, results reveal that average degree, the number of connections a word displays, is higher for *animals* than for *countryside*. This result, though expected from previous descriptive data, is very interesting, because not only are there fewer words in the semantic category *animals*, they are also much highly connected pointing to a more organized and structured semantic web than in the semantic category *countryside*. Further data point in the same direction, a shorter diameter and average path length, i.e. the number of nodes between two edges, indicate that the network can be navigated more easily, quickly and efficiently. That is, one can go from one end of the web to the other within few steps or intervening nodes/ words. In language terms, words can be found more easily, quickly and efficiently finding target items with ease. This is congruent with the higher token production in the semantic field *animals*, and with the association types found in this category, as we will see below. The connected components data, i.e. number of highly connected hubs or clusters, reveal that in general, the networks are highly structured and organized with many connections. However, again for *animals*, we can see stronger groups or connections, where the nodes are interconnected among themselves. In the second data set for *countryside*, on the contrary, we observe 2 different components clearly indicating that nodes are distributed into different unrelated groups, pointing again to a lesser organization of the network within this semantic category. Associations are freer, more difficult to systematize, more unpredictable and varied, and less related in the semantic field *countryside*. Finally, the figures for the clustering coefficient, i.e. the capacity of the web to connect nodes tightly, are again very revealing and insightful showing a higher clustering coefficient in *animals* meaning that nodes/ words within this web are strongly connected among themselves.

Table 3. *Descriptive Measures for the Data Groups*

	Animals_g1	Countryside_g1	Animals_g2	Countryside_g2
Tokens	1755	942	1392	1084
Types	228	353	238	430
Bigrams (Total)	1654	837	1307	1003
Bigrams (Not Repeated)	1059	752	938	916
Average Degree	4.645	2.13	3.941	2.13
Connected components	1 (strong)/ 23 (weak)	1 (strong)/ 71 (weak)	1 (strong)/ 17 (weak)	2 (strong)/ 87 (weak)
Diameter	11	23	13	20
Clustering Coefficient	0.171	0.067	0.191	0.043
Average Path Length	3.752	5.478	3.863	5.656

These results point to the two semantic categories having a slightly different structure in the learners' mental lexicon. Further analyses of the types of associations established to attach one word to another also shed more light in this direction.

Tables 4 and 5 present the breakdown of the types of associations for each semantic category and data set both with the bigram types (non-repeated bigrams) and the total number of bigrams. For all groups, associations resting on base words or coordinates is the most frequent type of association with over half of the associations belonging to this type. Metaphorical and experiential associations follow in frequency with less than 15% of the associations. The remaining types of associations feature very low numbers and represent low percentages as well, with derivation being the least frequent with only a couple of instances (less than 1%).

Table 4. *Raw Numbers and Percentage over Non-repeated of Association Types*

	Animals_g1		Countryside_g1		Animals_g2		Countryside_g2	
	No.	*%*	*No.*	*%*	*No.*	*%*	*No.*	*%*
Categorial Membership/ Paradigmatic	746	70.44	395	52.52	621	66.2	490	53.5
Hypernym/ Superordinate	38	3.58	22	2.93	40	4.26	27	2.95
Metonymic/ Functional/ Metaphorical	78	7.36	99	13.16	92	9.8	136	14.84
Shared Semantic Features	44	4.15	29	3.86	43	4.58	35	3.82
Synonyms and Antonyms	29	2.74	34	4.52	33	3.52	42	4.6
Derivation	1	0.095	3	0.4	0	0	1	0.11
Formal Similarity	27	2.55	11	1.47	13	1.39	25	2.72
Position-Based	16	1.51	26	3.46	12	1.28	26	2.83
Experiential	80	7.55	133	17.67	84	8.96	134	14.62

Table 5. *Raw Numbers and Percentage over Total of Association Types (Token-wise)*

	Animals_g1		Countryside_g1		Animals_g2		Countryside_g2	
	No.	*%*	*No.*	*%*	*No.*	*%*	*No.*	*%*
Categorial Membership/ Paradigmatic	1297	78.35	443	53	935	71.53	538	53.53
hypernym/ Superordinate	76	4.6	25	3	59	4.51	30	3
Metonymic/ Functional/ Metaphorical	82	4.96	138	16.5	95	7.26	150	15
Shared Semantic Features	44	4.3	31	3.7	62	4.74	50	5
Synonyms and Antonyms	32	1.93	34	4	36	2.75	44	4.38
Derivation	1	0.06	6	0.7	0	0	2	0.2
Formal Similarity	27	2.54	11	1.31	22	1.68	25	2.5
Position-Based	16	0.97	29	3.46	12	0.92	30	3
Experiential	80	5.5	135	16.13	86	6.56	134	13.36

Results of χ^2 tests of comparison reveal that the frequency of recourse to the association types is significantly different for all data sets (see Table 6). That is, they do not attach the words randomly, but they follow a system, prioritizing semantic-based attachment, specifically between base-words or coordinates of the same semantic category,

i.e. taxonomic categorization. Recourse to experienced and culturally-learned scenarios or image schemas follow in preference, i.e. thematic categorization.

Table 6. *Tables for χ² Tests of Comparison for the Association Types*

Test Statistics	Association Type
Chi-square	3808.725
d.f.	8
Sig.	.000
Animals_g1	

Test Statistics	Association Type
Chi-squre	1501.279
d.f.	8
Sig.	.000
Countryside-g1	

Test Statistics	Association Type
Chi-square	2524.789
d.f.	7
Sig.	.000
Animals_g2	

Test Statistics	Association Type
Chi-squre	1850.537
d.f.	8
Sig.	.000
Countryside-g2	

χ² goodness-of-fit post-hoc analysis with binomial test revealed that differences were significant among the category types[2] (see Table 7). This clearly showed that learners opted for coordinates when trying to recall new words followed by thematic relations based on pre-determined or already known schemas, some coming from general world knowledge and some deriving from their own personal experiences. Formal associations are the least used forms of attachment when bringing words to memory. Interestingly enough, when categories based on experiential associations and metaphorical or contextual associations are collapsed into a single thematic category, binomial test results show significant differences of a much lesser magnitude between taxonomic and thematic associations. It is necessary to make clear that all analyses could be conducted with SPSS making no loss of cases or values.

Table 7. *Table for χ² Goodness-of-fit post-hoc Analysis with Binomial Test.*

Binomial Test		Category	N	Observed	Tested	Sig. exact (bilateral)
Association	Group1	Taxonomic	745	.47	.50	.024
	Group2	Thematic	836	.53		
	Total		1581	1.00		

Additionally, on a qualitative observation of the data, we could identify some differences in the clustering solutions provided for each of the categories. *Animals* generates four main clustering types: 1) farm/ countryside animals (cow, sheep, dog), 2) wild/zoo animals (tiger, elephant, giraffe), 3) sea animals (shark, dolphin, whale) and 4) emotions, feelings (friend, calm, dangerous). On its part, *countryside* gives rise to four main clustering types as well, yet of a much different and varied kind, since here the notion of scenarios or schemas predominate: 1) countries and government scenario (Spain, France, queen, king, president), 2) vegetation and landscape scenario (tree, hill, mountain), 3) Spanish village related scenario (farm, summer, holidays) and 4) feelings and emotions (fun, relax, calm).

Our second research question asked for changes in association patterns as L2 proficiency progressed. Here, we reproduce Table 4 for the comparison across proficiency levels where we can see that in fact, association patterns change very little from grade 10 to grade 12 (see Tables 8 & 9). Thus, over two-thirds of the associations produced are of the paradigmatic type with over 66% for animals and over 50% for countryside. Again derivation is the least frequent type of association with less than 1% of occurrences. Metaphorical and experiential associations rank second and third with between 7 and 17% of presence.

Again, χ² tests of comparison revealed significant differences for the proficiency groups both for *animals* and *countryside* (see Table 10). Results point to a contradictory tendency in the two semantic categories tested with learners increasing their preference for thematic and experiential associations when recalling *animals* as they go up grade, but decreasing their recourse to experience-based associations in favour of coordinates when retrieving words related to *countryside*.

Table 8. *Association Patterns for Animals in Grade 10 and Grade 12*

	Animals_g1		Animals_g2	
	No.	*%*	*No.*	*%*
Categorial Membership/ Paradigmatic	746	70.44	621	66.2
Hypernym/ Superordinate	38	3.58	40	4.26
Metonymic/ Functional/ Metaphorical	78	7.36	92	9.8
Shared semantic features	44	4.15	43	4.58
Synonyms and antonyms	29	2.74	33	3.52
Derivation	1	0.095	0	0
Formal similarity	27	2.55	13	1.39
Position-Based	16	1.51	12	1.28
Experiential	80	7.55	84	8.96

Table 9. *Association Patterns for Countryside in Grade 10 and Grade 12*

	Countryside_g1		Countryside_g2	
	No.	*%*	*No.*	*%*
Categorial Membership/ Paradigmatic	395	52.52	490	53.5
Hypernym/ Superordinate	22	2.93	27	2.95
Metonymic/ Functional/ Metaphorical	99	13.16	136	14.84
Shared Semantic Features	29	3.86	35	3.82
Synonyms and Antonyms	34	4.52	42	4.6
Derivation	3	0.4	1	0.11
Formal Similarity	11	1.47	25	2.72
Position-Based	26	3.46	26	2.83
Experiential	133	17.67	134	14.62

Table 10. *Table for χ^2 Tests of Comparison for the Proficiency Levels*

Test statistics	Association Type
Chi-square	6753.907
d.f.	8
Sig.	.000

We also wanted to check whether the association patterns for lexical attachment differ across semantic categories. To that aim, we compared responses to the semantic categories *animals* and *countryside*. Table 11 offers the data. As we can observe from the percentages offered, patterns differ slightly between both. Base word associations are less frequent in countryside, whereas experiential and metonymic or metaphorical or contextual become more extended. This basically means that learners use different patterns to attach words depending on their semantic category. In *countryside*, attachment banks on experience and learned scenarios of what makes a *countryside* image or picture, i.e. thematic categorization. In *animals*, attachment bases on categorial membership, mostly, i.e. taxonomic categorization.

Table 11. *Association Patterns for Animals and Countryside*

	Animals_g1		Countryside_g1	
	No.	*%*	*No.*	*%*
Categorial Membership/ Paradigmatic	746	70.44	395	52.52
Hypernym/ Superordinate	38	3.58	22	2.93
Metonymic/ Functional/ Metaphorical	78	7.36	99	13.16
Shared Semantic Features	44	4.15	29	3.86
Synonyms and Antonyms	29	2.74	34	4.52
Derivation	1	0.095	3	0.4
Formal Similarity	27	2.55	11	1.47
Position-Based	16	1.51	26	3.46
Experiential	80	7.55	133	17.67

χ^2 tests of comparison revealed significant differences for *animals* and *countryside* as concerns the frequency of use of each category, with *countryside* giving rise to a less skewed and more regular distribution (see Table 12).

Nevertheless, for both categories coordinating is the most prominent recall mechanism followed by event-based associations based on world knowledge and personal experience with the remaining association types throwing marginal percentages for both categories.

Table 12. *Table for χ² Tests of Comparison for the Categories (Animals vs. Countryside)*

Test Statistics	Association Type
Chi-square	3346.317
d.f.	8
Sig.	.000

Finally, we include some graph figures (Figures 1-4) to illustrate the organization of the different semantic categories.

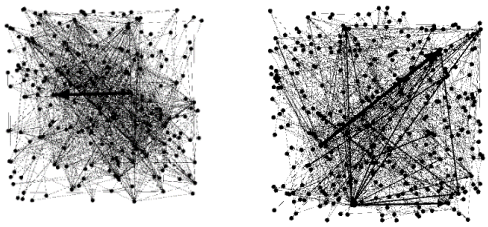

Figure 1. Graph for animals-g1 vs. countryside-g1 (weighted, directed networks)

From the graph figures in Figure 1 we can observe that the organization is denser (higher concentration of nodes in the middle part of the graph) in *animals* than in *countryside*. In order to get a better and clearer look of the organization, we prune to see the more connected nodes. The pruning gives the following figures, Figure 2 for pruning > 10, Figure 3 for prunin > 30 and Figure 4 for pruning > 40:

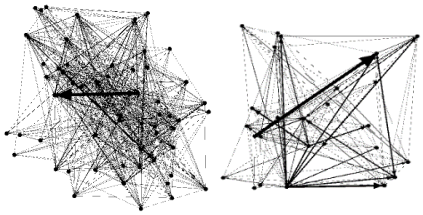

Figure 2. Graph for animals-g1 vs. countryside-g1, pruned to degree > 10 (weighted, directed networks)

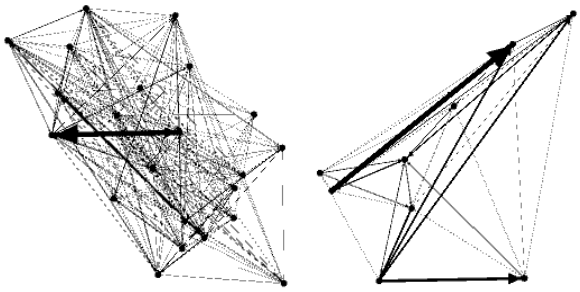

Figure 3. Graph for animals-g1 vs. countryside-g1, pruned to degree > 30 (weighted, directed networks)

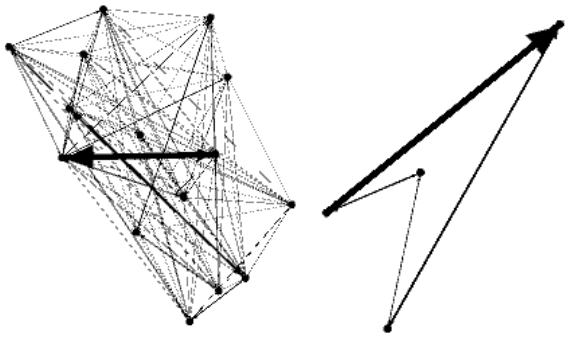

Figure 4. Graph for animals-g1 vs. countryside-g1, pruned to degree > 40 (weighted, directed networks)

From Figures 1-4 above, we can observe the internal structure of the two semantic categories is different. The organization in *animals* is denser and more compact, and therefore, more accessible; whereas in *countryside* the network is more diffuse, less tight or connected and connections seem less systematic, despite the higher presence of nodes. Graphs allow us to present information in a graphical way and they serve the metaphor of the mental lexicon as a semantic network. The distribution of the lexical information in the graphs and the size of the edges show the depth and strength of the word associations within the semantic categories, as well as the organization of the words in the mind, which points to the accessibility of each word within the network (words with many associations and associations that are repeated are highly accessible to the learner).

Finally, for illustrative purposes, a list of the most frequent bigrams and word-pair associations between responses, is presented. The number of occurrences appears in brackets. In Table 13 it can clearly be observed that taxonomic associations are the most frequent, with the category *animals* displaying a much higher level of homogeneity.

Table 13. *Most Frequent Bigrams*

Animals_g1	Countryside_g1
Dog-cat (56)	Mountain-river (10)
Cat-dog (29)	Tree-flower (6)
Lion-tiger (21)	Cow-horse (5)
Tiger-lion (20)	Tree-lake (4)
Pig-cow (9)	Tree-river (4)

6. Discussion

The main objective of the present study was to look into lexical attachment in the incorporation of information, new words, to the mental lexicon. The study aimed to find the most frequent mechanisms followed for the said attachment and see whether those association patterns are influenced by L2 proficiency of the learners and by the semantic category at stake. Results from descriptive graph metrics showed that *animals* and *countryside* throw different data pointing to the former being a more highly structured category than the latter. This matches previous findings related to the fact that taxonomic categories display higher levels of organization than more thematic ones, which are less structured (cf. Chaffin 1997, Peña et al. 2002, Sheng et al. 2006). Analysis of word association mechanisms corroborates this fact. Similarly, categories which display larger amounts of words (tokens) are also found to be more likely to show taxonomic organization, i.e. higher presence of coordinates or base words (Peña et al. 2002). This is also found in our present data. Although categorial membership or coordinates is the most frequent type of association altogether, it is indeed more frequent in *animals* than in *countryside*. Thematic (metonymic, metaphorical, contextual and experiential) associations follow in frequency with higher presence in *countryside* than in *animals*. These results clearly reveal the different nature of both semantic categories used in the present study, while at the same time confirming the prevalent use of coordinates (paradigmatic semantic associations, meaning-based (Cf. Fitzpatrick and Thwaites 2020)) as mechanisms for lexicon growth disregarding the type of semantic category (Chaffin 1997). Furthermore, our results also concur with previous findings, where *paradigmatic responses* [sic.] were found to be the most frequent (Cf. Fitzpatrick and Thwaites 2020); especially where the cue word is familiar (Wolter 2001). Results are also in line with previous studies (e.g. Henriquez

Guarín 2016) about the type of semantic information that guides lexical storage and retrieval: coordinates and contextual and experiential contiguity.

This result might be telling us that learners in academic contexts preferably refer to taxonomic associations, i.e. use of coordinates, to recall words independent of the semantic category (Lin and Murphy 2001, Mirman et al. 2007). Nevertheless, thematic categories give rise to more experiential and event-based associations clearly showing the cognitive flexibility of learners and their capacity to adapt their learning and production mechanisms to the situation (Sass et al. 2009, Coni et al. 2019). The type of association that predominates depends on the semantic category under scrutiny and on the type of preferred learning and recall mechanism, in our data, coordinates followed by event-based links and associations banking on experience and general knowledge background (cf. Lin and Murphy 2001, Murphy 2001). Probably, familiarity with the prompt calls for more coordinate associations, which also seem to be a faster link, and a less taxing procedure for lexical learning and semantic attachment; whereas when relations are less obvious or more based on experience attachment is slower (cf. Chaffin 1997, Lin and Murphy 2001, Wolter 2001). Furthermore, when scenes and schemas get repeated and learners gain experience with them, they are able to infer general characteristics and apply more taxonomic categorization based on those generalizations (cf. Peña et al. 2002). In our data, experiential associations decrease when repeated bigrams are considered, whereas taxonomic associations increase.

Apart from confirmation of previous studies, where a higher number of responses marked higher L2 proficiency (e.g. Zareva et al. 2005); increasing experience with the language and the language learning process has been found to result in significant differences concerning the frequency of use of the different attachment mechanisms. However, the order of preference for the said mechanisms remains stable across grades with learners opting for coordination first and event-based associations second (cf. Namei 2004, Zareva 2007). This result also mirrors previous findings by Rashidi and Mirsalari (2017), who found that learners at different proficiency levels appeared to have adopted the same strategies and did not differ much in their total use of strategies. The remaining categories present rather marginal frequencies of occurrence. This taxonomic preference is reflected in the literature (e.g. Lucariello et al 1992), which also shows, however, that thematic associating is present even in adult categorizing (e.g. Lin and Murphy 2001). Our results point to the same direction and allow us to conclude that FL learners alternate between coordination and event-based relations to adapt to the requirements of the learning situation. Storing and recalling words from the mental lexicon uses taxonomic mechanisms which are quicker for known words and guide the retrieval of semantic knowledge, but learners are able to resort to thematic associations based on general knowledge schemata or on schemata drawn from experience to store and recall new words and thus commit them to semantic memory.

Differences found in the metrics and association types of the semantic categories scrutinised here can be traced back to the real vocabulary size of the category (cf. Peña et al. 2002, Borodkin et al 2016). It seems reasonable to argue that *countryside* will give rise to more response types and associations types, because experiential knowledge can be source of more associations and also less limited associations than *animals*, which represents a more restricted category corresponding to the list or repertoire of animals found in the real world, mainly. In this sense, access might be easier and quicker for *animals*, since this "animal list" is pre-existing in the learners' mental lexicon, and probably has been acquired earlier in life (cf. Tomé Cornejo 2015: 295, 324). Responses are also more homogenous, i.e. coincidental within the group. Our results concur with this. On the contrary, *countryside* is a more abstract, diffuse semantic category which favours a higher number of response types (more heterogeneous or less coincidental responses) coming from this and other categories, e.g. also from *animals*, *food*, *emotions*, and different situations or events, as we saw in our data. Nevertheless, this characteristic also results in items not being so readily available and therefore the total number of responses is lower (cf. Tomé Cornejo 2015: 297, 323). Taxonomic associations are also frequent in this category, mainly because the amount of information within this category is so wide that thematic mechanisms for word attachment and retrieval might not be enough and learners need to resort taxonomic, coordinate associations to make sense of the category. Thus, both types of mechanisms are necessary (cf. Lin and Murphy 2001).

In this sense, it is interesting to wonder what results would be concerning categorization if fully adult learners were tested, or even if they were tested in the L1. We might wonder whether the meaning and the internal structure of the categories is an overriding factor over L1/L2 status or other individual characteristics of the learners (e.g. Aitchinson 1994, Mirman et al. 2017). Previous studies such as the one conducted by Borodkin et al (2016) where semantic networks in L1 and L2 were compared found that non-native networks were less organized and more difficult to navigate

disregarding the category analysed (communities are not tightly connected within). Some findings concerning lexical categorization have revealed that in order to account for the links between L2 words, learners borrow their L1 lexical-semantic network (e.g., Pavlenko, 2009), and as exposure to the L2 increases, the network undergoes reorganization to accommodate to the new language and cultural specificities (Malt & Sloman, 2003); however, to our knowledge, this is the first study to deal with these issues in relation to connectivity between words (apart from the above-mentioned Borodkin et al 2016).

Additionally, we can also speculate that it is probable that learners obtain the exemplar they recall from subcategories (e.g. sea animals), hence the numerous amount of coordinated or paradigmatic associations. Once a category is exhausted, they might move to another subcategory (e.g. wild animals, or types of habitat), and retrieve the desired word via other types of links or associations, e.g. collocations, world knowledge, or metonymy.

Our results also suggest that networks built from word associations can capture lexical storage and learning. We could clearly see that the most productive category showed a more compact and stable structure, perhaps because of both a larger vocabulary size and the presence of more systematic associations (taxonomic, paradigmatic). Although it might be an oversimplification of actual lexical development, it offers a crucial window into the mental lexicon (Fitzpatrick and Thwaites 2020). This study intends to present an interface between lexical studies and mathematical or technological approaches.

A final consideration in this brief discussion is called for concerning the role of usage-based (UB) linguistics in the interpretation of word associations. From our results, it does not seem ajar to argue that previous shared and individual experiences (linguistic and extra-linguistic) contribute to shaping linguistic experience and lexical development (cf. Fitzpatrick and Thwaites 2020). It would be interesting to try to map network analysis and association data with the tenets of usage-based conceptions of language and language acquisition in order to provide a new and probably enlightening perspective on the organization of the mental lexicon. The use of chunks or prefabricated word strings in a UB sense can be very helpful in examining lexical development and retrieval. Additionally, the idea of frequency in the input (and output) being central for the acquisition of an item, its entrenchment and routinization which ultimately leads to faster retrieval (Quick and Verschik 2019) is closely linked to the idea of centrality and anchor elements of network analysis. The more an item becomes a routine, the less demanding it is for the system according to UB theory, the more central an element, the easier and faster it can be retrieved and can aid in the retrieval of other related items says network analysis and LA trends. Finally, UB theory accounts well for individual variation, because although individual variation is very important in word pair associations, there is enough overlap in their frequency so that most frequent associations (bigrams) are shared by all (many) respondents (cf. Quick and Verschik 2019). Our results point in this direction. Individual differences are partially ruled out by the notion of (input and output) frequency from a UB perspective.

7. Conclusion

Two words are considered associated if one is produced when prompted by the other. The type of this association is yet a different matter, and individual and categorial differences give rise to idiosyncratic associations. However, in the present study, we tried to find the most common types of associations found within the semantic categories of *animals* and *countryside*. We could safely ascertain that although the type of relation that predominates may depend on the organization of the semantic category at stake, coordinate, taxonomic associations are the most frequent disregarding semantic category and proficiency level.

As learners increase their experience with the language, they display a better adaptative behaviour which results in an increase in the variety of the association types resorted to. Clearly, thematic associations are also important in the categorization ability of adults.

With increasing age and proficiency level, learners show a great capacity of flexibility in categorization to adapt to the task they are performing. We are led to think that this flexibility might be an advantage in FL learning. Learners store the words they know and learn as members of taxonomic categories, i.e. according to their internal properties, but also as part of events, scenes, or schemata, i.e. according to their external relations with other concepts, with cross-categorization as a very likely option. Both categorization types complement and enrich human cognition and learning and contribute to an organized and efficient system. It is interesting to note that learners make sense of the world around them through relations of belonging and through metaphors drawn from the physical world. Previous experiences

constitute the basis for lexical learning, organization, and word storage. The linguistic and extra-linguistic contexts when and where the words are encountered shape the way learners store and retrieve the words they need.

Interesting pedagogical implications derive from the results of the present research pertaining to the way and order in which FL vocabulary is taught. The present results hint at a didactic plan which fosters the teaching of coordinates and promotes links with previous experience, and world knowledge thus stimulating the activation of event- or situation-based mental schema to enhance vocabulary acquisition, e-g- through metaphorical associations (cf. Babko et al. 2019, Najjari and Mohammadi 2018). More advanced technological tools fed with corpus and association data can help predict WA and accordingly determine teaching sequences. Finding out how learners learn and store words together, especially those pertaining to the same semantic categories, can help us elucidate what vocabulary to teach, e.g. create thematic vocabulary lists to teach at the different levels, and also which words to teach together, i.e. those which are stored together. Additionally, teachers can provide learners with vocabulary learning strategies based on the word association types identified previously, such as metaphors, lexical paradigms, word families, derivations, and cultural, emotional or experiential words. There have been previous attempts at making proposals for vocabulary teaching arranged around thematic cores or semantic fields, such as the Curricular Plan of the Cervantes Institute for Spanish foreign language learners, or the Cambridge English Vocabulary Lists. These materials include lists adapted to the CEFR levels. Our data could largely contribute to these recommendation lists.

However, this study presents some limitations mainly pertaining to the assumptions made in categorization about the context and learners' mental lexicons. Associations are difficult to categorize, this categorization necessarily involves a certain degree of subjectivity, additionally, sometimes, word pair associations can be assigned to several categories and many, as we saw, derive from emotional or experiential background. Our categorization system is motivated by data; it uses multiple categories and enables identification of bigram associations in a rather transparent way.

Further research should concentrate on exploring how mathematical models, such as graph theory used here, and other deep learning technologies can help elucidate and better understand how lexical learning proceeds. Identifying central elements that anchor lexical knowledge stands as a promising objective in future studies within FL lexical acquisition.

This study is to be considered exploratory research only, with no claims to exhaustiveness.

Notes

[1]Conversely, the Collins and Loftus (1975) spreading-activation model of semantic processing can account very well for the data generated in the present study.

[2]We are grateful to the mathematician and computing engineer Dr. Mata for assisting us with the statistical decisions.

Acknowledgements

We want to express our special thanks to mathematician Dr. Gadea Mata who helped us with the analyses.

Funding

We acknowledge funding by Grant no. PGC2018-095260-B-I00, Ministry of Science and Innovation.

References

Aitchison, J. (1994): *Words in the mind: An introduction to the mental lexicon.* Basil Blackwell.

Babko, K., O. Samarkina, N., Kh. Fazlyeva, Z., & Carlson, C. (2019). The real psycholinguistic meaning of a word and metaphorically reinterpreted component correlation (based on English phraselogical Units). *Journal of Research in Applied Linguistics*, 10 (Proceedings of the 6th International Conference on Applied Linguistics Issues (ALI 2019) July 19-20, 2019, Saint Petersburg, Russia), 855-861. doi: 10.22055/rals.2019.15158

Bartol Hernández, J. A. (2006). La disponibilidad léxica. *Revista Española de Lingüística*, *36*, 379–396.

Borge-Holthoefer, J., & Arenas, A. (2010) Semantic networks: Structure and dynamics. *Entropy, 12*, 1264–1302.

Borodkin, K, Kenett, Y., Faust, M., & Mashal, N. (2016). When pumpkin is closer to onion than to squash: The structure of the second language lexicon. *Cognition, 15*(6), 60-70.

Cambridge English Vocabulary Lists (2012). https://www.cambridgeenglish.org/images/84669-pet-vocabulary-list.pdf

Chaffin, R. (1997). Associations to unfamiliar words: Learning the meanings of new words *Memory* & *Cognition* 25(2), 203-226.

Cherven, K. (2015). *Mastering Gephi network visualization*. Packt Publishing.

Collins, A.M., & Loftus, E.F., (1975). A spreading-activation theory of semantic processing. *Psychol Review, 82*(6), 407–428.

Collins, A.M., & Quillian, M.R. (1969). Retrieval time from semantic memory. *Journal of Verbal Learning and Verbal Behaviou, 8*(2), 240–248.

Coni, A.G., Ison, M., & Vivas, J. (2019). Conceptual flexibility in school children: Switching between taxonomic and thematic relations. *Cognitive Development, 52*, 100827.

Cuenca, M.J., & Hilferty, J. (1999). *Introducción a la lingüística cognitiva*. Ariel.

De Deyne, S., Kenett, Y. N., Anaki, D., Faust, M., & Navarro, D. (2017). Large-scale network representations of semantics in the mental lexicon. In M. N. Jones (Ed.), *Big data in cognitive science* (pp. 174–202). Routledge/Taylor & Francis Group.

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

Fitzpatrick (2006). Habits and rabbits: Word associations and the L2 lexicon. *EUROSLA Yearbook 6*, 121-145.

Fitzpatrick, T. (2007). Word association patterns: Unpacking the assumptions. *International Journal of Applied Linguistics, 17*(3), 319–331.

Fitzpatrick, T., & C. Izura. (2011). Word association in L1 and L2: An exploratory study of response types, response times, and interlingual mediation. *Studies in Second Language Acquisition 33*(3). 373–398.

Fitzpatrick, T., Playfoot, D., Wray, A., & Wright, M. (2013). Establishing the reliability of word association data for investigating individual and group differences. *Applied Linguistics*. doi: 10.1093/applin/amt020.

Fitzpatrick, T., & Thwaites, P. (2020). Word association research and the L2 lexicon. *Language Teaching, 53*(3), 237-274. doi: 10.1017/S0261444820000105.

Gudmundson, A. (2020). The mental lexicon of multilingual adult learners of Italian L3: A study of word association behavior and cross-lingual semantic priming. In Camilla Bardel & Laura Sánchez (eds.), *Third language acquisition: Age, proficiency and multilingualism*, (pp. 67–109). Language Science Press.

Henríquez Guarín, M.C, Mahecha Mahecha, V., & Mateus Ferro, G. (2016). Análisis de los mecanismos cognitivos del léxico disponible del Cuerpo humano a través de grafos. *Lingüística y Literatura, 69*, 229-251.

Hernández Muñoz, N. (2005). La disponibilidad léxica: una herramienta fronteriza para el estudio del léxico en Lingüística y Psicología. *Estudios sobre adquisición del lenguaje*. (pp. 942-954): Ediciones Universidad de Salamanca.

Hernández Muñoz, N. (2014). Categorías en el léxico bilingüe: perspectivas desde el priming semántico interlenguas y la disponibilidad léxica. *RAEL: Revista electrónica de lingüística aplicada,13*(1), 19-38.

Hernández Muñoz, N. & López García, A. (2014). Análisis de las relaciones semánticas a través de una tarea de libre asociación en español con mapas auto-organizados. *RLA: Revista de Lingüística Teórica y Aplicada, 52*(2), 189-216.

Higginbotham, J. (2010). Individual learner profiles from word association tests: The effect of word frequency. *System, 3*8, 379-390.

Jiménez Catalán, R.M. (Ed.). (2014)**.** *Lexical availability in English and Spanish as a second language*. Springer.

Jiménez Catalán, R.M. (2022). Lexical profiles of children and adolescent EFL learners in the semantic domain of animals. In Vanessa De Wilde & Claire Goriot (eds.), *Second language learning before adulthood: Individual differences in children and adolescents*, (pp. 133-154). De Gruyter Mouton. .

Karuza, E. A., Thompson-Schill, S. L., & Bassett, D. S. (2016).Local patterns to global architectures: Influences of network topology on human learning. *Trends in Cognitive Sciences,* 20(8), 629–640.

Lin, E.L., & Murphy, G. L. (2001). Thematic relations in adults' concepts. *Journal of Experimental Psychology: General, 130* (1):3–28.

Lucariello, J., Kyratzis, A., & Nelson, K. (1992). Taxonomic knowledge: What kind and when? *Child Development 63*(4), 978-998.

Mahecha, V., & Mateus, G. (2017). El léxico disponible y sus mecanismos de asociación: Un análisis con grafos. In Florencia del Barrio de la Rosa (ed.), *Palabra, Vocabulario, Léxico. La lexicología aplicada a la didáctica y a la diacronía*, (pp. 123-142). Università Ca' Foscari.

Malt, B.C., & Sloman, S.A., (2003). Linguistic diversity and object naming by non-native speakers of English. *Bilingualism: Language and Cognition, 6*(01), 47–67.

Meara, P. (1982). Word associations in a foreign language: A report on the Birkbeck vocabulary project. *Nottingham Linguistic Circular*, *11*(2), 29-37.

Mirman, D., Landrigan, J.F., & Britt, A.E. (2017). Taxonomic and thematic semantic systems. *Psychological Bulletin, 143*(5), 499–520.

Morais, A.S., Olsson, H., & Schooler, L.J. (2013). Mapping the structure of semantic memory. *Cognitive Science, 37*(1), 125-145.

Murphy, G.L. (2001). Causes of taxonomic sorting by adults: A test of the thematic-to-taxonomic shift. *Psychonomic Bulletin & Review*, *8*(4), 834–839.

Najjari, R., & Mohammadi, M. (2018). Metaphorical conceptualization of SPORT through TERRITORY as a vehicle. *Journal of Research in Applied Linguistics*, *9*(1), 127-147. doi:10.22055/rals.2018.13408

Namei, S. (2004). Bilingual lexical development: A Persian–Swedish word association study. *International Journal of Applied Linguistics 14*(3), 363-388.

Instituto Cervantes (2006). *Plan curricular del Instituto Cervantes. Niveles de referencia para el español*. Madrid: Instituto Cervantes- Biblioteca nueva.

Pavlenko, A. (2009). Conceptual representation in the bilingual lexicon and second language vocabulary learning. In Pavlenko, A. (Ed.), *The bilingual mental lexicon: Interdisciplinary approaches* (pp. 125-160). Multilingual Matters.

Peppard, J. (2007). *Exploring the relationship between word-association and learners' lexical development.* Center for English Language Studies: University of Birmingham.

Peña, E.D., Bedore, L.M., & Zlatic-Giunta, R. (2002). Category-generation performance of bilingual children: The influence of condition, category, and language. *Journal of Speech Language and Hearing Research, 45*, 938–947.

Precosky, K. (2011). *Exploring the mental lexicon using word association tests: How do native and non-native speakers of English arrange words in the mind*? (PhD thesis). University of Birmingham.

Pranoto, B.E., & Afrilita, L. K. (2018). The organization of words in mental lexicon: Evidence from word association test. *TEKNOSASTIK, 16*(1), 26-33.

Quick, A., & Verschik, A. (2019). Usage-based contact linguistics: An introduction to the special issue. *Applied Linguistics Review*, *10*, 1-13.

Rashidi, N., & Mirsalari, S. (2017). Investigating the relationship between Iranian EFL learners' use of strategies in collocating words and their proficiency level. *Journal of Research in Applied Linguistics*, *8*(2), 93-118. doi: 10.22055/rals.2017.13093

Rosch, E. (1978). Principles of categorization. In E. Rosch, & B. B. Lloyd (Eds.), *Cognition and categorization* (pp. 27–48). Erlbaum.

Roux, P. W. (2013). Words in the mind: Exploring the relationship between word association and lexical development. *Polyglossia*, *24*, 80–91.

Sass, K., Sachs, O., Krach, S., & Kircher, T. (2009). Taxonomic and thematic categories: Neural correlates of categorization in an auditory-to-visual priming task using fMRI. *Brain Research*. 1270:78–87. doi: 10.1016/j.brainres.2009.03.013.

Sánchez-Saus Laserna, M. (2016). *Léxico disponible de los estudiantes de español como lengua extranjera en las universidades andaluzas.* Universidad de Sevilla.

Schmitt, N. (1998) Tracking the incremental acquisition of second language vocabulary: A longitudinal study. *Language Learning*, *48*, 281-317.

Séguin, M. (2015). Exploration of the relationship between word-association and learners' lexical development with a focus on American L1 and Croatian L2 speakers. *ExELL (Explorations in English Language and Linguistics)*, *3(*2), 80-101

Sheng, L., McGregor, K.K., & Marian, V. (2006). Lexical–semantic organization in bilingual children: Evidence from a repeated word association task. *Journal of Speech, Language, and Hearing Research 49*(3), 572–587.

Shivabasappa P., Peña E. D., & Bedore L. M. (2017). Typicality effect and category structure in Spanish–English bilingual children and adults. *Journal of Speech, Language, and Hearing Research*, *60*(6), 1577–1589.

Šifrar Kalan, M. (2014). Disponibilidad léxica en diferentes niveles de español/lengua extranjera. *Studia Romanica Posnaniensia*, *41*(1), 63–85.

Steyvers, M., & Tenenbaum, J.B. (2005). The large-scale structure of semantic networks: Statistical analyses and a model of semantic growth. *Cognitive Science*, *29*, 41–78 16.

Tomé Cornejo, C. (2015). *Léxico disponible. Procesamiento y aplicación a la enseñanza de ELE*. Universidad de Salamanca, tesis doctoral.

Vitevich, M. (2008). What can graph theory tell us about word learning and lexical retrieval? *Journal of Speech, Language, and Hearing Research*, *51*(2): 408–422.

Wolter, B. (2001). Comparing the L1 and the L2 mental lexicon: A depth of individual word knowledge model. *Studies in Second Language Acquisition*, *23*(1), 41–69.

Zareva, A., Schwanenflugel, & Nikolova Y. (2005) Relationship between lexical competence and language proficiency: Variable sensitivity. *Studies in Second Language Acquisition*, *27*, 567–95.

Zareva, A. 2007. Structure of the second language mental lexicon: How does it compare to native speakers' lexical organization? *Second Language Research*, *23*(2), 123–153.