The impact of semantic mapping technique on the organization of bilingual mental lexicon and L2 utterance fluency of Iranian EFL learners

Sara Ebrahimi

PhD

University of York
Education

June 2021

Abstract

To set a scene on how second language oral fluency can be developed effectively, repertoire of lexical knowledge has been scrutinized for long; however, the impact of organization of the mental lexicon, concentrating on network knowledge, has yet to be explored. To this end, semantic mapping has been introduced as a technique supporting the development of already known lexical items, providing chances for restructuring the network knowledge. Accordingly, the primary aim of this study was to investigate the extent to which semantic mapping had an impact on L2 fluency of speech. This had been researched in tandem with the impact of same technique on organizational structure of mental lexicon, exploring the connectivity and strength of the connections among items stored. Finally, the relationship between organizational structure of the mental lexicon and L2 oral fluency was explored.

A quantitative study was conducted with thirty-three Iranian intermediate EFL students, randomly divided between two groups: experimental and control groups. All participants took part in speaking and productive word association tests in three stages of pre-tests, immediate post-tests, and after two weeks' time interval in delayed post-tests. The experimental group participated in six-session training on semantic mapping, concentrating on practicing and retrieving the previously met lexical items. The control group didn't receive any training on semantic mapping while being presented with the same material. Participants' oral performances were analysed in terms of a range of fluency measures. In addition, the data collected from the word association tests was analysed regarding the overall network quality, which was used as a gauge to assess if developments in the lexical network was envisaged.

The findings revealed that overall word association score increased between pre-test, immediate post-test, and also delayed post-test only for the semantic mapping group. Likewise, the results for measures of oral fluency indicated a similar pattern, namely a significant increase in almost all measures, except for the mean duration of silent pauses. Consequently, the results of correlational analysis showed a relationship between organization of mental lexicon and oral fluency. Findings are explained in terms of efficacy of the technique by looking into model of L2 lexical development while inspecting the contribution of the shifts in number and strength of connections, and finally the way strengthening the connections create an impact upon the bilingual speech production process.

Table of Contents

List of tables

List of figures

Dedications

To my father, Iraj, for his love and measureless support from heaven
To my mom, Khorshid, for her unconditional love and support from miles and miles away
To my brothers, Habib and Hossein, for their endless love, support, and encouragement
And to my love, Reza, for being there for me in the best and the worst of times

I love you all

Author’s declarations

I hereby declare that this thesis is the result of my own work and I am the sole author. The work referred to in this thesis has not been previously submitted by the author for an award at this, or any other, university or institution. All sources are acknowledged as References.

Acknowledgements

Writing up this thesis would not have been feasible without the support of people I have had behind me, cheering me on what I needed it most. To all of you, I owe huge thanks.

First, I would like to express my deep appreciation and gratitude to my wonderful PhD supervisor, Dr. Zoe Handley, who has been with me throughout the entire process from the early stages of conception to the last point of finalizing the thesis. She has challenged me to find my voice, supported me at every single stage of my studies, provided me insightful comments, constructive feedback, and I am eternally grateful.

I would also like to thank Dr Nadia Mifka-Profozic, my Thesis Advisory Panel (TAP) member for her support, advice and remarkable comments on the documents presented in the TAP meetings. I would also to express gratitude to Dr. Mohammad Javad Ahmadian for all his thoughtful and constructive input and remarkable comments on my methodological design and perspective.

Central to the aim of my study was that the research was grounded in pedagogical authenticity; meaning, I was working with genuine students in intact classes. Hence, I am indebted to the students involved in this study, the pilot phases, and all of the students whom I experimented with the tests and techniques described in this thesis. I would like to thank Dr. Ali Raeesi, the language school director and supervisor, for having confidence in me and allowing me to carry out this research in the school.

I wish to give my special gratitude forever to my amazing family, my mom whose love, even in her physical absence, always lives on in me. My two brothers, for their never-ending emotional support, and for always telling me how proud they are of me. There are no words to let me express how much I love you and I am thankful for you.

My deepest and most heartfelt thanks go to my partner, Raza. Thank you for your enduring love, commitment, patience, and companionship from the beginning to the very end of this research project. Thank you for your kind words of affirmation, concern for my safety and always providing for, and being there whenever I needed you. You stayed true to your words; 'you need not worry about a thing for as long as you have me. I am always here for you'. Every single word of this thesis I have written with you in my mind and your love in my heart though we were miles apart most of the time. I could not have gotten this far without your love and support. During my late-night studies on campus, walking back home in all four British weathers at odd hours in the dark, you were the company and blanket of safety I felt. I know the distance between us and this journey as a whole was not easy, but hey, we made it.

Many thanks to my PhD colleagues and friends at the University of York for keeping me sane and optimistic;

To Ayse, for late nights in Derwent and being my statistics buddy,
To Hatice, for being my mom and making me delicious soups when I didn't even have time to cook,
To Nadia, for being more than a friend, for brightening my view when all I could see was darkness,
To Shiva, for being truly supportive, helping me in revising my thesis
To Sihem, for being the kindest, with her sweetest French accent saying 'Bonne nuit Sara Ebrahimi',
To Sina, for helping me settle in York during the first year of my PhD, and being there for me in the worst of the times
To Shermin, for motivating me and giving me confidence,
To Yara, for being with me all during this PhD, most particularly the last few months in the hell of times, If I could give our friendship a name since the first days in Berrick Saul, it would be *Memories Unlimited,*
To Zozo, for being the most helpful and supportive friend, helping me to move from one house to the other,
To Kathy and Agata, for being early morning and late-night study buddies at Berrick Saul

And last but not least, to Jessica, for being with me during the Covid-19 lock down and didn't let me feel lonely and scared at home.

1. Chapter 1: Introduction

1.1. Rational of the study

Every day we spend long hours talking and conversing to others, narrating stories, describing, explaining, complaining, and many other activities dependent on speaking skill. It is an ability which is never steady, since very early childhood to older stages. This ability is expanding to satisfy different needs at different phases of life. Speaking can be defined as an interactive process of constructing meaning which involves receiving, processing, and producing information (Brown, 1994; Burns & Joyce, 1997).

What many L2 language learners strive for is being able to become competent in the language they are learning. That is, EFL learners hope for gaining high proficiency and full mastery of English language, which can be conceptualized as "the highest point on a scale that measures the spoken command of a foreign language" (Lennon, 1990, p. 389). More specifically, they have a strong desire to speak the language smoothly and effortlessly, without paying conscious attention to its production. In other words, EFL learners have an ultimate goal of thinking and talking simultaneously in the second language[1] under the real time pressure as they do in their first language; that is, to attain L2 speech fluency (De Jong & Perfetti, 2011). However, with all effort they make, it is still not an easy task to achieve, and they grapple with it even after long years of L2 study.

One of the main factors contributing to the importance of oral fluency for EFL learners originates from the dominance of the communicative approach in teaching besides the salience of the ability to communicate the intended meaning clearly in real time for both educational and professional purposes (De Jong, 2018). The substantial importance of fluency in language teaching context has also affected the criterion of both high-stake and low-stake tests, such as Foreign Language internet Based Test (TOEFL iBT), the International English Language Testing System Academic (IELTS), and the Pearson Test of English Academic (PTEA), against which successful language

[1] Second language (SL), learning, and teaching terms are used interchangeably with L2 learning/teaching and foreign language (FL) learning/ teaching in this thesis.

communication is evaluated (De Jong, 2016; 2018; Fulcher, 2003). Accordingly, the highlighted role of fluency within such tests of English further motivates language learners to pay closer attention to this aspect of language more attentively.

In addition to this all, various significant reasons highlight the contribution of understanding and probing fluency from the perspective of the listeners and their perception of dysfluent speech. First and foremost, attending to dysfluent speech is most often considered to be “tiring and annoying” for the listeners (Derwing, 2017, p. 247; Varonis & Gass, 1982), as a result of which they are reluctant to become engaged in conversations when the language is not smooth and fluent. The inevitable outcome of such a perception would be less communicative opportunities leading to deficiency and paucity of input and opportunities to speak, which are the prerequisites for learning a language. Furthermore, listeners perceive dysfluent speakers to be less intelligent and make negative judgments on the basis of their non-fluent language. Thomson and Isaacs (2011) reported strong correlation between temporal measures of fluency and listeners judgments of the speakers’ level of intelligence. Finally, fluency is considered to be a key factor in predicting both L2 speaking proficiency (e.g. Iwashita, Brown, McNamara, & O’Hagan, 2008) and overall proficiency (e.g. Baker-Smemoe, Dewey, Bown, & Martinsen, 2014) .

Given the importance of this elusive concept, one of the concerns in second language acquisition (SLA) research has been understanding L2 fluency in terms of its definition, how it can be achieved by language learners and improved by language teachers, and finally how it can be assessed in L2 language testing context. In order to respond to this growing interest in studying L2 fluency of speech, this area has been researched in great depth over the past few decades with regard to the conceptualization and operationalization from the perspective of the psycholinguistic processes involved in L2 fluent production (Freed, Segalowitz, & Dewey, 2004; Lennon, 1990; Riggenbach, 1991; Segalowitz, 2010, 2016), based on which various techniques, instructional activities, and tasks, contributing to the development of L2 fluency have been developed (e.g. Galante & Thomson, 2016; Seifoori & Vahidi, 2012; Tavakoli, Campbell, & Mccormack, 2016; Wood, 2010).

In spite of the continued interest in speech fluency, it appears that not much research has been carried out exploring the impact of various pedagogical techniques and interventions on the development of fluency, for the case of the learners who are studying the language in an EFL context (Tavakoli, et al., 2016), and this is more highlighted in regard to fluency-based pedagogical interventions focusing on the knowledge of the lexis (Hilton, 2008).

To date, researchers have provided tentative conclusions in relation to the association between various measures of knowledge of lexis and L2 fluency of speech. More specifically, a comparison of the linguistic knowledge of lexical items (Bosker, Pinget, Quené, Sanders, & de Jong, 2012; Bosker, Quené, Sanders, & de Jong, 2014; De Jong & Bosker, 2013; De Jong, Steinel, Florijn, Schoonen, & Hulstijn, 2013; Zareva, Schwanenflugel, & Nikolova, 2005) and the processing skill of lexical retrieval (DeJong & Vercellotti, 2015; Hilton, 2008; Kahng, 2014; Towell, Hawkins, & Bazergi, 1996) with L2 fluency of speech have been investigated, and it has been reported that productive vocabulary knowledge can be considered as a likely predictor of L2 speech fluency. Although the link between knowledge of the lexical items and real time spoken fluency is well-proven and evident, there remains areas not systematically investigated, resulting into unanswered questions.

Firstly, although lack of lexical knowledge and easy access to this knowledge have been argued to be the biggest impediments in the way of reaching fluent production of the language (Hilton, 2008), it is quite disappointing to see not many fluency-based pedagogical techniques and activities have been proposed to help EFL learners with developments in the knowledge of vocabulary, as a way to improve L2 fluency of speech. Hence, this study attempts to provide empirical support for the application of semantic mapping technique as a fluency-based activity, fulfilling two main aims: firstly, strengthening the connections among already known lexical items in the process of productive retrieval (as a result of higher frequency of production and co-activation), and subsequently, creating higher number of links within and between levels of knowledge representation.

Secondly, not any other study to the extent of my knowledge, have attempted to investigate the relation between the organizational structure of the mental lexicon, as one of the main knowledge stores in speech processing system, and L2 fluency of speech. Previous studies have primarily concentrated on exploring productive vocabulary knowledge by examining the number of lexical items stored in the lexicon; that is, the size/breadth dimension of knowledge, when relating it to the fluency of speech (e.g., Clenton, De Jong, Clingwall, & Fraser, 2020; De Jong, et al., 2013; Kahng, 2020; Uchihara & Clenton, 2020; Uchihara & Saito, 2019; Uchihara, Saitio, & Clenton, 2020). Based on such studies, generalizations have been made regarding the organization of the mental lexicon, suggesting that learners with larger vocabularies are equipped with more structured lexicon; therefore, they can retrieve the required lexical items faster to apply in real-time communication (Uchihara & Saito, 2019). However, the organizational structure of the mental lexicon, manifested in the number, strength, and length of the connections among the lexical items stored in the lexicon (Henriksen, 1999; Read, 2004), have never been researched out as an entity of its own. In addition, the organization of the lexicon might also have implications for fluent production of the language, which has never been thoroughly explored.

Accordingly, the current study aims to weave the lines between L2 fluent speech production and the structure of the bilingual mental lexicon from a network perspective, concentrating on strengthening the links among its items within the hierarchical system of knowledge. In accordance, given the strong desire of EFL teachers in helping students to achieve more fluent production of language, the particular technique of semantic mapping is proposed for the purpose of this study to maintain smoother and faster production of speech, that is more fluent outcome.

1.2.Aims and research questions

The current study was motivated by a strong desire to explore the precise nature of the impact of a pedagogical technique, semantic mapping, on the development of second language (L2) fluency among EFL learners, which was studied by investigating the overall organizational structure of the mental lexicon from a network perspective. More specifically, it aims to empirically investigate the effect of semantic mapping technique on the structural properties of the mental lexicon and subsequently explores whether the changes in connectivity and strength of the links among items bring about any developments in various temporal measures of L2 fluency of speech.

This study therefore has a dual purpose and motivation: to understand and advance knowledge of the relationship between L2 fluency and organization of one of the main knowledge stores in the speech processing system, the bilingual mental lexicon (theoretical direction), and to explore a practical way of integrating a lexical practice approach, that is semantic mapping technique, into EFL teaching practice as a way to improve fluency as a result of the underlaying changes in the lexicon (a pedagogical direction).

This overarching aim was achieved by targeting the following objectives: (1) exploring the impact of semantic mapping on the structure of the bilingual mental lexicon by assessing the overall quality of network knowledge (2) exploring the impact of semantic mapping in the sense of assessing L2 utterance fluency of speech by measuring empirically established means of gauging L2 fluency (3) investigating if there is a relationship between the structural properties of the mental lexicon and L2 fluency of speech. A quantitative experimental design was employed to get a detailed understanding of the underlying processes and make judgments accordingly. To achieve the main overarching aim, the following contributing research questions were proposed to explore the impact of semantic mapping:

RQ1: To what extent does semantic mapping affect organization of bilingual mental lexicon as measured through free productive word association?

RQ2: To what extent does semantic mapping affect L2 speech fluency measured through empirically established means of gauging L2 fluency?

RQ3: Is there a relationship between organization of bilingual mental lexicon and L2 speech fluency?

1.3.Originality of the study

The originality of the study lies in the fact that it explores the impact of the level of structuredness of the bilingual mental lexicon on L2 fluency of speech by application of the semantic mapping technique. More specifically, it investigates the application of semantic mapping as a technique to provide changes to the organization of the lexicon, by providing support for the development of the depth of knowledge of the known lexical items, which is an extension of the application of the same technique as an instructional tool and teaching technique of new words only (e.g. Johnson, Pittelman, & Heimlich, 1986; Sánchez, 2004; Stahl & Vancil, 1986). Subsequently, it probes into any developments in L2 fluency of speech. Whilst the majority of research has been conducted applying semantic mapping technique as an instructional tool for teaching new words in EFL classroom context (Badr & Abu-Ayyash, 2019; Khoii & Sharififar, 2013; Kogok & Ahamed, 2017; Margosein, Pascarella, & Plfaum, 1982; Mohammed & Malo, 2020; Morin & Goebel, 2001; Saragih, 2019; Schmitt & Schmitt, 1995; Stoller & Grabe, 1993; Svenconis & Kerst, 1994; Zahedi & Abdi, 2012), this technique has rarely been explored as a way to provide practice opportunities for students with the lexical items they have already met, with the aim of creating changes to the structural properties of the lexicon.

Drawing on literature, lexical items are argued to be one of the greatest impediments in production of fluent language (De Bot, 1992; Hilton, 2008), and sematic mapping has been argued to be an effective instructional tool to increase the knowledge of lexical items; however, its impact on L2 fluency of speech have not been explored. There is only one single study (Ghonsooly & Hosseinpour, 2009) conducted with Iranian EFL learners assessing the impact of semantic mapping on the spontaneous use of vocabulary and fluency of speech. This study, however, investigated the application of semantic mapping as a way of teaching new vocabulary items. Besides, although this study claimed that improvements in oral fluency is the direct result of increase in vocabulary knowledge, it was not empirically tested and proved. Finally, the effectiveness of this technique has always been discussed and justified by relying on the depth of processing hypothesis (e.g., Ghonsooly & Hoseinpour, 2009; Khoii & Sharififar, 2013; Mohammed & Malo, 2020; Sagarra & Alba, 2006; Zahedi & Abdi, 2012), suggesting that the efficiency of this technique is deeply rooted on the grounds that the more semantic manipulations involved with a lexical item, the more elaborate routes will be available for subsequent retrieval (Craik & Lockhart, 1972; Craik & Tulving, 1975); however, the current study investigates the efficacy of the technique also by looking into the theory of L2 lexical development.

As it is evident, different studies applied semantic mapping technique as an instruction tool to teach new vocabulary items and also a rare number of studies assessed its impact on fluency of speech (Ghonsooly & Hoseinpour, 2008) and organizational structure of lexical knowledge (Sánchez, 2004); however, no study has explored the efficacy of the technique, as a way to productive retrieval of the lexical items, on developments of L2 fluency of speech as a result of changes in structural properties of the bilingual mental lexicon.

1.4. Context of the research

Iran is a Middle Eastern country with compulsory school attendance for all boys and girls from the age of six. The predominant and official language of the country is Persian (Farsi), and English is instructed as a foreign language (FL) and an obligatory subject during the last seven years of K-12 program in this country, starting from Grade 2 of the junior high school. In addition to schools, opportunities for learning English are also provided in private language schools all over the country due to the high demand of learning this language in more practical ways. As

Iran is an Islamic country, schools are sex segregated by the government; however, the case for private language schools is totally different. Children and adults from different genders can sign up in language schools all over the country and sit in the same class, regardless of their gender.

Mastery of speaking skill and the ability to use English for communication is one of the main reasons many Iranians study English, and accordingly the effectiveness of the courses taken and success of learning in such context is mainly evaluated base on how well they have improved in speaking proficiency (Sadeghi & Richards, 2015). Despite the high popularity of this skill among language learners, Curriculum Development Centre at the Ministry of Education in Iran primarily design coursebooks based on Reading Method (RM) and Situational language teaching (SLT) methods of teaching. Students coursebooks are mainly composed of long reading passages along with several grammatical points and a list of vocabulary to be memorized. Speaking and conversational skills receive no emphasis in the formal education at all. Overall, the curriculum designed for learning English at schools have a set goal, which is preparing students for the English section of the National University Admission Examination.

More recently, the materials used in classrooms for teaching English have started undergoing major shifts based on recent accounts of planning, implementation and evaluation of materials and move towards general English proficiency instead (Foroozandeh & Forouzani, 2015), incorporating communicative language teaching principles. That is, the pre-university coursebook, which was the first undergone changes in 2003, include more activities with the main purpose of enhancing communicative skills. However, there still exist challenges in the implementation of the material causing unsatisfactory results, particularly regarding speaking skill. For instance, factors such as limited number of hours for teaching English in the schools' curriculum (2 hours per week for a total of 26 weeks in an academic year), large number of students in classes from different English language backgrounds, and also heterogeneity of the population of students in classes would make it very difficult to practice speaking. Consequently, it appears that in spite of the growing desire in learning the language for communicative purposes, very few schools have been successful helping students to achieve this goal (Aghagolzadeh & Davari, 2014).

More recently, it seems that studying English has proven "itself to be a necessity, rather than a mere school subject" (Zandian, 2015,p. 113), as it is a vehicle to professional and educational advancement (Farhady, Hezaveh, & Hedayati, 2010). Accordingly, growing number of private language schools have been established all around the country. In order to sign up for taking classes in private language schools, language learners have to sit for a placement test prior to registration, and accordingly, they are assigned into different levels of language proficiency.

Despite the popularity of Reading method (RM) and Situational language teaching (SLT) methods of teaching at schools, the private sector has adopted primarily more communicative language teaching (CLT) approach as a core methodology in teaching English. Large number of books used in private language schools are very updated and recent books in English language teaching (ELT) market, such as *American English File, New Interchange Series, American Headway, English Result*, etc.

Although language schools provide higher chances for the language learners to develop their skills and become more proficient in English by offering great ELT resources and application of communicative based methodology, learners don't seem to be satisfied with the outcome as it doesn't meet their expectation (learning how to speak the language) (Mirhosseini & Khodakarami, 2015). In a study conducted in seventeen EFL classes in five different language schools in Shiraz, Iran, approximately only 14% of the total time of the class was spent on speaking activities (Razmjoo & Riazi, 2006). It can be summarized that although language schools in Iran claim to be preparing learners for communicative situations, this is not quite what goes as expected for the students (Mirhosseini & Khodakarami, 2015).

As an EFL teacher in classroom context of Iran, I am totally familiar with the realities of how speaking is practiced in classroom context. Speaking is predominantly practiced through the platform of "free discussion", in which students receive incidental and implicit teaching of oral language. Although the number of learners in classes are not large, often between 5 to 12, and it is very much feasible to allocate enough time to practice speaking, teachers mostly receive no training on how speaking should be practiced, and it seems that the same traditional, teacher-centered, product-based pedagogy, which is applied in delivering other skills and subskills, is practiced with speaking activities as well. Indeed, teachers recruited in language schools are required to pass compulsory training courses, involving both theoretical and practical perspectives of teaching prior to taking any classes; however, this is a single-shot workshop, which appear to be not fruitful. In addition, Iran's educational system is much exam-oriented and this is even evident in procedure of teaching and delivering content in language schools as well. Both formative and summative assessments are written exams including mainly sections on grammar, vocabulary, reading comprehension. Accordingly, speaking is only indirectly assessed through written pronunciation items (Farhady, et. al., 2010), which is another reason this skill is not much attended to in teaching procedure.

The problem even becomes more evident when it gets to L2 fluency of speech in Iranian language classroom pedagogy; however, this also seems to be true in a wider context of foreign language teaching in general, as Derwing (2017) posits that there seems to be a general agreement that "many L2 students do not have much opportunity to enhance their spoken fluency in classrooms" (p. 253). In the traditional context of Iranian classroom, L2 fluency is mostly practiced through the same speaking activities of "free discussion", which are applied to practice speaking skill. Accordingly, it seems that L2 fluency and speaking ability are used interchangeably. Tavakoli and Hunter (2018), investigating the ways EFL teachers' practice speech fluency in classroom, reported that "communicative free production activities" is by far the most popular type of activity used in classroom to promote fluency of speech (p. 339). Nevertheless, Rossiter, Derwing, Manimtim and Thomson (2010), reviewing the teaching materials in ESL context available to language teachers, argue that there is little empirical evidence supporting L2 fluency enhancement through the platform of free discussion; however, this is among the only ways fluency of speech is practiced in Iran context of teaching and learning.

Perhaps, it would be more realistic to acknowledge the gap between L2 fluency research and its practice in classroom in a wider context of second language teaching in classroom context. Therefore, the current study aims to address the gap in general, and in Iran classroom context in particular by providing chances for the EFL learners to receive further practice on vocabulary.

Vocabulary is often given attention in the Iranian classroom context of private language schools, and opportunities for learning new items are provided; however, teaching vocabulary is mainly incidental; that is, it occurs as a "by-product of a meaning-focused task" (Webb, 2020, p. 225). The main medium of teaching vocabulary is the reading texts in ELT coursebooks. While reading the texts, students are invited to look up the meaning of words they don't know in bilingual dictionaries (Hamzah, Kafipour, & Abdullah, 2009). However, student frequently memorize the meaning of the words and won't use it in innovative tasks later on (Mazdayasna & Molaei, 2015). That is, although chances for leaning lexical items are provided, opportunities for further practice in the development of lexical knowledge is not. As a result, many Iranian student do not have an affluent lexical knowledge even after years in learning English (Naeimi & Foo, 2014). Mazdayasna and Molaei (2015) suggested that students need guidance on how to expand the domain of their lexical knowledge as many difficulties they experience originate from "either limited vocabulary or its inappropriate application" (p. 54).

Consequently, the current study aims to address Iranian EFL students' difficulties in enhancing the fluency of speech by providing further chance to practice the lexical items they have already learnt through the application of semantic mapping technique.

1.5. Structure of the thesis

Having introduced the main aim of the current research along with introduction of the fundamental constructs and the context of the research, the remainder of the thesis aims to present an overview of the layout.

Chapter 2 presents a review and critique of the related research literature on four main concepts of L2 fluency (see Section 2.2), speech production and second language fluency (see Section 2.3), bilingual mental lexicon (see Section 2.4), Oral fluency and bilingual mental lexicon (see Section 2.5) and finally teaching L2 fluency (see Section 2.6). Initially, the first section (2.2.1) explores the construct of L2 fluency of speech and distinguish it from L2 proficiency. It is attempted to provide a detailed understanding of L2 fluency of speech by presenting its various conceptualizations from different perspectives. Subsequently, the concept of L2 fluency is approached from a more theoretical perspective, drawing on Segalowitz' (2010) framework (see Section 2.2.2). Accordingly, three senses of *perceived, utterance, and cognitive* fluency are discussed along with an explanation of the relationship between the three senses, drawing on various number of studies in this realm (see Section 2.2.3).

In order to examine fluency of speech comprehensively, the following section (2.3) presents models of bilingual speech production along with an in-depth discussion of the underlying cognitive processes, grounding on fluent production of speech. However, in the same section, various points of dysfluency, which can be originated from difficulties in the underlying processes of speech production, are discussed rigorously. After having discussed the process of speech production as well as the critical points in the process giving rise to dysfluencies, it is noted that the significance of the knowledge stores, that is their nature and function, in the process of fluent speech production is overlooked. Hence, the following section (2.4) explores bilingual mental lexicon as one of the most significant and influential knowledge stores in the process of fluent speech production. In order to gain a clearer understanding of the bilingual lexicon, this thesis has adapted a network perspective rather than probing into the properties of single lexical items stored in this knowledge store. Accordingly, various levels of knowledge representation (conceptual and lexical levels) are introduced, along with a discussion of the relationship between the items stored in each of these levels as well as the strength of the connections, which are all explained in the light of spreading activation model. Finally, at the end of this section (2.4.4), word association task is introduced as a reliable means of measuring the network knowledge, in order to explore and find out about the structure and organization of this multi-layered integrated network.

Consequently, the chapter ends bringing the concepts of fluency and lexical knowledge together (see Section 2.5), while addressing the pedagogical aspect of fluency, and proposing a creative approach of semantic mapping to implement both theories of speech fluency and mental lexicon in a pedagogical manner (see Section 2.6).

Following this, chapter 3 presents the research methodological approach employed in the current study to answer the research questions proposed. It begins by discussing the research strategy adopted, and justifications for the application of experimental design (see Section 3.2). Following this, the participants of the study are introduced (see Section 3.3). Next, the design of the study adopted to answer the research questions effectively is explained rigorously (see Section 3.4). Subsequently, the assessment instruments (C-test, picture narrative, productive word association task) and intervention material are explained (see Section 3.5) along with the procedure followed (see Section 3.6). In section 3.7, the pre-analysis procedures adopted prior to the main phase of data analysis are explained in two subsections. Firstly, the pre-analysis procedures of data obtained from the word association task is explained. which is followed by description of data collected from picture narratives. In particular, comprehensive information on how the speech samples collected from participants, analyzed using PRAAT computer software, is presented (Boersma & Weeninck, 2018). Section 3.7.2 presents the dependent variables of the study; that is, the measure of overall word association score and various measures of L2 utterance fluency (syllable run, phonation run, syllable duration, and mean duration of silent pauses). Each variable along with the way it is calculated for the purpose of the study, is presented thoroughly. In addition, the main reasons for their inclusion in the current study are justified. Consequently, the remaining of the chapter discuss the ethical considerations (see Section 0) and lessons learnt from conducting the pilot phase of the study (see Section 3.9).

As the productive word association task discussed in chapter 3 has been particularly developed for the purpose of the current study, chapter 4 presents a detailed account of word association behavior. As already mentioned, in order to meet the aims of the current study, it is decided to use word association task to examine the structure of the bilingual mental lexicon. Accordingly, this chapter presents a short review of related literature on word association behavior along with the main points of methodological concern, which are required to be considered while designing the test (see Section 4.2). Secondly, the design of the test is presented by explaining all the single steps taken one after the other in order to have a reliable and accurate instrument for eliciting data (see Section4.3). In the last part of this chapter (see Section 4.4), the coding scheme adapted from Henriksen (2008) was rigorously explained along with the changes applied to meet the general aim of the current study.

The principal purpose of chapter 5 is analytical. This chapter is divided into three subsections, each answering one of the research questions posed. Therefore, the first subsection (5.2) presents an overview of the word association test, in addition to presenting the quantitative analysis conducted to explore the impact of semantic mapping technique on the organization of the bilingual mental lexicon. Beside the quantitative analysis of word association results, a descriptive part is added to extend the analysis and discussion by exploring various type of associations provided by the participants in each word association test. This section is mainly added to probe into the underlying differences between type of responses provided as a single score cannot be as fully representative. Next, section (0) presents the results of the quantitative analysis of various measures of L2 fluency in order to investigate whether semantic mapping have a significant impact on L2 utterance fluency. Finally, in the last section (5.4) the results of the correlational analysis between overall word association test and measures of fluency is presented.

The discussion chapter (Chapter6) returns to the original hypotheses and investigate them in relation to the findings of other studies in the same realm. Potential explanations for any deviation from existing patterns in literature along with points of similarities are offered while supporting them with theoretical explanations.

Chapter 7 concludes the thesis by reiterating the main aims of the research (section7.2), a summary of the study (see Section 7.3) and the findings (see Section 7.4). Subsequently, the remaining part of the thesis presents the implications of the current study (See Section 7.5), including theoretical implications (see Section7.5.1), methodological implications (see Section 7.5.2), and pedagogical implications of the study (see Section 7.5.3), along with the limitations (see Section 7.6) and suggestions for future research (see Section 7.7). This thesis ends by sharing a final word with the readers (see Section 7.8).

2. Chapter 2: Review of Literature

2.1. Overview

One of the main aims of many language learners is to speak the language smoothly and effortlessly in the same manner with which they deliver talk in their native language. Nevertheless, vast majority of L2 language learners struggle in reaching this goal, even after considerable length of time in language classes with teachers tending to bypass efforts to facilitate the process. In fact, one of the most frequent comments made by L2 language learners is "although I totally understand it, I cannot say it in English" (Hunter, 2017). Accordingly, in order to be able to help language learners to reach this goal, understanding how speech is produced and studying the underlying mental processes involved in its production is highly significant. The current research mainly concentrates on one of the main knowledge stores in L2 language processing; that is, the bilingual mental lexicon and it aims to investigate whether the structure of this knowledge component has any significant impact of L2 fluency of speech. Accordingly, this chapter is divided into four main sections investigating; Second language fluency, Speech production and second language fluency, Bilingual mental lexicon, and finally Teaching fluency via consideration of the network structure of the lexicon.

To follow this path, the first main part of this chapter (see Section 2.2) presents an explanation of what L2 fluency of speech is and how it is defined in literature from different perspectives, while differentiating it from L2 proficiency (see Section 2.2.1). Next, in order to scrutinize L2 fluency from a theoretical perspective rather than focusing solely on practical approaches, Segalowitz's (2010) conceptualization of L2 fluency with a cognitive basis will be introduced (see Section 2.2.2). This triple conceptualization is a distinguished view moving the discussion beyond the audibility features of fluency and take into account the underlying cognitive processes responsible for fluent production of language. Accordingly, in the following section (see Section 2.3), the underlying speech production processes in bilinguals will be explored by presenting current theories of bilingual speech production along with a discussion of the potential underlying processing difficulties, which might give arise to L2 dysfluencies in the speech processing system. In this section, it will be argued that linguistically encoding one's intentions to produce fluent language in such a complex system of speech processing is one of the most difficult activities to be carried out in a manner not requiring conscious attention, that is automatically, and all this process is more critical in case of second language speakers (Kormos, 2006a); hence, various rotes to automatic and parallel processing of language are presented.

Based on the models of bilingual speech production and the proposed "critical points in the architecture of speaking system" (Segalowitz, 2010, p. 17) giving arise to fluency difficulties, it appears that points of dysfluencies originating from the nature and the function of knowledge stores in the speech processing system has remined overlooked. That is, the processes involved in the production of fluent language has been in the spotlight of extensive research while fluency problems associated with the organization and properties of knowledge components have been neglected in the literature. Accordingly, mental lexicon is introduced as one of the most influential knowledge stores in speech processing system, since it is the store for lexical items which are regarded as "the greatest impediment to spoken L2 fluency" (Hilton, 2008, p. 162), and also based on the theories of automaticity discussing that research on the processes of automatic lexical encoding and retrieval has been limited in literature (Segalowitz, 2003).

Following this, the third main part of the chapter discusses the bilingual mental lexicon (see Section 2.4). In order to examine the bilingual mental lexicon, a network perspective is adapted, concentrating on a deep understanding of the lexicon rather than attending to the individual lexical items, as there is a "danger of losing sight of the wood through concentrating too hard on the individual trees" (Meara, 1996, p. 9). In light of this argument, bilingual mental lexicon is introduced as a gigantic network of connections and associations, with two levels of knowledge representation; conceptual level and lexical level (see Section 2.4.1). Next, the contribution of different levels of knowledge representation in the process of lexical development and processing in the bilingual mind is discussed, while concentrating on a model taking the practical constraints of L2 lexical development into account (see Section 2.4.2).

Accordingly, these levels of knowledge form a multi-layered integrated network, with information spreading not only within each level of knowledge representation but also between the two levels, which is discussed in the light of spreading activation model (see Section 2.4.3). Consequently, the way information is flown in this network through various number node of items and connections with varying degrees of strength among them have strong implications for parallel and automatic production of the language; as a result of which, it is of outmost importance to be able to measure the network knowledge to come to conclusion regarding the structure and organization of the bilingual lexicon, which is possible via the application of word association task (see Section 2.4.4).

Subsequently, in the last section (see Section 2.5), the growing body of research and empirical studies in the area are presented. Apart from the theoretical significance of the lexical items in the process of fluent production of L2, the empirical studies conducted on the central properties of mental lexicon (that is, dimensions of lexical knowledge from a lexical research perspective) are scrutinized, and the relationship between the two dimensions of breadth and depth of knowledge and L2 fluency of speech are presented.

Following this section (see Section 2.6), implementation of the theories of fluency and bilingual mental lexicon to language teaching and learning context and the pedagogical implications are presented by a discussion of Nation's (2008) strands with respect to vocabulary learning and teaching, while diverting the focus of attention from the techniques normally associated with improving fluency such as repeated practice. The current study introduces a creative approach of semantic mapping (see Section 2.6.1), incorporating theories of bilingual mental lexicon, providing L2 language learners with a type of practice reducing difficulties with fluency of speech, in congruence with models of speech processing, suggesting that the structural properties of the bilingual lexicon might also be a way to support L2 fluency of speech.

2.2. Second language fluency

"What does it mean, in ordinary language, to say that someone is fluent in an L2?" (Segalowitz, 2010, p.3). This is a basic question that is of considerable interest to many language learners, teachers, testers, researchers, and material developers. The challenge posed by such question is grounded on the complex and multifaceted nature of the concept of fluency in regard to its definition, ways of improvement, and measurement. Hence, the first section of this chapter tries to provide a clear understanding of L2 fluency by defining it in qualitative and quantitative terms from various perspectives in literature. This part firstly involves defining fluency by contrasting it from language proficiency, presenting a dichotomous and binary perspective of broad and narrow senses of fluency (Lennon, 1990), as well as putting forward the interrelated approaches grounded on teachers' point of view on L2 fluency (Tavakoli & Hunter, 2018). The second subsection explores how L2 fluency can be defined, categorized, and subsequently measured drawing on Segalowitz' (2010) framework, by discussing the interrelated conceptualization of three dimensions of perceived, utterance, and cognitive lenses of L2 fluency, which was originally an attempt to consolidate various senses into a single model.

2.2.1. L2 Fluency versus proficiency

One of the most common terms in second language teaching context is fluency of speech. Yet in spite of its popularity, it is not an easy concept to define precisely. Various ways of defining and categorizing L2 fluency of speech have been proposed (Lennon, 1990; Segalowitz, 2010; Tavakoli & Hunter, 2018). Most often in ordinary life, fluency has been considered equivalent to overall language proficiency, referring to a second language speaker "having a good command of a language" (Chamber, 1997, p. 536). Speaking fluently in this sense refers to "error-free grammar, a large vocabulary and/or native-like pronunciation" (Bosker, et al., 2012, p.160). This extended interpretation of fluency shall be differentiated from the one used by language specialists (e.g. researchers, teachers, curriculum designers, etc.) in a much narrower sense, being only one of the essential components of overall second language proficiency (Kahng, 2014) and one of the most prominent differences between L1 and L2 speakers (Kormos, 2006a). Despite the importance of the narrow sense of L2 fluency in the domain of language assessment and proficiency, there is a lack of clarity in its definition and language of understanding of the concept (Kahng, 2014; Kormos & Dénes, 2004; Segalowitz, 2010).

The aforementioned senses of L2 fluency, either as overall language proficiency or one of the components of language proficiency, best fits into the traditional Lennon's (1990, 2000) distinction of fluency as broad and narrow senses. In this categorization, fluency in board term is characterized as overall speaking proficiency (Chambers, 1997), including both accuracy and complexity of the utterance (De Jong & Perfetti, 2011). Likewise, Sajavaara's (1987) definition of fluency can be categorized as same broad conceptualization of fluency, defining it as “the communicative acceptability or the speech acts, or ‘communicative fit’” (p.62). This sense of L2 fluency might be considered difficult to operationalize, unlike the narrow sense, as it includes variety of factors to be categorized as fluency and it mostly covers qualitative issues in defining it. Besides, this sense includes both the language proficiency as well as speaking proficiency, which are two very different phenomena. In addition, as mentioned by Derwing (2017), this type of fluency is “not the intent of most applied linguistic studies” (p. 246). Rather, the main point of interest is the narrow sense of fluency, which is restricted to the degree of speech flow and the extent to which the flow is interrupted (Derwing, 2017).

Lennon’s (1990) definition of the narrow sense is entirely a performance-based phenomenon. Fluency is considered as one of the components of oral proficiency. Lennon (2000) defined it as “rapid, smooth, accurate, lucid, and efficient translation of thought or communicative intention into language under the temporal constraints of on-line processing” (p. 26). Another definition, which is a clarification of narrow sense of Lennon’s one, was proposed by Schmidt (1992), labelling fluency as “automatic procedural skill”, due to the fact that "the psycholinguistic processes of speech planning and speech production are functioning easily and efficiently" (Lennon, 1990, p. 391); hence, it calls for different type of rapid and quick access to input from different knowledge stores, such as the mental lexicon (Chambers, 1997), and is characterized primarily by automatic and parallel processing (Levelt, 1989). Although the narrow sense concentrates on specific aspects, it still reflects the application of various perspectives of linguistics knowledge, neurological and muscular mechanisms (Segalowitz, 2010).

Lennon's (1990, 2000) definition of L2 fluency proposed a dichotomous and binary categorization, while a more recent detailed study put forward four different associated and interrelated approaches based on teachers’ appreciation of fluency, presented in Figure ***2.1***. It has been proposed that fluency of speech “often seem to inhabit the space somewhere between a broad and narrow definition" (Tavakoli & Hunter, 2018, p. 341). In this study, English language teachers were asked to define fluency and their perception was analyzed by the use of questionnaires both qualitatively and quantitatively. The results show that a new framework of fluency can be developed, a framework of four different types of very broad, broad, narrow and very narrow speech fluency.

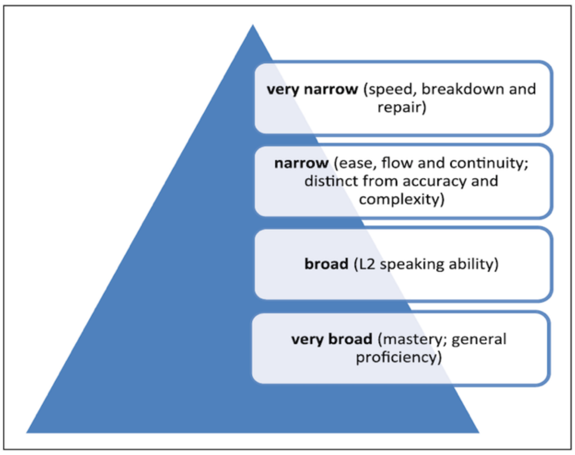


Figure 2.1 Approaches to defining fluency (Tavakoli & Hunter, 2018)

Two of the largest approaches to defining fluency are allocated to two categories of very broad and broad. At the bottom of the pyramid, the very broad approach, defines fluency as L2 proficiency in general, which can be used interchangeably with 'proficiency' or 'mastery'. The second largest category characterizes a fluent speaker as one "who can speak confidently and communicate their intended message well in the spoken mode" (Tavakoli & Hunter, 2018, p. 343). The third perspective to defining fluency is related to "ease, flow, and continuity of speech" (p. 343), which is very close to the definition of fluency proposed by Koponen and Riggenbach (2000) suggesting that fluency is the "flow, continuity, or smoothness of speech" (p. 6). Finally, the last perspective, which only attracted small proportion of teachers' answers, was surprisingly devoted to fluency as the objective measurement of speech in terms of temporal measures of speed, silence, and repair.

From the above-mentioned definitions of L2 fluency, it seems that fluency of speech in general includes both performance characteristics (being rapid and smooth) and linguistic competence (being accurate) simultaneously. The current study operationalizes the term L2 fluency from the very narrow perspective by taking only the performance characteristics of this definition, as it concentrates on the ease with which information, particularly lexical items, can be accessed and retrieved to communicate meaning, which is parallel to the definition by Housen and Kuiken (2009). They defined fluency as "learners' control over their linguistic L2 knowledge, as reflected in the speed and ease with which they access relevant L2 information to communicate meanings in real time" (p. 462).

Definition of fluency in this section concentrates on "atheoretical perspective" (Segalowitz, 2010, p. 5), focusing mainly on the practical approaches in defining it as a measure of smoothness, fluidity, and effectiveness of speech in all senses. In the next section (see Section 2.2.2), fluency is conceptualized from a larger context of cognitive science perspective inquiring into the nature of language, proposed by Segalowitz (2010), which goes beyond mainly the audibility feature of fluency of speech.

2.2.2. *Threefold lenses of fluency*

For long fluency was securitized in the narrow sense as mainly being related to the ease and smoothness of speech delivered as a component of L2 oral proficiency, till later L2 researchers in an attempt to unravel L2 fluency of speech in its narrow sense resort to cognitive perspective, arguing that fluency in this sense is a reflection of using linguistic knowledge along with "execution of the neurological and muscular mechanisms that a speaker has developed over an extended period of time through socially contextualized, communicative activities" (Segalowitz, 2010, p.7).

In such a paradigm, there exists three different ways of interpreting the single term of L2 fluency, covering all three senses more or less in an imprecise manner. Segalowitz (2010) clarifies the point on these three senses by making an example; "Noriko speaks Inuktitut quite fluently for a Japanese" (p. 47). Different people can have different interpretations of this single sentence hearing it. Such a sentence can mean;

- She has the "ability to mobilize her cognitive system for speaking in a highly effective and efficient manner", that is similar to talking in a native language.
- The sentences she produces has certain characteristics of speech flow, such as pauses, hesitations, speed of delivery, repairs and reformulations, etc.

- It can also target the listeners hearing Noriko talking and their perception of her as having "highly efficient cognitive skills for speaking the language"

These three interrelated lenses of L2 fluency, as categorized by Segalowitz (2010) into cognitive, utterance, and perceived fluency, are defined in Table ***2.1***.

Table 2.1 Three lenses of L2 fluency (Segalowitz, 2010)

Cognitive fluency	"the efficiency of operation of the underlying processes responsible for the production of utterances"
Utterance fluency	"the features of utterances that reflect the speaker's cognitive fluency"
Perceived fluency	"the inferences listeners make about speakers' cognitive fluency based on their perceptions of their utterance fluency" (p. 165)

The present study aims to embrace Segalowitz's view of L2 fluency, while taking a cognitive basis. To this end, the three lenses of L2 fluency in their narrow sense are discussed in depth in the following sections.

2.2.2.1. Perceived fluency

Drawing on Lennon's (1990) claim on fluency of speech, "an impression on the listener's part that the psycholinguistic processes of speech planning and speech production are functioning easily and efficiently" (p. 391), it is easy to see an allocation of a central role to the importance of the listener's perception on the speaker's fluency. Such interpretation is based on making subjective judgments of the speakers' fluency based on the efficacy of the underlying cognitive processes of speech production, which is represented in the speech sample. This sense is mostly presented in transferring and communicating ideas without hesitations and pauses in an acceptable speed of delivery, which are the actual properties of an utterance, and embraces Segalowitz's (2010) dimension of a narrow sense of perceived fluency. Though another interpretation of perceived fluency shapes up consulting into second language acquisition research, referring to a broader interpretation. Perceived fluency in its broader sense represents the subjective judgment listeners make based on their own idiosyncratic explanation of what fluency is; thus, each listener can interpret a sample of speech differently from other listeners and judge accordingly.

In order to reduce the ambiguity in understanding and application of perceived fluency in terms of how it is interpreted, providing instructions to the listeners can be a solution. Surprisingly, fluency judgements from the raters, who are regarded as experts and trained, are not significantly different from the untrained natives (Bosker et al., 2012; Derwing, Rossiter, Munro, & Thomson, 2004; Rossiter, 2009); as long as they are instructed what to focus on (De Jong et al., 2013); otherwise, different raters' perceptions of what constitutes fluency can bring various interpretations into the ground (Kormos & Dénes, 2004; Cucchiarini, et al., 2002), as they refer back to their own set criteria of what fluency is.

2.2.2.2. Utterance fluency

The second dimension of Segalowitz' (2010) triple lenses of fluency is utterance fluency. Utterance fluency can be defined as "the fluidity of the observable speech as characterized by measurable temporal features" (Segalowitz, 2016, p. 81). It is the quantifiable aspect of fluency, which has attracted substantial attention (e.g. Bosker, et al., 2012; De Jong, 2016; De Jong, et al., 2012, 2013; Derwing, Munro, Thomson, & Rossiter, 2009; Derwing et al. , 2004; Freed, 2000; Ginther, Dimova, & Yang, 2010; Iwashita, et al., 2008; Kormos & Dénes, 2004; Préfontaine, 2013; Riggenrach, 1991; Rossiter, 2009; Segalowitz & Freed, 2004; Towell, et al., 1996).

The main advantage associated with this dimension of fluency is the existence of variables reflecting utterance fluency, bringing about measurability and objectivity. It has been proposed that different measures of fluency can be brought together in terms of speed fluency, breakdown fluency, and repair fluency (Skehan, 2003) (see Table ***2.2***).

Table 2.2 Typical measures of fluency

	Typical measures
Speed	Speech rate, articulation rate, mean length of run
Break down fluency	Number and duration of silent and voiced pauses
Repair	Number of corrections, reformulations, repetitions, false starts

Subsequently, this framework was revisited and repackaged into temporal fluency, and repair fluency (Tavakoli & Skehan, 2005), as it was proposed that both speed and pausing characteristics of speech represent temporal aspects while repairs is a tendency to correct and reformulate ones' speech. An important point to be taken into account in this framework is the existence of measures which are related to one another, overlapping with other measures; that is, composite measures (Kormos, 2006a; Skehan, 2014; Tavakoli & Skehan, 2005). Speech rate capturing the speed and pausing aspects is among such measures (De Jong, et al., 2012). Using composite measures might lead to difficulties in interpretation of the results as it is not easy to identify if changes in fluency is caused and attributed to which of the measures assessed. Therefore, depending on the purpose of the study, one can choose among the pool of various measures, for instance, studies focusing on global measures of fluency are mainly interested in composite measures (Tavakoli, et al., 2016).

Accordingly, such taxonomies can be applied to categorize various indicators of utterance fluency into groups. In a review of the current literature, the most commonly proposed and employed measures proposed in literature along with their definition can be summarized in Table ***2.3***.

Table 2.3 Most common measures of utterance fluency

	Measures	**Definition**
Speed	Pruned speech rate	Total number of syllables divided by total time (See De Jong & Vercellotti, 2015; Derwing, et al., 2004; Freed, 2000; García Amaya, 2009; Ginther et al., 2010; Huensch & Ventura, 2017; Iwashita, et al., 2008; Lennon, 1990; Llanes & Muñoz, 2009; Préfontaine, 2013; Tavakoli, 2016; Tavakoli et al., 2016)
	Pruned syllable rate	Total number of syllables minus repair syllables divided by total time (See De Jong, Groenhout, Schoonen, & Hulstijn, 2015; Derwing et al., 2004; Rossiter, 2009; Smemoe et al., 2010)
	Articulation rate	Total number of syllables divided by phonation time (total time minus unfilled pauses) (See Cucchiarini, Strik, & Boves, 2000; De Jong & Perfetti, 2011; De Jong & Vercellotti, 2015; Ginther et al., 2010; Kormos & Dénes, 2004; Llanes & Muñoz, 2009; Mora & Valls-ferrer, 2020; Préfontaine, 2013; Préfontaine, et al., 2016; Suzuki & Kormos, 2019; Tavakoli, 2016; Tavakoli et al., 2016; Towell, et al., 1996)

	Syllable duration	Phonation time divided by total number of syllables (inverse of Articulation rate) (See Bosker, et al., 2012; De Jong, et al., 2013)
	Phonation time ratio	Percentage ratio of time speaking to time to take the whole speech sample (See Cucchiarini, et al., 2002; De Jong & Perfetti, 2011; De Jong & Vercellotti, 2015; Ginther et al., 2010; Kormos & Dénes, 2004; Mora & Valls-ferrer, 2020; Préfontaine, 2013; Tavakoli et al., 2016; Towell et al., 1996)
Break down	Silent pause rate	Total number of silent pauses divided by the total time (See Bosker et al., 2012; Ginther et al., 2010)
	Filled pause rate	Total number of filled pauses divided by the total time (See Bosker et al., 2012; Ginther et al., 2010; Iwashita et al., 2008; Kormos & Dénes, 2004; Tavakoli et al., 2016)
	Pause rate	Total number of pauses divided by the total time (See Préfontaine, 2013; Préfontaine, Kormos, & Johnson, 2016; Tavakoli, 2016)
	Mean length of silent pauses	Total silent pause duration divided by number of silent pauses (See Bosker et al., 2012; Cucchiarini et al., 2002; De Jong, Schoonen, & Hulstijn, 2009; De Jong et al., 2013; Ginther et al., 2010; Hilton, 2009; Kahng, 2014; Kormos & Dénes, 2004; Riazantseva, 2001)
	Mean length of filled pauses	Total filled pause duration divided by number of filled pauses (See Ginther et al., 2010; Hilton, 2009)
	Mean length of pauses	Total pause duration divided by the total number of pauses (See De Jong et al., 2013; De Jong & Perfetti, 2011; Préfontaine, 2013; Préfontaine et al., 2016; Tavakoli, 2016)
	Mean length of run/ Syllable run	Total number of syllables divided by the total number of pauses (See Baker-Smemoe et al., 2014; Brien, Segalowitz, Freed, & Collentine, 2007; Cucchiarini et al., 2002; De Jong & Perfetti, 2011; De Jong & Vercellotti, 2015; Derwing et al., 2004; García Amaya, 2009; Ginther et al., 2010; Huensch & Tracy–Ventura, 2017; Kormos & Dénes, 2004; Lennon, 1990; Mora & Valls-ferrer, 2020; Préfontaine, 2013; Préfontaine et al., 2016; Segalowitz & Freed, 2004; Tavakoli et al., 2016; Towell et al., 1996)

Repairs	Mean number of dysfluencies	Total number of dysfluencies (repetitions, restarts, repairs, false starts) divided by the total time (See Baker-Smemoe et al., 2014; Bosker et al., 2012; Cucchiarini et al., 2002; De Jong & Vercellotti, 2015; Huensch & Tracy–Ventura, 2017; Suzuki & Kormos, 2019; Tavakoli, 2016; Tavakoli et al., 2016)

The main reason for such a long list of measures, along with many more not included in this list, is the very fact that discussion on the best indicators of utterance fluency and their operationalization have failed to reach a consensus, and the very primary reason is that fluency is far from "a single unitary concept" (Koponen & Riggenbach, 2000, p. 17). In the present study, in order to find out which measures are more sensitive to pedagogic intervention, and are able to manifest the underlying changes in the cognitive processes, a range of above measures are applied, which will be discussed in more depth in chapter 3 (see Section 3.7.2.2).

2.2.2.3. Cognitive fluency

Having explored two senses of perceived and utterance fluency, the last dimension of the triple view is cognitive fluency. It applies to the speaker's ability to accommodate and mobilize various processes that underlie speech production towards production of language in real time communication (see Section 2.3). Cognitive fluency is characterized by the speaker's abilities to "efficiently plan and execute his speech" (De Jong et al., 2013, p. 894). The high speed and efficacy of the processes are carried out with regard to lexical access and also application of linguistic resources (Segalowitz, 2016). The underlying mechanism responsible for cognitive fluency is automaticity, referring to the absence of attentional control in the process of the simultaneous activities of planning and execution of speech (Kahneman, 1973), which will be discussed more profoundly in section 2.3.5.

2.2.3. The relationship between three lenses of fluency

Within the realm of second language acquisition, bulk of research has concentrated on the potential relation between the three senses of fluency; that is, the objective measures of utterance fluency, subjective ratings of perceived fluency, and measures of underlying cognitive fluency.

Accordingly, many studies in this domain share same overarching research questions: "what can objective measures of a speakers' utterance fluency tell about their measures of cognitive fluency?" or "what can subjective measures of a speakers' utterance fluency tell about their underling cognitive processes?" In this section, the possible relation between different senses of fluency and the related pedagogical studies conducted are discussed.

2.2.3.1. Utterance fluency and perceived fluency

As it has already been discussed, the distinction between two senses of utterance and perceived fluency, both aiming to represent the aspects of effectively translating thoughts to speech, is the way each sense is gauged. Utterance fluency is a sense which is objectively measurable while perceived fluency is subjectively gathered from ratings (Segalowitz, 2010). Relating these two senses reveal "the aspects of fluency that are perceptually salient and the aspects of (fluency in) the speech signal listeners deem important" (De Jong, 2018, p. 242).

In an early study, Riggenbach (1991) sampled speech from only six nonnative speakers, three of whom were rated as fluent and three as nonfluent, in terms of (hesitation, repair, and speech rate), and investigate the relation between utterance and perceived fluency. The results obtained are in general compatible with the findings in other later studies, revealing that measures of speed and breakdown are strong indicators of perceived fluency (e.g. Bosker et al., 2013; Cucchiarini, Strik, & Boves, 2002; Derwing et al., 2004; Kormos & Dénes, 2004; Lennon, 1990; Préfontaine et al., 2016; Rossiter, 2009; Saito, Ilkan, Magne, Tran, & Suzuki, 2018; Segalowitz & Freed, 2004).

Later, Kahng (2018) focused on the qualitative difference between L1 and L2 speech regarding the distribution of pauses as one of the significant indicators of perceived fluency and tried to find out if the location of pauses has any impact on the perception of fluency. The results suggest that the listeners are more sensitive to the frequency of mid-clause pauses. The results of this study was in line with Suzuki and Kormos (2019), presenting similar findings on the strong association between frequency of mid clause pauses and perceived fluency. Likewise, Saito et al. (2018), examining spontaneous speech of 90 Japanese learners, found out that frequency of mid-clause pauses can distinguish between mid- and high-level performance.

Although all above studies are in favor of the relative contribution of the subjective ratings in explaining utterance fluency, the results should be interpreted with caution as the multicollinearity of the fluency measures have been neglected, which in turn can lead into spurious relation between the two senses (De Jong, 2018). In a creative attempt to overcome this obstacle, Bosker et al. (2014) purposefully manipulated the speech samples collected for different measures of fluency, controlling for the confounding factors, prior to rating. On this account, the variance in the results could be attributed to fluency differences between the different conditions. The results suggest frequency and length of pauses along with speech rate are the measures having a significant impact on perception of fluency.

Taken all together, it seems that the association between aspects of fluency, which are perceptually salient and the efficient ability of translating thoughts to speech gauged objectively, is less straightforward than assumed, and the results of the studies, depending on the research methods applied, shall be interpreted with caution (De Jong, 2018).

2.2.3.2. Utterance fluency and cognitive fluency

Until recently, not many empirical studies have explored the underlying cognitive processes in charge of fluent production of language in relation to utterance fluency; however, more recent literature seems to concentrate on this aspect in L2 fluency research. To meet this end, two different approaches have been adopted to investigate the relationship between cognitive fluency and measures of utterance fluency.

The first approach is through careful consideration of various types of knowledge and processes involved in speech production by operationalizing various elements of L2 cognitive fluency and relating them to measures of utterance fluency. Segalowitz (2010) proposed that speed, processing stability, and attentional flexibility to be a manifestation of the cognitive processes underlying L2 use. In an early attempt, Segalowitz and Freed (2004) examined the association between both speed and efficiency of lexical access (through semantic classification task) and attention control (via speed responses of matches and mismatches to cue words) as measures of the underlying processes and measures of fluency of speech. The results of analysis revealed a significant relationship between both the speed and efficiency of lexical access with mean length of fluent runs, while duration of pauses was not associated with any of the measures of L2 cognitive fluency. Later in a large-scale study, De Jong et al. (2013) aimed to find out the measures of utterance fluency which are the best manifestation of the underlying processes. The 179 participants taking part in this study, were asked to complete 8 monolingual speaking tasks to extract measures of utterance fluency. In addition, various L2 linguistic knowledge (vocabulary, grammar, and pronunciation) and processing skill tasks (morpho-phonological processing, lexical selection, and articulation) were designed based on theories of language production to tap on the underlying cognitive processes required for planning and execution of speech. The results indicated that mean duration of syllables (inversion of articulation rate) is the best indicator of L2 linguistic knowledge and processing skills, explaining as much as 50% of variance in the scores. In line with the results of the previous study, mean duration of pauses again proved to bear the weakest relation (approximately 5% of variance explained).

Examining measures of L2 utterance fluency, which are associated with L2 cognitive fluency in this approach, leaves to speculate if all L2 utterance fluency measures are reliable indicators of L2-specific cognitive fluency rather than manifesting L1 fluency behavior, paving the way in turn to the second approach in this domain. This approach is less direct, exploring the relation between cognitive fluency and utterance fluency by investigating different aspects of L2 utterance fluency, which are associated with speakers' L1 fluency behavior. That is, in case L2 measures of utterance fluency can be predicted from L1 measures, then they can be mostly attributed to idiosyncratic speaking styles rather than underlying cognitive processes. De Jong et al. (2015) examined two groups of Turkish and English L1 speakers, learning Dutch as their second language, and found out nearly all measures of fluency can be predicted by means of L1 fluency behavior. Among various measures, mean silent pause duration was the strongest reflection of personality traits (accounting for only 0-3% of the variance of L2 proficiency), while syllable duration was only moderately associated with L1 behavior and accounted for 21% of the variance from the L1 measure. Derwing et al. (2009) likewise argue that "close relationship between a speaker's L1 and L2 temporal characteristics would suggest that fluency is governed by an underlying trait" (p. 533). In accordance with the above studies, Kahng (2020), replicating De Jong et al. (2013) while correcting L2 cognitive measures for L1 baseline measures, suggests that "L2 utterance fluency is dependent upon L2 cognitive fluency and has also been shown to be related to L1 utterance fluency" (p. 476). The results of this study on 44 Chinese learners of English in both their L1 and L2 revealed that mean syllable duration can be considered as a gauge of L2-specific cognitive fluency measure as it revealed high correlation with both processing skills of speed of syntactic encoding and lexical retrieval. On the other hand, mean silent pauses duration and number of filled pauses are largely corresponding with L1 utterance fluency instead. Finally, in light of research in this domain, it can be concluded that "L2 utterance fluency is dependent upon L2 cognitive fluency and has also been shown to be related to L1 utterance fluency" (Kahng, 2020, 476).

Taken together to synthesize findings from research in this field, the importance of utterance fluency is undeniable on account of both its theoretical and pedagogical significance (Kahng, 2020), and it has been found that the association between rate, breakdown, and repair measures of utterance fluency with both perceived fluency and cognitive fluency is notable. To continue on the same path, growing number of scholars examine the underlying cognitive processes responsible for fluent speech performance, and relate various aspects of performance to different stages of speech production. In the next section, the major models of speech processing are presented to better appreciate the processes responsible for smooth and fluent production of speech.

2.3. Speech production and second language fluency

In order to discuss fluency of speech in more depth and get to know the language processes and knowledge components involved in the production of fluent speech in L2 speakers, the best starting point can be a model of speech production. Several models have been posed capturing L1 and L2 speech production processes (Kormos, 2006a), the most famous and well-known of which is Levelt's "blueprint" (1989, updated and revised in 1999). This model has been originally devised for monolinguals, adapted later by De Bot (1992) to match the bilinguals speech production system. Many L2 models of speech processing draw on Levelts'(1999) hugely influential work as their starting point (De Bot, 1992; Kormos, 2006a; Segalowitz, 2010), while modifying some of its theoretical underpinnings. Following the same path, Kormos (2006a) proposed the most comprehensive model of bilingual speech processing, which guides the discussion of fluency in this section. Figure ***2.2*** presents Kormos' (2006a) bilingual model of speech production.

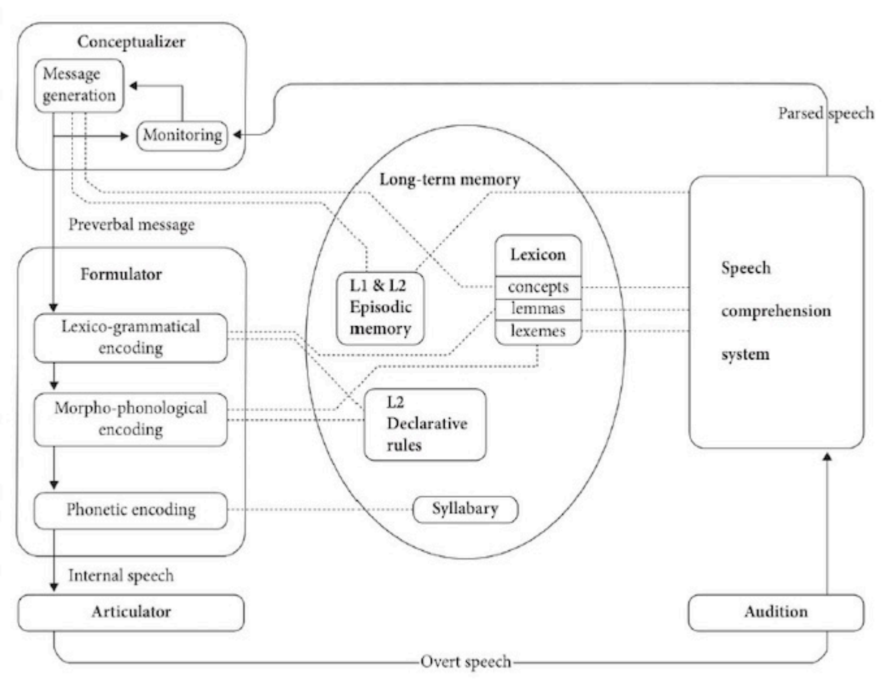


Figure 2.2 The model of bilingual speech production (Kormos, 2006a)

In general, regardless of the model adopted, there is a strong consensus among researchers in this realm that speech processing is composed of the four processing components, as originally put forward by levelt's model;

- Conceptualization, planning what to say;
- Formulation; encoding the syntactic, lexical, and morphological part of the message;
- Articulation; production for the actual speech sounds,
- Self-monitoring; checking the appropriateness and correctness of the output.

Although both L1 and L2 language processes involve the four above mentioned processing components, the way information is processed varies. This difference is a source of significant distinction in the speed of production of the two languages. L1 speech processing is largely automatic in the formulator and articulator processing components, and can, as a result, run in parallel, making the talk substantially smooth and fast, while this is not the case in L2 speech processing (Kormos, 2006a). As a result of degree of attention and control exerted over these two phases of encoding in L2 production, which will be discussed more thoroughly in alter section (see Section 2.3.5), the output is processed serially instead, and the speech is considerably slow.

In addition to Kormos' (2006a) model of bilingual language processing, in the realm of second language acquisition, a recent model concentrating particularly on fluency of speech has been introduced by Segalowitz (2010). This model not only accepts and incorporates the main tenets of Levelt's (1989, 1999) blueprint of the monolingual speaker as well as De Bot's (1992) implications of second language speakers, but also advances the discussion by calling attention to potential critical points in the speech processing system that underlie processing difficulties related to L2 fluency of speech. Figure ***2.3*** presents the bilingual model of speech production.

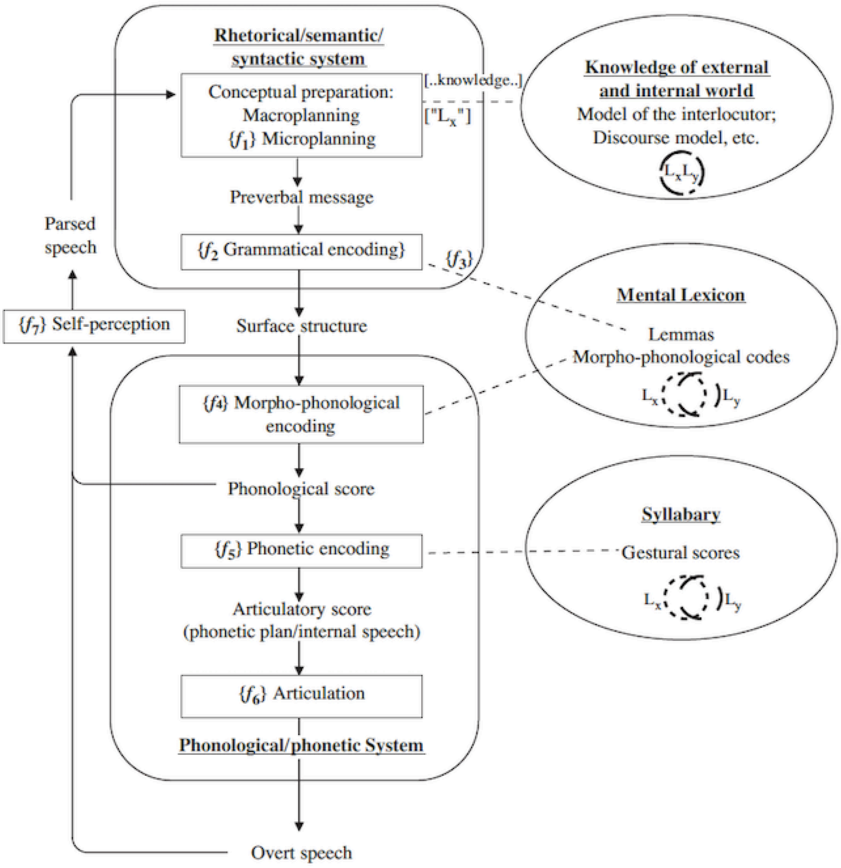


Figure 2.3 The L2 speech production process (Segalowitz, 2010)

In the following sections, each processing component is discussed in more detail while relating the process to L2 fluency of speech and fluency vulnerability points.

2.3.1. Conceptualization and fluency vulnerability points

The first phase in the process of speech production is the stage of conceptualization. This is the stage for conceptual preparation of the message, which involves deciding on what to say, that is planning the content of the message and activating concepts to be further processed in later stages (Levelt, 1989). Primarily, this stage maintains generating concepts drawing on knowledge of the discourse situation, the power relation between the speaker and the addressee, norms of the interaction and politeness, and finally encyclopedic knowledge of the world (Kormos, 2006a). Not only the concepts but also the language of communication is selected in conceptualization stage; accordingly, both L1 and L2 concepts, which are mainly shared and stored together in the semantic memory, compete for selection in this stage (Kormos, 2006a).

Breaking down the conceptualization stage, it is composed of two subcomponents of macroplanning and microplanning, as referred to by Levelt (1999, p. 89). Macroplanning includes elaboration of the communicative intention, selection and retrieval of the information required to express the communicative goals, and also deciding on the content of the speech act (Levelt, 1989). The second subcomponent, microplanning, deals with further shaping the speech acts, and generating language specific surface structure, while it is still conceptual in nature. It involves making decisions on the perspective, accessibility status, new and old information consideration, the order of information presentation, and providing propositional content to the message (Levelt, 1989, 1999; Kormos, 2011). Finally, the result of the conceptualization stage is the preverbal plan, which consists of information concerning word meaning (De Groot, 2000; Kormos, 2011). To each of the activated and then selected concepts a language cue is added, which has already been selected mainly on the basis of discoursal model in macroplanning and specifically in microplanning, as discussed (De Bot, 1992; Kormos, 2011).

The process of conceptualization is not much straightforward for bilinguals due to the choice of language added to the whole process (De Bot, 1992; Paradis, 2004), which can be a source of difficulty and origin of vulnerability to the fluency of speech. As already discussed in this section, macroplanning applies the encyclopedic knowledge of the world; hence, it is not differentiated by different languages (De Bot, 1992; Levelt, 1989, 1999). For example, the knowledge that *Spring* is one of the seasons in which vegetation begins to appear is not held in any particular language store, and the choice of retrieving *Spring* in English or بهار (*Bahar*) in Persian will be based on sociopragmatic aspects of the context, which would reflect the next stage of microplanning. Accordingly, Segalowitz (2010) assumes that "no L2-specific fluency issues arise at this level because the processes involved are presumed to be non-language specific" (p. 10). However, the decision that L2 speakers need to make regarding what to say in this stage, which is based on their level of language proficiency and their understanding of their oral proficiency, can nominate macroplanning as a candidate for L2 fluency vulnerability point in the process of language production (Kahng, 2014). The reason is that fluency is very vulnerable to changes in the topic and content of speech, and macroplanning stage is the phase where shifts in topic can occur, hence, this stage of preparation needs to be finalized in the early process of production in the first place to avoid fluency problems (Kahng, 2014; Robertson & Kirsner, 2000; Tavakoli & Skehan, 2005).

Next point worth discussing is disfluencies during microplanning stage, which are caused as a result of avoiding problematic lexical concepts and formulating the preverbal message in a way to be encodable in the next stages. Sometimes L2 speakers do not have enough lexical resources available for some concepts to draw on, or they can't decide on the lexis required to be retrieved in the next stage of speech processing (Segalowitz, 2010). For instance, the preverbal plan might include the lexical item *Spring*, and a Persian L2 speakers of English does not actually know the lexical item in English. Occasionally, the bilinguals are aware of this shortcoming; hence, they try to avoid such concepts from this early stage (Poulisse, 1997), and select different lexical concepts which can be further grammatically encoded in the formulation stage. Therefore, the strategic microplanning applied during conceptualization can lead to dysfluencies of speech.

2.3.2. Formulation and fluency vulnerability points

The preverbal message as the output of the conceptualizer, is the input for the second stage of language processing, formulation, changing it into internal speech. This stage entails matching all conceptual specifications of the preverbal message plus the language cue, with the suitable lexical entry by accessing the mental lexicon (Kormos, 2006a, 2011). Mental lexicon is a depository of all lemmas and lexemes in a form of a gigantic network of interrelated links and connections (See section 2.4 for a detailed discussion of bilingual mental lexicon). This processing component is a store of both L1 and L2 languages, and "there is no need to postulate that a speaker's repertoire of different language lemmas must be segregated in the brain in distinct, neural regions" (Segalowitz, 2010, p.13). Both languages are assumed to be located in "neurally related regions" of the brain (Paradis, 2004) just as synonyms, near synonyms, and related words do. Accordingly, not only links are created among items of the same language, but also between items of different languages.

Accordingly, the selected concept sends activation to lemmas in both L1 and L2 (which is the case of shared or partially shared concepts in L1 and L2), and they compete for selection. The lemma matching all or most of the conceptual specifications and the language cue receives the highest level of activation and is selected for further processing. A point worth mentioning is that not all L2 speakers can automatically access the lemmas during the creation of the surface structure, and they might occasionally face difficulties in retrieving them, which is considered "the most salient characteristics of a non-balanced bilingual" (De Bot, 1992, p. 14).

Next, the activated and selected lemma enters the stage of syntactic processing. Syntactic processing entails two processes of firstly activating the syntactic properties of the lemma, which is carried out using declarative knowledge stored in the lexicon, and secondly, structuring the phrases and clauses and putting them in an appropriate order, which can be done either using procedural or declarative knowledge depending on the level of language proficiency of the L2 speaker (Kormos, 2006a). It is important to note that in the case of unbalanced bilinguals, sometimes it is difficult to apply the required linguistic resources to construct the syntactic surface structure. Dysfluencies during syntactic encoding can arise from various factors such as drawing on declarative knowledge, which is applied consciously, or lack of knowledge of the syntax which results in conscious transfer from L1 (Segalowitz, 2010). Finally, the result of this stage is creation of the surface structure, which is the "linguistic shape" of the message (Segalowitz, 2010, p. 12).

2.3.3. Articulation and fluency vulnerability points

After selection of the appropriate syntactic form, the outcome is required to be converted into a form that can be distinguished as overt speech (Segalowitz, 2010). This requires the phonological shape to be retrieved from the mental lexicon, which in turn needs activation of the phonological form, division of the words into syllables, and setting loudness, pitch, and intonational duration parameters (Kormos, 2006a, 2011). Each lemma, having a specific phonological code (lexeme) in the mental lexicon, sends activation to both L1 and L2 counterparts in order to create the phonological score. Phonemes are activated by phonological forms in a serial fashion (Kormos, 2011); that is, activation begins from the first phoneme and ends on the last, one after the other. Levelt (1989) poses that the process of morpho-phonological encoding is automatic for L1, requiring no conscious attention, while the unbalanced bilinguals might rely on declarative knowledge in the case of different phonemes in L1 and L2, stored with different and separate representations, making the process dysfluent with hesitations (De Bot, 1992; Segalowitz, 2010).

In turn, the phonetic encoding phase activates the syllable programs of both L1 and L2, the gestural scores for converting the phonological score into articulatory score (Levelt, Roelofs, & Meyer, 1999), which are stored together in a common depository called the syllabary. In this phase, there is no access to declarative knowledge, and all the process is supposed to be carried out in an automatic manner, while the low proficiency L2 speakers and unbalanced bilinguals in the case of not having access to L2 syllable program rely on the L1 syllable program, which is the reason most L2 speakers have accent (De Bot, 1992). In addition, any effort in choosing the syllable program and executing it can be another source of dysfluency for L2 speakers as it again requires conscious attention (Segalowitz, 2010). The outcome, phonetic plan, is the input for the articulator, by which the intended speech plan is converted into actual speech.

2.3.4. Monitoring and fluency vulnerability points

All through the process of speech processing, there are instances that things might go wrong at any of the conceptualization, formulation, or articulation stages; hence, speakers monitor their speech to resolve problems. This is, when the "internal speech" (Levelt, 1983, p. 469) is listened to, comprehended and reformulated. As we are the first people who listens to our own speech, monitoring internal representations can happen while encoding (Levelt et al., 1999). The process is similar for both L1 and L2 in the sense that both requires conscious attention for the process of monitoring.

The whole process is accomplished in three loops being in charge on the inspection of the outcome; first of which is comparing the outcome of conceptualization stage, the preverbal massage, to the intended communicational goals. This type of monitoring and repair occurs as the speaker might need to change topic to a new or totally different one or might need to modify the syntax (content of the information currently formulating). Second type of repair is monitoring the phonetic plan before the phase of articulation, which is called covert monitoring. It can happen due to various factors such as words being activated erroneously, incorrect grammatical structure applied, or even due to selection of a wrong phoneme. The final loop is responsible for checking the final generated outcome after articulation, that is the utterance is being monitored externally (Levelt, 1989). In case any errors encountered, the production mechanism starts for the second time from conceptualization stage all over again (Kormos, 2006b).

A fluency point needs to be taken into consideration in here at the stage of monitoring. Monitoring by its nature interrupts fluidity of speech to go back, check for errors, and repair the problem, though L1 speakers do not interrupt their speech at the moment they find an error and prefer fluency over accuracy (Seyfeddinipur, Kita, & Indefrey, 2008). The case for L2 speakers is different, depending on their level of language proficiency and the circumstances under which one is speaking. They might stop to reformulate speech as soon as they detect an error as prioritize accuracy over fluency, which is equal to frequent pausing (Segalowitz, 2010).

Consequently, from the discussion above, it is evident that speech production processes in L1 and L2 have main points of similarities; that is, the basic underlying psycholinguistic mechanisms are very similar, though there exist points of differences mainly causing L2 speech dysfluent. As the definition of fluency indicates, fluent language is “fast” and “smooth”, which is mainly due to the parallel and automatic manner of processing of language discussed. As L2 process requires conscious attention and is not primarily automatic, various difficulties are encountered in speech production. Accordingly, the main difference between the processes involved in the two languages origins from the automaticity. The following section will follow what is meant by automaticity of speech processes while several theoretical views and routes to automaticity are discussed. Finally, theories of automaticity are applied to the process of lexical retrieval from memory, which is the main focus of the current research.

2.3.5. Achieving automatic control over speech production

Automaticity has been studied broadly within the realm of cognitive skill acquisition (e.g. Anderson & Lebiere, 1998; Anderson, 1983; Levelt, 1989, 1999; Logan, 1988) and more specifically and rigorously as respects to one aspect of skill acquisition in the field of second language acquisition literature (e.g. DeKeyser, 2001; DeKeyser, 2017; Hulstijn, 2001; Ellis, 2002; Schmidt, 2001; Segalowitz, 2003; Segalowitz & Hulstijn, 2005). But what can a practical definition of automaticity be?

Automaticity has been recognized as a multi-layered construct referring to efficacy of taking various steps in the underlying cognitive processes of speech production in the absence of attentional control, which can lead to fluency (Derwing, 2017). Automaticity refers to processing which is “unintentional, uncontrollable, unconscious, efficient, and fast” (DeKeyser, 2017, p.17). Such features of automaticity are in line with a very early definition by Schneider, Dumas, and Shiffrin (1984) defining it as “a fast, parallel, fairly effortless process that is not limited by short-term memory (STM) capacity, is not under direct subject control, and is responsible for the performance of well-developed skill behaviors” (p.1).

Although there is a relative consensus regarding the definition and main characteristics of automaticity, there exist different theories how automaticity emerges; that is, under what conditions a process becomes automatized. These theories have been mainly classified into rule-based approaches, which consider the conversion of declarative knowledge to procedural, and the item-based approaches, involving a single step access of memorized items (Derwing, 2017; Kormos, 2006a).

Prior to explaining the rule-based approach to automaticity, it is significant to know about different types of knowledge, as there is a general consensus that not all types that speakers have are the same. Declarative knowledge is knowledge “about the language”, while procedural knowledge is knowledge “of the language” (DeKeyser, 2017). Contrary to declarative knowledge, procedural knowledge is unconscious and is available to the speaker in real time communication situations. Although both declarative and procedural knowledge can be verbalized, declarative knowledge is consciously known to the speaker, while procedural type can be applied in spontaneous communication unconsciously.

Rule-based approach consider the “development of automaticity as the transformation of factual knowledge into production rules” (Kormos, 2006a, p. 40). That is, at the first stage of learning, some type of knowledge used in communication is the declarative type, resulting in dysfluencies as the speaker needs time to process it in spontaneous real time. Practice is the next step on the way to reach automaticity. DeKeyser (2017) points out that “as a result of practice, they become better at putting their knowledge to use, using it more correctly, more easily, more frequently, in a wider variety of contexts” (p. 16). Finally, the rules can be applied with no conscious attention automatically. Kormos (2006a) illustrates the way by taking an example of using articles “a” and “an” before singular nouns. At first, the speaker has to pay conscious attention to the way the articles are used; that is, using the declarative knowledge of applying “a” before nouns with a consonant and “an” before nouns starting with a vowel sound. Gradually, by putting this knowledge into use more frequently and as a result of practice, the theory claims that it might be converted into “production rule”. That is, the rule might be applied automatically, unconsciously, and efficiently.

One of the most comprehensive accounts in which the application of rule-based approach is provided is by Anderson's adaptive control of thought (1983), ACT theory and his revised (1995) ACT-R theory. This theory proposes that developments in automaticity is not only a quantitative change, involving the speeding up of the processes, but also qualitative modifications and reorganization of the procedures already acquired. Anderson (1995) argued that there are five learning mechanisms playing a major role in the development of automaticity: composition, proceduralization, generalization, discrimination, and strengthening.

The first mechanism, composition, refers to breaking down the sequences and creating “macro-productions” or chunks. Proceduralization mechanism refers to the retrieval of a whole chunk from memory while no declarative knowledge needs to be activated for execution. As the rule has become proceduarlized, the declarative knowledge might not be anymore remembered, and the application can become automatic. Three mechanisms contributing to the fine tuning of procedural knowledge, leading to qualitative changes, are generalization, discrimination, and strengthening. These three processes involve being able to apply new production rules wherever they are suitable, and *only* when they are suitable, and that better (more efficient) rules are selected, and weaker ones are abandoned. The first of these three, generalization, refers to broadening the scope of application of the declarative rules, while discrimination narrows down the scope of the rule application. Consequently, these more general rules are strengthened through use. Kormos (2006a) explains it is “the increased likelihood with which a particular production procedure is selected” (p.41). Accordingly, Anderson’s model provides a detailed and comprehensive theory of rule-based approach to automaticity. Hulstijn (1990) poses that “the more sub-procedures get subsumed into overall procedures, the more language use can be said to take place fluently and automatically, requiring less attention” (p.32).

Although conversion of declarative to procedural knowledge is one of the plausible ways to automaticity, it is not the only way. The aforementioned stages in language learning, akin to cognitive skills development at large, poses a strong-interface position where finally the procedural knowledge becomes fluent, spontaneous, and effortless (see e.g., Anderson, 1982; DeKeyser, 2007). Unlike the strong-interface position, which has a single incarnation, the proponents of the other interface position, the weak-interface position, assign different weights to the role of consciousness. R. Ellis (2005; 2006) proposes that in the early stages of learning in instructed environment of the classroom, explicit knowledge can turn into implicit when the learners are developmentally ready and are at the correct developmental stage for that feature. Unlike this position, ascribing a much lesser role to consciousness, N. Ellis (2005; 2007) maintains that learning is largely implicit, assuming to be occurred based on extensive meaning-focused exposure to the target language and its frequent usage. It goes on explaining that the unconscious learning processes, which occur automatically during language usage, are necessary in developing the rationality of fluency (N. Ellis, 2005).

Next, theories of item-based approaches are another route to automaticity, considering automatic processing to be equal to memory retrieval in a single step. Logan's (1988) instance theory was one of the first models arguing that not all learning involves the conversion from declarative to procedural learning. The proponents of this theory claim that another possible way to automaticity is memory retrieval, occurring as a result of practice of strengthening the associations between the problem and the memory traces of the solution, leading to speeding up of the retrieval. The scope of this model is rather narrow since triggering a memory-based retrieval process happens only if a stimuli encountered is totally identical to the one encoded in memory, otherwise in case the stimuli is only similar and not identical, same results cannot be maintained (DeKeyser, 2001). In support of this claim, Myles, Hooper and Mitchel (1998) examined the data from a two-year longitudinal study of 16 beginner learners of French for occurrences of chunks, exploring their contribution to the development of creative language. The results revealed that school-aged children in this study, used parts of the gradually unpacked chunks in generating new utterances. This is in-line with the findings of Robinson and Ha (1993) demonstrating that "the two processes are not either/or options, but there seems to be an interface between them and both might contribute to the development of automaticity" (Kormos, 2006, p. 159).

Besides Logan's (1988) instance theory, strength theory (e.g. MacKay, 1982) also attributes development of automaticity to memory. In such models, memory retrieval is based on the notion of strengthening (Cohen, Dunbar, & McClelland, 1990; MacKay, 1982; Schneider, 1985; Schneider & Detweiler, 1988). The main difference between strength theories and instance theory is that there is no accumulation of response representations, but instead there is a continuous strengthening of the connections among responses and stimuli. It assumes that the stimuli and responses develop strong connections given enough practice. Mackay (1982) argues that automatic behavior can become more flexible with practice, transferring readily from one response mechanism to another, which takes place in hierarchical order from the lower level to the automaticity and fluency of the higher-level tasks.

Consequently, as discussed in this section, the role of automatic processes has been broadly investigated over the past few years, and it appears that the major point of commonality among all of the aforementioned theories is the significant role practice plays in attaining automaticity, though each of such models might be applicable and more appropriate to certain areas of performance. Looking back into these theories, strength theory of automatization seems to bear more relevance to the development of cognitive fluency in the process of lexical encoding and lexical retrieval in L2 production. In this theory of automatization, quick and efficient retrieval of the lexical items comes across with the creation of strong connections among concepts and lexical items and the search procedure is taken over via a direct single step memory retrieval. To summarize, “the major process of automatization in lexical encoding involves the strengthening of links between concepts and L2 lexical items” (Kormos, 2006a, p. 160).

Summary

Linguistically encoding one’s intentions can be argued to be one of the most difficult and complex activities to be carried out in automatic manner without conscious attention, and this is more critical in case of second language (Kormos, 2006a). In the above section, bilingual speech processing mechanism was addressed thoroughly along with various points where difficulties in processing might bring about problems in fluency of speech. The L2 fluency vulnerability points proposed by Segalowitz (2010), mainly discuss the critical aspects underlying the cognitive processes triggering dysfluencies, that is the ones associated with problems while conceptualizing, formulating, articulating the message, and even monitoring the outcome, though the knowledge stores, which “do not share processing functions” (Kormos, 2006a, p. 7) and their impact on reaching automatic control over the process of speech production and fluency, was never examined. Accordingly, the notion of storage and the knowledge store components which are critical in the study of L2 spoken language fluency (Singleton, 2009), have been missed out.

Mental lexicon, as one of the most influential knowledge components, is a store for various pieces of lexical knowledge, while lexical knowledge is regarded as one of the greatest barriers in the production of L2 fluent language (Maclay &Osgood, 1959; Hilton, 2008; Kahng, 2014). For a fluent production of L2 spoken language based on theories of automaticity, it appears crucial to take various properties of the lexicon into consideration, since it is postulated that maintaining an efficient use of language in real communication situations requires bilingual mental lexicon to be of certain size and quality for optimal access during online encoding of the lexical items (Henriksen, 2008; Hilton, 2008). Despite the fact that lots of research has been conducted in the realm of fluency and factors having an impact on its improvement, what hasn’t been explored is the impact of the quality of the network structure of the bilingual mental lexicon; that is, its structure and organization, on L2 speech fluency.

The current study aims to weave the lines of fluent bilingual speech production and the structure of the bilingual mental lexicon. Together, this might be explained with the help of strength theory of automaticity, focusing on strengthening the connections within the hierarchical system of language. This can maintain smoother and faster production of speech, that is more fluent outcome.

Next section mainly concentrates on a detailed description of the organization of the bilingual lexicon and the quality of the connections among items stored within and between hierarchical levels of knowledge.

2.4. Bilingual mental lexicon

Without grammar very little can be conveyed
Without vocabulary nothing can be conveyed
Wilkins (1972, p.111-112)

Mental lexicon is one of the most fundamental components of language processing, storing various pieces of word knowledge. The term was first introduced and developed by Oldfield (1966) as a passive data structure, and since then it has been frequently researched and redefined from different perspectives. Fay and Culter (1977) used the *dictionary metaphor* to describe the lexicon, focusing on the analogy of matching sound representations with their meanings. This early definition mainly concentrates on what knowledge is stored in the lexicon and does not go beyond to various ways lexical terms are accessed and retrieved; hence, it can be rather a simplistic perspective to “the cognitive system that constitutes the capacity for conscious and unconscious lexical activity” (Jarema & Libbe, 2007, p. 3). Although providing a concise definition is a challenge (Jarema & Libbe, 2007), what cannot be underestimated about lexicon is the pivotal role it plays in automatic speech production.

Large body of research has been carried out on central properties of mental lexicon, which are also regarded as dimensions of lexical competence or vocabulary knowledge from a lexical research perspective (Qian, 1998, 1999; Read, 1988, 1989; Wesche & Paribakht, 1996). The two primary ones are breadth (the number of lexical items stored in the lexicon) and depth (the quality of knowledge). Breadth is the number of known connections between form-meaning; that is, the number of links between lexical concepts and the lexical items in the lexicon, while depth inquiries about the properties of the lexical items, how well they are known and have been integrated in the lexicon with other items. Based on such definition, depth of knowledge can be more specifically conceptualized from two different perspectives: word-centered and lexicon-based. Whilst the former defines depth as “the quality of understanding a word” (Anderson & Freebody, 1981), the latter focuses on the network knowledge; that is the number and strength of the connections among the lexical items stored in the lexicon (for an overview, see Henriksen, 1999; Read, 2004), which requires restructuring of the network for any development (Meara, 1996; Dóczi & Kormos, 2016).

The current study concentrates on the lexical network perspective, as Meara (1996) argues that research in the depth of knowledge requires to concentrate on a deeper understanding of the lexicon rather than focusing too hard on finer details of lexical knowledge and individual lexical items. “Vocabularies are interlocking networks” (Meara, 2004, p.137), stored in mental lexicon, and exploring them separately is not in line with the nature of the lexicon. Accordingly, from a lexical network perspective, depth is conceptualized as “a lexical network in L2 learners’ mental lexicon”, aiming to investigate the extent to which lexical items are linked to one another in the lexical network (Henriksen, 1999; Meara, 1996, 2009; Meara & Wolter, 2004). This conceptualization corresponds best to Henriksen (1999) and Read’s (2004) definition of depth, inquiring in the context of network knowledge and network building. Taken together, it mainly refers to the degree of structuredness of mental lexicon; that is, the extent to which lexical items are linked to one another in an organized manner (Read, 2004), which can best be defined as “the process of discovering the sense relations or intentional links between words – that is, fitting the words together in semantic networks” (Henriksen, 1999. p. 308).

It is widely agreed that lexical items are learned incrementally, and not in a dichotomous fashion of either acquired or nor acquired (Dóczi & Kormos, 2016; Meara, 1983; Schmitt, 1998b), and this assumption calls for a sort of organization, in which “newly acquired words need to be accommodated within a network of already known words, and some restructuring of the network may be needed as a result” (Read, 2004, p.19). This approach mainly concentrates on development of connections among lexical items in the network of mental lexicon. Aitchison (2012) argues in favor of the logical structure of “gigantic multi-dimensional co-web” (p.70), providing two reasons. Firstly, the speed with which tens to thousands of lexical items are searched, found, and retrieved implies a structured organization. In addition, the “flexible and extendable” nature of the memory calls for organized and structured type of information (Aitcheson, 2012, p. 5). In case of random information, it is extremely difficult to remember them; though, even very large quantities of information can be remembered quickly, provided that it is structured and organized. Accordingly, it can be argued that lexical items are stored in a dynamic ever-changing network of lexicon in a structured and logical way, which makes it possible to recall and produce tens to thousands of words with a high speed, providing quick access to all items during language use. Aitchison (2012) asserts that “enormous quantities of data can be remembered and utilized as long as they are well-organized” (p.5). L1 mental lexicon is a representation of a well-organized web with many links and pathways among lexical items, enabling fast and easy retrieval of the words due to the fact that “the access routes in the lexical store are varied and well-established" (Henriksen, 2008, p. 20). The many lexical items in L2 networks are not stored as structured, though it is still organized in a logical way (Meara, 2009).

In light of this argumentation, the organization of the bilingual mental lexicon, with a focus on the network properties, is explored in this section. That is, different levels of knowledge representation in mental lexicon are introduced, which is followed by discussion of the models of lexical development and representation in the bilingual mind. Next, models of flow of information within and between these levels in the network is presented. Finally, the technique to investigate the lexical network, word association, is introduced in brief, which will be discussed in more depth in chapter 4.

2.4.1. Levels of knowledge representation in the bilingual mental lexicon

Whether hierarchical or network, it can be assumed that the bilingual mental lexicon comprises two levels of knowledge representation: the conceptual level and the lexical level. As the mental lexicon is a gigantic network of links and associations, it is assumed that various associative links are established among items in each of these levels as well as links across them. Prior to discussing the nature of the links, it is essential to scrutinize the underlying assumptions regarding the structure of knowledge representation adopted in each of these levels. Accordingly, this section mainly concentrates on the nature and type of knowledge stored in each level of knowledge representation, in addition to the relation among the items in each of the levels across different languages.

2.4.1.1. Conceptual level of knowledge representation

As already discussed (see Section 2.3.1), the first level, conceptual level of knowledge representation, contains experiential knowledge, being mostly derived from experience and knowledge of the world. Concepts are "multimodal mental representations that include visual (mental imagery), auditory (sound), perceptual (texture) and kinesthetic (sensory-motor) information stored in implicit memory" (Pavlenko, 2009, p.132). Hence, they are considered to be mainly language neutral, due to representing shared background of encyclopedic knowledge (Dufour & Kroll, 1995). Among concepts can exist links which are created as a result of "cooccurrence or contiguity of objects and actions in the world as well as through shared or private experiences of events" (Henriksen, 2008, p. 30).

Over the past few years, considerable literature has been emerged on bilingual models of language processing, with the exception of Distributed Feature Model (De Groot, 1992), bearing evidence that concepts are largely cross-linguistically shared between languages (cf. Costa, 2005; Kroll & Stewart, 1994). If this is the case, Pavlenko (2009) argues "linguistic categories mediated by languages A and B share both category structure and boundaries" (p. 33). Pavlenko (2008) made an example of equivalent/near equivalent concepts by describing the feeling of fear in English and Russian. L1 Russian speaker, talking about a person experiencing fear, prefers to apply reflexive emotion verbs (ispugat'sia, boiat'sia), while L1 English speaker favors emotion adjectives or pseudo-participles (afraid, frightened, terrified). Thus, it was mentioned that although English and Russian speakers apply alternatives which are structurally different, they are still conceptually same and shared. However, not all lexical concepts are fully shared between languages. Instead, there exist many concepts which are partially equivalent, such as abstract nouns. The verb *to fall* in English roughly corresponds to two Finnish words. The first Finnish translation equivalent is *pudota,* meaning to fall down/off from a high to a low altitude, like when you drop down a phone, and the second is *kaatua,* which is used in the sense of falling from a vertical to a horizontal position, like when a tree falls (Jarvis, 2003). In such cases, speakers need to make fine-grained differentiations and take the contrasts into account while encoding the concept. Finally, the last group are the linguistic categories in language A which do not have any counterpart in the other language, and such concepts are culturally specific or might be completely idiosyncratic. In this case, Pavlenko (2003) illustrate the two terms *privacy* and *personal space* in English, which do not have any corresponding conceptual and translation equivalent in Russian language. In this case of conceptual nonequivalents, L2 learners can use the lexical item, such as *privacy,* if they acquired the nonlinguistic concept (such as imagery or scripts) through interactions in natural setting which can be supplemented by classroom learning.

Second issue in the structure of the conceptual level, which needs to be discussed and taken into account, is the type of information stored in it, as it is a depository for various types of information. There is a strong consensus among researchers that pragmatic and sociolinguistic knowledge, such as apologizing, requesting, etc., are stored at this level. Research in L2 pragmatic implies that this type of information is difficult for L2 learners to acquire; hence, even the higher proficiency L2 speakers most often rely on L1 pragmatic and sociolinguistic knowledge, and this case is more prominent in the decontextualized classroom setting (Kasper, 1992).

As regards to other types of knowledge, there is still disagreement if the semantic specifications are stored along with the conceptual information at this same level or if the semantic and conceptual representations are required to be distinguished and stored at different levels. There are two different views with respect to it: the one-level view and two-level view. The advocates of latter argue that semantic and conceptual features of the lexical items are distinct entities, and are stored at different levels of representation (Paradis, 1997, 2000; Pavlenko, 2009). Pavlenko (2000), as one of the proponents of the two-level approach, indicates that "conceptual representations should be treated as related but not equivalent to word meanings" (p.3). Paradis (1997) maintains that the semantic component of the lexical items is stored outside of the lexicon as declarative form of knowledge. The necessity of having separate levels comes from two sources of research; one of aphasic patients who cannot name objects while they can characterize them (Paradis, 2000) and the second from studies by Pavlenko (2009) on cultural relatively perspective of bilinguals. On the other hand, the proponents of the former propose that there is a single level where conceptual and semantic features are stored together (Levelt at al., 1999; De Groot, 1992). The model of lexical access proposed by Roelofs' (1992) and Levelt et al.'s (1999), the WEAVER++ model, serves as an example of one-level view, proposing that conceptual representations can also code semantics (for a detailed discussion of this issue, see Roelofs, 2000). In addition, Roelofs (2000) challenges the assumption of separation of the two levels based on aphasic patients failing to access the lexical representations while they can retrieve the conceptual specifications. For instance, an aphasic patient having difficulty naming a dog, while is able to identify a dog perceptually and infer that a dog can bark, can be diagnosed with having problems with concept-lemma connection rather than the inability of conceptual identification and the process of conceptual retrieval. Consequently, as the relationship between the semantic and conceptual terms seem to be ambiguous and difficult to discern in empirical way (Tytus, 2013); for the purpose of the current study, the two terms are decided to be used interchangeably (Francis, 2005, 1999; Levelt, 1999).

Finally, it is noteworthy to mention that the number of concepts, both linguistic and nonlinguistic, stored in the conceptual level is higher than the number of the lexical items due to the fact that many of the concepts do not have any linguistic representation to be encoded, as discussed above. The focus in this study is mainly on *lexical concepts,* which are the linguistic categories seen as "conglomerate of interrelated memory traces consisting of information concerning word meaning" (Kormos, 2006a, p.169). Finally, following Levelt (1999), the term *lexical concepts* is applied in this study, covering both types of conceptual and semantic representations, which can be linguistically encoded. Having defined the structure of the conceptual level of knowledge representation, the discussion now focuses on the second level of knowledge, the lexical level.

2.4.1.2. Lexical level of knowledge representation

The second level of knowledge representation, lexical level, is assigned to the lexical items. This level is a depository of the lexical entries of a language, which are "dynamic, episodic, and therefore inherently unstable" (De Bot & Lowie, 2010, p. 117). As discussed earlier (see Section 2.3), this level is composed of syntactic, morphological, idiomatic information in the lemma level, and phonological knowledge in lexeme level, which are both very much distinct from lexical concepts (Pavlenko, 2009), as they are mostly linguistic representation of the knowledge.

In an overview of bilingual models of language processing and representation, there is a general consensus that meaning and concepts at the conceptual level are mostly shared between languages, while empirical evidence has shown that lexical level of knowledge representations are stored separately (Pavlenko, 2009), though not entirely apart as they are stored in close related manner to one another across the two language (Zeng, Branigan, Pickering, 2020). Singleton (2007) argues that "languages differ widely in formal terms" (p. 5); that is, they have very different morphological and phonological systems at the lexical level, and the searches bilinguals do in finding analogies referring to the structure of the new language provides evidence for the separation at this level. The degree of overlap at this level mainly depends on the relative closeness of the languages. In particular, cognates, (i.e., orthographically/phonologically similar lexical items, having similar meanings in the two languages) are examples of lexical items which are considered to be sharing representations at the conceptual level, and to a certain degree in lexical level, as much of syntactic and morphological information is still language specific.

In accordance with bilingual models of lexicon and empirical data they are based on, it can be argued that although lexical levels of different languages are not totally shared, lexical networks can be intertwined (for a review, see Kroll and Tokowitz, 2009). Thus, the links created are not only among items of one language (i.e. lemmas, lexemes, and the concepts as discussed in 2.3.1.1) but also there exist connections across items of different languages at different levels of knowledge representation (i.e. lemma in language A can be associated with a lexeme in language B) (e.g. De Bot, 2004; Francis, 2005; Jiang, 2000; 2002; 2004).

In addition to cross-linguistic lexical associations in this level, a point worth noting is varying degrees of strength and asymmetricity of the connections within and across languages among items, which assign the lexical items to a central or peripheral position in the network of the connections. The strength of the connections can be affected by factors associated with the lexical items such as frequency of use, recency of use, degree of similarity between the items (Michael & Gollan, 2005, Pavlenko, 2009) or associated with the level of language proficiency and stage of learning in the learning process (in the early stages the links are mostly formed between L1 and L2 items at the lexical level while later links among L2 items can become strengthen) (Kroll & Stewart, 1990, 1994), which can in turn affect the asymmetricity of the connections. That is the direction of the links pointing between L1 and L2 can to an extent be determined by the level of language proficiency (Kormos, 2006a, 2011), when most often L2 items primes the lexical items of the corresponding L1 item more strongly than the reverse (Dóczi & Kormos, 2016).

Knowing specifically the structure of conceptual and lexical levels of knowledge representations, and the underlying issues considering whether concepts and lexical entries in these levels are integrated or separated in the bilingual mental lexicon, it is essential to find out how this can affect lexical development and processing in bilinguals.

2.4.2. Model of adult L2 lexical development

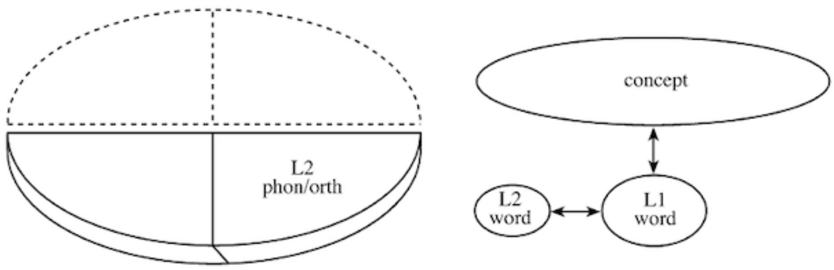
As mentioned earlier, various pieces of information of lexical entries (semantic, syntactic, phonological, morphological, etc.) are represented across two partially independent levels of conceptual and lexical knowledge representation in a highly integrated lexicon for bilingual (e.g., Levelt et al., 1999). Nevertheless, degree of integration is not same across the two levels, with higher degrees of overlap in conceptual level in comparison to the lexical level (e.g., Costa, 2005; Kroll & Stewart, 1994; Pavlenko, 2009). These levels of representation have high inter-level and intra-level connections, making the activation and use of lexical items feasible in the process of language production. Consequently, the full integration of various pieces of information and knowledge in these levels of representation leads to storage of readily available and accessible lexical items in natural communication.

There has been increasing number of studies concerning the development of L2 lexical knowledge in bilinguals (e.g., Hilton, 2008; Jiang, 2000; 2002, 2004; Laufer & Hulstijn, 2001; Meara & Fitzpatrick, 2000). Among them all, Jiang (2000) proposed a model of L2 lexical development for adult bilinguals, providing a conceptual framework, which particularly applies to the classroom second language learners. This model has its basis on the internal structure of the lexical entries and levels of knowledge representation and takes the practical constraints of L2 instructional setting into account. That is, this model considers the fundamental differences between the process of L1 and L2 lexical development, which is grounded on "the lack of contextualized input and the presence of an existing conceptual and L1 system" (Jiang, 2004, p. 417). On the basis of this model, lexical development in adult bilinguals can be divided into three stages of formal stage (lexical association stage), L1 mediated stage (hybrid stage), and full integration stage.

The initial stage in the process of L2 development occurs in a unique condition, which is under the influence of a strong tendency to rely on L1 lexical and conceptual systems in addition to the limited contextual exposure and scarcity of input in traditional L2 instructional setting. In this setting, new lexical items are mainly acquired through association with corresponding L1 translation equivalent or by means of definition. This in turn leads to primarily remembering the spelling and pronunciation of the words, while other pieces of knowledge of the lexical items are minimally created and established within the lexical entry.

Accordingly, the lexical entry in this level is composed of formal specifications (phonological/orthographic) along with a pointer to its L1 translation counterpart in the lexical level of knowledge representation. From the representational perspective, the lemma structure of such lexical items are empty and lexical entries are 'registered in mental lexicon' for the first time as items only containing formal specifications without lemma (Jiang, 2004, p. 417). However, this does not imply that the lexical items at this level of lexical development do not have any semantic or syntactic information available to them. Little semantic and syntactic information are available to such items through the existing L1 conceptual and linguistic structure; that is, through activating L1 translation equivalent of the lexical items. In addition, the syntactic properties can also be accessed through the explicit grammatical rules acquired in the classroom setting. Nevertheless, none of these pieces of information is integrated in the lexicon but are located out of lexicon as part of episodic memory, leading to lack of automaticity in accessing these pieces of knowledge and as a result, the lexical entry itself.

From a processing perspective, the outcome of conceptualization stage, the preverbal message, initially activates the L1 lexical item whose semantic specifications match the preverbal message. Subsequently, the L1 lexical item activates the associated and corresponding L2 lexical items through the lexical link between them in the lexical level of knowledge representation (Figure ***2.4***). This is also postulated in the word association model of bilingual lexical representation, posed by Potter, So, Von Eckardt, and Feldman's (1984), which was suggested to be more applicable to low proficiency language learners, as their reaction time had proven to be faster to L2 translation equivalents in comparison of the concepts and images of the words presented to them (French & Jacquet, 2004).



(a) (b)

Figure 2.4 Lexical representation (a) and processing (b) at the initial stage of lexical development (Jiang, 2000)

Subsequently, with more frequent use of L2 lexical item, continued exposure to L2 input in tandem with productive use of the lexical item, stronger associations between the L2 lexical item and its L1 translation equivalent are established. This implies coactivation of the L2 word along with the corresponding L1 lemma structure only, since the lexeme specifications of L1 is deactivated due to continued lack of activation in the process of production. Finally, the L1 lemma structure is copied and attached into the empty space of L2 lexical entry. The new item is a mixture of L1 lemma along with L2 formal specification. This stage is called L1 lemma mediation stage from a processing perspective and a hybrid entry stage from, representational viewpoint. The new representation of the L2 lexical entry may be used "with more fluency and automaticity" due to the creation of a direct link between the L2 item and the conceptual level of knowledge representation. However, the connection in the L1 mediated stage is still active and concept can be accessed through L1 translation equivalent as well. Therefore, it can be argued that there exist two different ways from the lexical level to the conceptual level of knowledge representation. The L2 lexical item can be connected to the conceptual level either through the L1 translation equivalent or through the L1 lemma structure within the lexical entity (Figure ***2.5***).

Jiang (2000) argues that as the link between the L1 lemma in L2 lexical entry is more straightforward and direct, it becomes the default one gradually.

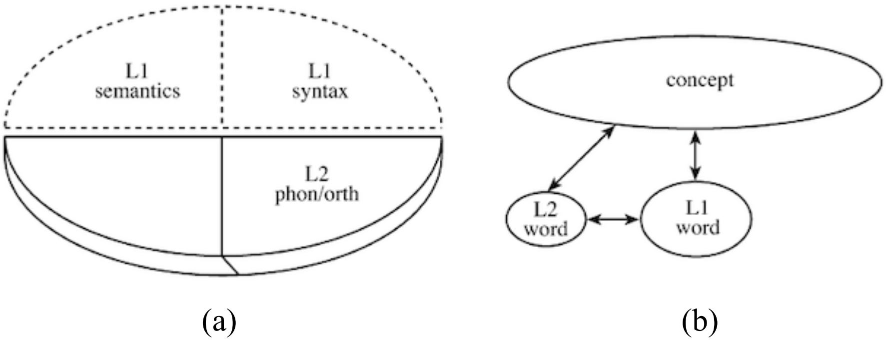
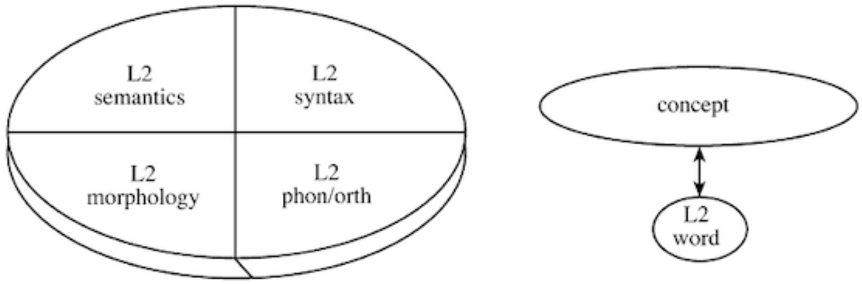


Figure 2.5 Lexical representation (a) and processing (b) at the hybrid entry stage of lexical development (Jiang, 2000)

Finally, the last stage of L2 lexical developments in adult bilinguals is the full integration of the L2 lexical entry into the mental lexicon as a single entity and discarding L1 specifications (Figure ***2.6***). This stage corresponds to the highest level of automaticity and idiomaticity (Jiang, 2004), as the influence of L1 linguistic and conceptual knowledge representations are minimum. This is a stage in which no difference can be spotted between the L1 and L2 lexical items as both are fully integrated into the lexicon. Nevertheless, Jiang (2000) maintained that not many words may enter this last stage and most often lexical items stop on either the first or second stages of development.



(a) (b)

Figure 2.6 Lexical representation (a) and processing (b) at the integration stage of lexical development (Jiang, 2000)

In general, this model is concerned with the way lexical items evolve in the process of learning and development. It doesn't provide a general picture for how bilingual mental lexicon develops over time in L2 traditional classroom setting as a whole; nevertheless, these concepts are closely intertwined. Accordingly, probing into the links and connections created between and within conceptual and lexical levels of knowledge representation, it is essential to find out how all this information is structured and organized in this gigantic network, allowing words to be looked for, found, and retrieved with such an impossible speed, making speech fluent and communication flowing smoothly.

2.4.3. The relationship between the items in the lexicon – focus on network knowledge

By the introduction of the theory of semantic memory in 1960s (e.g., Collins & Quillian, 1969), the idea of integrated mental lexicon in bilinguals has received great support, and also the ways through which information flows in such integrated network received considerable attention. As mental lexicon is a gigantic network of connections, it is viewed as "a web of interconnected nodes that collectively produce activations of lexical items" (Neff, 1991, p.199); that is, there exist various degrees of relationship within and between conceptual and lexical levels of knowledge representation already discussed (see Section 2.4.1), which has implications for parallel and automatic language processing. The flow of information was proposed to be accomplished through a model of spreading activation, being originally proposed by Collins and Loftus (1975), which is based on the connectionist's models. The model originally describes the flow of information in semantic networks in which semantic memory traces, being represented as nodes, can organize large network of concepts, and the connections among different nodes are determined by the degree of strength between the links, rather than hierarchy of the connections (Kersten, 2010).

The principal element of this approach is explained by Singleton (1999) posing that, "in the language processing a multiplicity of nodes are excited by the arousal of a node by which they are connected" (p. 125). Subject to activation of any given node via "stimulus presentation or via internal direction of attention" (Balota, 1994, p. 342), the activation spreads through associative links to the related and connected nodes comprising the network. The nodes are interlinked through unidirectional and bidirectional associative links within semantic networks, which are weaker in case of connections among nodes in different conceptual networks in comparison to the related networks. A sample subnetwork is presented in Figure ***2.7***, showing the notion of relatedness of the concept nodes.

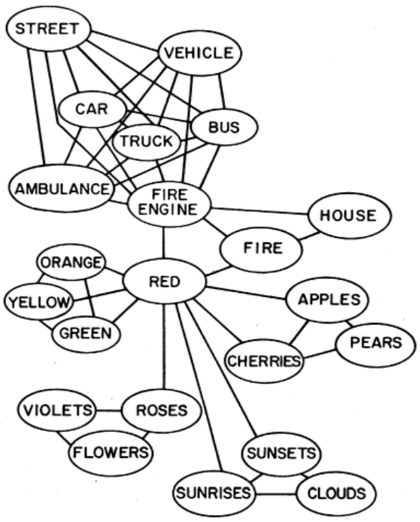


Figure 2.7 Sample of spreading activation model (Collins & Loftus, 1975)

Collins and Loftus (1975) illustrate the notion of concept relatedness in the figure above by exemplifying the concept of *vehicles,* with various items and connections of various length among them (the strength of the concept nodes is shown by the length of the edge linking the nodes; that is, longer edges represent weaker while shorter ones indicate stronger relationships). In this case with activation of the concept *vehicle,* activation will be spread to all types of vehicles. In other words, if the concept *ambulance* is primed, it will then activate *bus, fire engine, car, truck,* etc., and then each of them will in turn activate others related. Yet, if the concept *red* is primed, the activation spreading to *fire engine* will not activate *flowers* or *pears*, as there are not many links between these concept nodes. Alternatively, *fire engine* will tend to prime other vehicles and *roses* to prime other flowers. As a result, a similar amount of activation with a diluted strength will be spread out among a greater number of concepts. Likewise, this entails the cases for bilinguals when separate concepts exist for an identical notion in L1 and L2 (as discussed in section 2.4.1). If so, both concepts will be activated; although, based on the bilingual speech production models (see Section 2.3.1), only the intended concept in the selected language will be chosen for further processing (e.g. Bloem, van den Boogaard, & La Heij, 2004; Kormos, 2006a; Levelt, 1989).

Accordingly, the strength of the links between different semantic representations varies on the basis of the properties they share; that is, those having close and many mutual semantic features have stronger connections in comparison to those sharing minimal or no shared representations (Hashimoto & Frome, 2011). It is noteworthy to mention that although within a same semantic network, the links are fairly strong, the strength of the connections is not stable and can be partially determined by the frequency of production of the associative links, which has been best described as "fire together, wire together" (Shatzz, 1992, p.26). Finally, stronger connections among concept nodes can spread activation to a network faster in comparison to a network which does not have as strong associative links among its nodes. The spread of activation will then gradually diminish over time as the number of pathways increases (Collins & Loftus, 1975).

The main point of criticism to Collins and Loftus's (1975) model was the absence of any clarification on separate and distinct levels of lexical organization. Accordingly, a revised version of spreading activation model adding a language level to the previous semantic model was proposed (Bock & Levelt, 1994). The extended model takes different levels of lexical organization into account as well as the conceptual level. Accordingly, there are two main levels at which it operates; the conceptual level and lexical level, which is in turn composed of lemmas (intermediate level), lexemes (language level). Conceptual level of knowledge works as a semantic network described with concept nodes linked. As explained earlier (see Section 2.4.1.1), semantic relationships among lexicalized concepts also exist in the conceptual realm. The lemma level, the "the hidden abstract level of knowledge" (Randall, 2007, p. 115), includes syntactic and morphological information, and finally the lexeme level is the actual language level, providing the phonological information required for production of language.

The revised spreading activation model, like Collins and Loftus's model, put great emphasis on the strength of connections while integrating linguistic features, which is the point of divergence between the two models, as the first model works independently of the language, while Levelt and Bock's (1994) model integrates different processes at interactive levels of knowledge. It is assumed that information spreads through links which are established within each of these two levels of knowledge as well as between them. In case of the bilinguals, activation spreads from the selected concept to both L1 and L2 lexical items; that is more specifically, it is spread to both L1 and L2 lemmas (Hermans et al., 1998; Lee &Williams, 2001), which compete for selection. So, in this case according to theories of bilingual language production and spreading activation, not only the lemma and the semantically related lemmas from the intended language but also lemma from the non-intended language along with the semantically related ones all receive various degrees of activation, which in turn spread activation to the lexemes in the same manner (Kormos, 2006a). Consequently, the various levels of knowledge representation intersect in order to provide a representation of the required knowledge in successful language production.

Items in this integrated network are connected to one another as a result of "well-worn paths among words that tend to be used together or because of shared information (phonemic, syntactic) among the items" (Murphy, 2003, p. 238). In particular, the lexical entry *sun* might have associative links to other lexical items in the semantic memory; for example, to the words *moon*, *sky*, *day* and *sunny*, which are developed and strengthened via repeated exposure to the contexts in which such items co-occur. The type of connections in lexical level is qualitatively different from the links in the conceptual level. For instance, *black* and *white* may be linked in the lexical level, but they are only recognized as antonyms at the conceptual level (Murphy, 2003).

Consequently, by examining the associations, existing between conceptual and lexical items both within and between levels, that is measuring the network knowledge, it is possible to explore the structural properties of the lexicon, which is one of the aims of this study.

2.4.4. Investigating and measuring network knowledge - word association

As already discussed (see Section 2.4.1), lexical network is defined as "intricate, multi-layered structure or word web with numerous levels of interconnections between lexical items" (Henriksen, 2008, p.28), and all this network of lexis is built upon and connected to the semantic network at the conceptual level. This is all a dynamic network in which "webs of meanings and associations constantly shift and re-adjust" (McCarthy, 1990, p.42 as cited in Wolter, 2001). Assuming the importance of the network knowledge, we need to be able to measure it.

A line of investigation which has revealed a lot about the lexical relations and structure of the bilingual mental lexicon is the well-known elicitation tool, free word association (Entwistle, 1966; Palermo, 1971; Meara, 1978; Politzer, 1978; Söderman, 1993; Singleton, 1999; Wolter, 2001; Namei, 2002, 2004; Henriksen, 2008; Kormos & Dóczi, 2016). The extent to which word association provide insight into the structure of the lexicon is based on theories of spreading activation (Dell, 1986). As discussed earlier (see Section 2.4.3), it can be argued that lexical items used more frequently have higher levels of activation, which in turn help them to be selected and retrieved more rapidly and easily (Levelt, 1989, 1999; Kormos, 2006a), which is reflected in the result of word association tests.

Word association is considered to tap on both conceptual and lexical levels of knowledge representation and provide a comprehensive measure of the network knowledge. Unfortunately, not many studies address this issue, and the underlying assumption implies that word association mainly taps the connections in the lexical level of knowledge among the lexical items (Henriksen, 2008). Nevertheless, Kormos and Dóczi (2016), inspecting changes in the structure of the bilingual mental lexicon with changes in proficiency, argue that word association can manifest changes in the conceptual level as well, as this level of knowledge undergoes "substantial reorganization with new meaning senses being added and L1-based meaning associations refined" (p. 116). In addition to these two levels of knowledge representation, the links between the conceptual level and the lexical level undergoing changes and modifications is tapped via word association as well. This can be shown by less frequent use of L1 responses to L2 cue words, which is a sign of changes in routes between concepts and L2 lexical items which are strengthened, and the importance of L1 lexical items becoming less prominent in accessing L2 items (e.g. Dóczi & Kormos, 2016; Navracsics, 2007). Accordingly, it can be summarized that word association is the best manifestation of the changes undergone in both conceptual and lexical levels with connections in each level as well as connections between levels (Kormos & Dóczi, 2016), and can be used in the current study as a measure to reflect the changes occurred in the network knowledge.

Knowing that word association is a manifestation of the changes in the network knowledge, the next step is adapting ways by which such reorganization and changes can be envisaged. Both qualitative and quantitative paradigms have been applied widely to reflect bilingual mental lexicon and lexical development. Reorganizations in the bilingual mental lexicon network manifested through word association can be conceptualized in two ways; first the number of connections among items and secondly, type of connections among them (Henriksen, 2008).

Number of items and connections among items in the bilingual mental lexicon is noteworthy as a denser network of lexical items with many links among them, can play an extremely important role in dealing with real communication. This number is not stable, and constantly change due to various factors such as practice, exposure, experience, leading to continuous establishment of connections between items in the lexicon (Wolter, 2001). In general, increase in L2 competence and more exposure to L2 input aid the incremental process of link-building mechanism (Kormos & Dóczi, 2016). However, Wilks and Meara (2002) speculate that "there is not a constant build-up of links for any one item, but rather that some associations may atrophy or become dormant as others are formed" (p.321). Accordingly, the process is not always a purely incremental one with new nodes gradually added to the network or nodes receiving higher levels of activation. In the process of lexical processing and learning, new links between lexical items might be created or strengthened, while some others gradually become weaken and finally disappear due to lack of activation (Meara, 1983, 2004). Accordingly, as the number of items and links are not constant, the changes in the number of them can be a reflection of modifications in the structure of the lexicon. A denser and more structured network is what is expected in the process of lexical organization.

As a result of these rotations and replacements, nature and type of connections is another point which is susceptible to changes in the process of restructuring the internal structure of the network. This line of analysis concerns the investigation of lexicosyntactic patterns. Traditionally, the categories identified are paradigmatic, syntagmatic, and clangs (Entwisle, 1966). Paradigmatic responses are the ones with same grammatical functions as the cue word and are of five different types of superordinates, subordinates, coordinates, synonyms, and antonyms. Syntagmatic have sequential relationship with the cue word and are not of the same word class, and finally, associations with phonological relationship, bearing no semantic relatedness, are clangs. Taking the word *accident* as an example, a paradigmatic response can be *crash,* which is a synonym of the cue word and of the same word class, a syntagmatic one might be *cause*, and *access* can be considered as a clang type of response having only phonological relationship with the cue word.

All these different types of lexical items with various types of connections are assumed to be stored differently in the bilingual mental lexicon (Van Hell & De Groot, 1998), though the point of commonality is that they are mostly stored according to their meaning relation they have to one another (Navracsics, 2007). Some items in the network might take a more central position with higher number of connections of various lengths, that is degrees of strength, to other nodes as a result of more frequent activation. Such close and strong links are assumed to bear mostly paradigmatic connections, and become more conventionalized and stereotypical, that is more canonical. However, some other nodes might take a more peripheral position in the network with fewer number of links and lower level of activation (Wilks & Meara, 2002). The lexical items in the center are of more significance as they have higher number of connections with other items in the network or their position in the lexical network is "of more strategic significance in the overall structure of the association network by acting as link points between different clusters of associations within the network" (Wilks & Meara, 2002, p.322). These items found in the center are more likely to have paradigmatic and meaning-based links with other items in the network, while the ones further from the center tend to have mostly syntagmatic and form-related connections. Finally, clangs found in the periphery, also called "miscellaneous" (Navracsics, 2007, p. 20), are assumed to be based on similar phonological sounding or even random memories, retrieved based on no linguistic association (Wolter, 2001).

Accordingly, in an attempt to investigate word-association domains in the current study, it is postulated that both qualitative and quantitative paradigms, that is shifts in the number of lexical items and links in addition to the type of the connections, can be studied through word association. This helps in manifesting the internal structure of the network, that is representing connections within and between conceptual and lexical levels of knowledge representation. Both number and type of connections in the lexicon are subject to change as the lexical items either become more integrated to the network and develop higher number of connections with other items or become dormant and fade away. The ones in the center with many connections develop more retrieval paths, assisting more automatized lexical retrieval and more rapid use of the lexical items in language use. Such changes to the structure of the lexicon can be determined by the type of instructions L2 learners receive (Kormos & Dóczi, 2016).

2.5. Oral fluency and bilingual mental lexicon

Very recently, fluency of speech has been extensively researched in relation to the knowledge of lexical items (Clenton, et al., 2020; De Jong, et al., 2012; 2013; Kahng, 2014; 2020; Koizumi & In'nami, 2013; Segalowitz & Freed, 2004; Uchihara & Saito, 2019), the area which previously very few number of studies attempted to explore, regardless of the evident and clear link between them (e.g., De Bot, 1992). Hilton (2008) argued that one of the main reasons lied on the relative "complexities of analyzing just what is going on when an individual performs the complex task of talking in a foreign language" (p. 153). In addition, multicomponential nature of lexical items and what specifically constitutes lexical knowledge (e.g., Meara, 2005; Schmitt, 2010; Webb, 2007) intensify the complexities. This section particularly deals with the role, lexical items play as one of the sources of dysfluency of speech in addition to the relationship between the rapidity and smoothness of speech production and knowledge of lexical items.

In one of the earliest pieces of second language research, Maclay and Osgood (1959) conducted an explanatory investigation of hesitation phenomena in English spontaneous speech and reported that lexical items are the greatest impediments in fluent speech. In this study, 103 spontaneous utterances, including 5000 words by 13 speakers, from a conference were transcribed for analysis. The selection of the utterances was not random, but the researchers chose longer ones with more than 80 words in each. Four types of hesitations were identified; repeats, false starts, filled pauses, and unfilled pauses. Among all transcribed hesitations, only the ones that two judges agreed on the type and location, were chosen for further analysis. Their findings revealed that pauses of both types, filled and unfilled pauses, most frequently occur before lexical item. In addition, false starts typically involved lexical items. According to the results of this study, one of the main problems of dysfluent talk is linked to problems with lexical encoding. In line with this, in a more recent study Kahng (2014) examined cognitive fluency of 17 participants qualitatively via stimulated recalls, in order to have insights regarding the underlying L2 fluency problems. The results were in congruence to the previous study, indicating that participants faced more significant challenges with vocabulary and the content of the talk, in comparison to grammar, phonology and pragmatics. The problem with vocabulary was of various qualitative types; some were caused by the lack of knowledge of the L2 words by translating them from their L1, and some others were due to doubts between various choices of words or multi-words, they were not sure which lexical item serves the meaning they were looking for conveying or fits the context better. It can be concluded that as Hilton (2008) claims "lexical knowledge is the greatest impediment to spoken L2 fluency" (p. 162).

Apart from the theoretical importance of lexical items in fluent production of language, more recently, empirical studies have been conducted to explore the relationship between the knowledge of lexical items and narrow sense of fluency of speech. As already discussed (see Section 2.4), knowledge of lexical items is a multidimensional concept, and can be scrutinized from two different dimensions of mental lexicon properties; that is, the breadth and depth of knowledge. Table ***2.4*** summarizes a few number of previous studies exploring this relationship quantitively from both dimensions.

Table 2.4 Previous studies analyzing relationship between oral fluency and lexical knowledge

Study	*Participants*	*Fluency measures*	*Vocabulary measures*		*Results*
Clenton, de Jong, Clingwall, & Fraser (2020)	30 pre-intermediate Japanese EFL learners	- Silent pause duration between AS units - Silent pause duration within AS units -Number of silent pauses per second -Number of filled pauses per second -Number of repetitions per second -Number of corrections per second -Mean syllable duration	Lex30		$r = -.06$ $r = -.15$ $r = -.39^{*}$ $r = -.17$ $r = .22$ $r = -.02$ $r = .12$
			X_Lex		$r = .24$ $r = .18$ $r = -.16$ $r = -.14$ $r = .13$ $r = .12$ $r = .02$
Kahng (2020)	44 Chinese EFL learners	-Syllable duration -Duration of silent pauses -Number of mid-clause silent pauses -Number of Final-clause silent pauses -Number of mid-clauses filled pauses -Number of final-clause filled pauses -Number of repetitions -Number of corrections	Breadth of lexical knowledge		$r = -.13$ $r = -.12$ $r = -.41^{*}$ $r = -.33^{*}$ $r = -.25$ $r = -.18$ $r = -.17$ $r = -.19$
			Test of lexical retrieval		$r = .50^{*}$ $r = .22$ $r = .30^{*}$ $r = .20$ $r = .32^{*}$ $r = .07$ $r = .02$ $r = .05$
Uchihara, Saito, & Clenton (2020)	40 Japanese EFL learners	Articulation rate Filled pause ratio Silent pause ratio	Lex30		$r = .48^{*}$ $r = -.03$ $r = -.43^{*}$
Uchihara & Saito (2019)	39 Japanese EFL learners	Optimal speech rate	Lex30	Raw score	$r = .34^{*}$
				Percentage score	$r = .28$

De Jong, Steinel, Florijn, Schoonen, & Hulstijn (2013)	179 EFL learners of Dutch	-Number of silent pauses -Mean silent pause duration -Number of filled pauses -Number of corrections -Number of repetitions -Mean syllable duration	Lexical knowledge (Test of breadth and depth)		$r = -.39^*$ $r = -.02$ $r = -.33^*$ $r = -.43^*$ $r = -.24^*$ $r = -.58^*$
			Processing skill of lexical retrieval		$r = .20^*$ $r = .16^*$ $r = .32^*$ $r = .25^*$ $r = .16^*$ $r = .32^*$
Koizumi & In'nami (2013)	224 Japanese EFL learners	Speech rate (a) Repairs (b)	Processing skill of lexical retrieval		(a) $\beta = .57$ $R^2 = .32$ (b) $\beta = .36$ $R^2 = .13$
			Depth of lexical knowledge	Derivation	(a) $\beta = .47$ $R^2 = .22$ (b) $\beta = .30$ $R^2 =.09$
				Antonyms	(a) $\beta = .49$ $R^2 = .24$ (b) $\beta = .31$ $R^2 = .10$
				Collocations	(a) $\beta = .41$ $R^2 = .17$ (b) $\beta = .26$ $R^2 =.07$
	87 Japanese EFL learners	Speech rate (a) Repairs (b)	Breadth of lexical knowledge		(a) $\beta = .69$ $R^2 = .64$ (b) $\beta = .34$ $R^2 = .16$
			Lexical organization test		(a) $\beta = .68$ $R^2 = .48$ (b) $\beta = .34$ $R^2 =.12$
			Lexical access time test		(a) $\beta = .21$ $R^2 = -.46$ (b) $\beta = .05$ $R^2 =-.23$
Hilton (2008)	56 non-native speakers from various L1 and L2s	- words per min - mean length of run - mean length of hesitations - Percentage of production time spent hesitation - Rate of hesitation - Rate of retracing	Breadth of lexical knowledge (DIALANG)		$r = .58$ $r = .67$ $r = -.39$ $r = -.55$ $r = -.66$ $r = .52$

As it is clear from the table, this line of research has been largely devoted to the exploration of the relationship between oral fluency and the breadth dimension of the mental lexicon; that is, the correlation between various measures of utterance fluency and number of form-meaning connections in the mental lexicon of the bilingual speakers. For example, Uchihara et al. (2020) investigated the relationship between the productive knowledge of lexical items and second language oral ability. They administered two independent tests of speaking and vocabulary to elicit each measure separately. As for the productive test of vocabulary, Lex30 word association task was administered, which was created by Meara and Fitzpatrick (2000). Lex30 is composed of 30 cue words asking four associated responses for each cue. This test is widely employed for research purposes as it is argued to be an “indication of learners’ lexical resources in addition to fluency” (Uchihara, et al., 2020, p. 151). Besides, spontaneous speech was elicited using eight-frame picture narrative task with one-minute planning time. The results of the correlation between six objectives measures of utterance fluency and Lex30 revealed that Lex30 raw scores was only significantly but moderately correlated with articulation rate (r = .48) and silent pause ratio (r = -.43).

In contrast to the findings of the study above, Clenton et al. (2020) found out that none of the measures of utterance fluency in their study, except for the number of silent pauses per seconds, significantly correlated with Lex30 productive vocabulary size test. In this study, 30 pre-intermediate undergraduate Japanese EFL students took part in the study. Participants performed three different speaking tasks, adapted from De Jong et al. (2013), from which various measures of utterance fluency were elicited. Besides, Lex30 was used as a measure of productive lexical knowledge. They found that the only measure of utterance fluency correlated significantly with the Lex30 task was the number of silent pauses per seconds; that is, the lower scores in Lex30 was associated with the higher number of silent pauses in speech. In broad terms, the results of this study are in congruence with De Jong et al. (2013) and Kahng (2020), arguing that higher scores on productive lexical test correlates significantly and negatively with measures of silent pausing. De Jong et al. (2013) explored the relationship between oral fluency and both linguistic and processing lexical skills in a large-scale study (179 speakers of Dutch as second language), while taking the individual differences into account. Various measures of utterance fluency were elicited from eight speaking tasks. In addition, participants took part in a test of lexical knowledge (The Productive Vocabulary Levels Test (PVLT), Laufer & Nation, 1999) and lexical retrieval to assess the linguistic processing skill. The results revealed weak to moderate strength but nevertheless significant correlation between fluency measures and both productive vocabulary knowledge and lexical retrieval skill.

Apart from measures of silence in these studies, mixed results were reported on pure measures of speed. Although Uchihara et al. (2020) and De Jong et al. (2013) reported significant correlation between the syllable duration (inversion of articulation rate) and productive measure of lexical knowledge, Clenton et al. (2020) and Kahng (2020) found no relationship.

Consequently, reviewing the studies in this realm reveals that there is little consistency in the results reported. One of the main reasons can be attributed to the multidimensionality of knowledge of lexical items and lack of consistency in assessing which aspects of lexical knowledge each study targets (Koizumi & In'nami, 2013). In this regard, four of the above mentioned studies (Table ***2.4***) primarily explore the breadth dimension of the mental lexicon (Clenton, et al., 2020; Hilton, 2008; Uchihara & Saito, 2020; Uchihara, et. al., 2020), with three studies measuring both breadth and depth of knowledge (De Jong, et al., 2013; Kahng, 2020; Koizumi & In'nami, 2013). Furthermore, another factor can be contributed to the claims made considering the aspects being measured. For example, De Jong, et al. (2013) argued that the test of the PVLT administered as a way to explore the productive lexical knowledge of the participants; however, the test was not only a test of breadth since as many as 22% of the questions tapped into the depth of lexical knowledge instead. Among the mentioned studies in the above table (table 4), the only study specifically focused on the depth of lexical knowledge was conducted by Koizumi & In'nami (2013). This large-scale study aimed to find out the degree to which L2 fluency of speech can be predicted by depth of lexical knowledge. The study was conducted in two separate phases. The first part targeted the extent to which speed and repair measures of utterance fluency could be explained by knowledge of derivations (20 items), antonyms (17 items), and collocations (18 items) among 224 novice to intermediate Japanese learners of English. In addition to the test of depth of lexical knowledge, participants were asked to complete 5 different speaking tasks, which result into 15 minutes of real-time monologues. They found that depth of lexical knowledge predicted speed fluency 17% to 24% and repair fluency, only 7% to 10%. More specifically, knowledge of derivations predicted speed fluency strongly ($\beta = 47$), explaining 22% of the speed fluency factor variance; repair fluency moderately (β = 30), with 9% only explained. Antonyms revealed similar proportions of variances of derivations, explaining 24% of the speed and 10% of repair fluency. Finally, collocations predicted the least proportion of variances of speed fluency 17% and repair fluency only 7%.

The results suggested that although speed fluency could be predicted by knowledge of the lexis to a substantial degree, repair fluency could not. They argued that an explanation for the gained results can be grounded on the notion that "L2 learners with larger and deeper vocabulary knowledge and faster access to it, can perform lexical searches more easily and quickly" (Koizumi & In'nami, 2013, p. 911), which can finally contribute to more rapid and smoother processing and production of the second language. However, this study is limited in the fact that the depth of lexical knowledge was investigated from merely a decontextualized word-based perspective by analyzing only three aspects of depth of lexical knowledge, and generalizations were made accordingly. In addition, repair fluency in this study was measured by the number of dysfluent markers, such as repetitions, self-repairs, and voiced pauses; although, each of these dysfluency markers can be considered as an indication of a specific problem in speech processing system and grouping them together as a single dysfluency measure leads to inaccurate results.

Consequently, it can be summarized that the scarcity, divergence and lack of consistency in the results calls for further research into the relationship between various aspects of lexical knowledge and fluency of speech. Many of such studies on the breadth dimension of the mental lexicon suggested that a larger lexicon and higher number of form-meaning connections in bilingual mental lexicon might equip L2 learners with well-organized lexical networks (Meara, 1990; 2006 cited in Uchihara & Saito, 2019, p. 69) with the help of which the process of lexical retrieval would be carried out in more economic and immediate manners in real time communication. However, as it has been already discussed, a well-structured bilingual lexicon is the one with various number of strong links between and within the network. Therefore, making claims about the structure of the network knowledge and the strength of connections between its items and the impact on L2 oral fluency, merely on the basis of the number of form-meaning connections in the lexicon may yield to misleading results.

Therefore, in order to scrutinize the structuredness of the lexicon and assess its impact on L2 fluency of speech, conducting research on depth of knowledge would be advisable; however, the contribution of depth of knowledge have been mainly researched by having a word-centered perspective previously; that is, exploring the collocations, antonyms, derivations, etc. Consequently, adopting a lexicon-based perspective, concentering on the connectivity and strength of connections among lexical items in the bilingual mental lexicon, would provide a more comprehensible understanding of the degree of structuredness and organization of the lexicon, which can have implications for the fluent production of speech.

In addition, although extensive research in this realm indicated that there is a relationship between knowledge of lexical items and fluency of speech (e.g., Clenton, et al., 2020; De Jong et al., 2013; Kahng, 2020; Segalowitz & Freed, 2004; Uchihara & Saito, 2019), and also the fact that fluency can be effectively predicted by vocabulary knowledge (e.g. De Jong, et al., 2012; Koizumi & In'nami, 2013), further experimental studies are necessary to examine if enhancing lexical knowledge, concentrating on network knowledge, actually leads to an increase and improvements in oral fluency (Koizumi & In'nami, 2013, p. 911). The following section provides a pedagogical perspective to the teaching of fluency, concentrating on the knowledge of lexical items from a lexicon-based perspective.

2.6. Teaching fluency

While understanding what fluency is, how it can be measured, the dominant language processing theories, fluency vulnerability points which in turn lead the discussion to the idea of bilingual mental lexicon and lexical networks, are important, a crucial factor to bear in mind is that how all these can be implemented into language teaching and learning context; that is, how fluency can be developed through formal instruction. Derwing (2017) points out;

> Applied linguists typically want to identify ways in which learners' fluency can be enhanced through manipulation of tasks in the classroom, the effects of study abroad or other forms of immersion in the L2, the fluency trajectories of learners, and the interrelationships that effect the fluency of utterances produced by L2 speakers. (p. 248)

Various studies have tried to pin down the potential sources which are associated with having an impact on fluency of speech. Some of the sources have been categorized as having internal impact such as the age of the speaker when exposed to L2 (e.g., Llanes & Munoz, 2013), the level of language proficiency (e.g. Kormos & Dénes, 2004), working memory (e.g. Kormos & Safar, 2008), while some other sources can have external impact, like task repetition (e.g. Ahmadian & Tavakoli, 2011; Bygate, 2001; De Jong & Perfetti, 2011; Goldman-Eisler, 1968; Nation, 2008; Thai & Boers, 2016; Wang, 2014; Wood, 2010), fluency strategy instruction (e.g. Rossiter, 2003; Tavakoli et al., 2015), planning time (e.g. Ahmadian & Tavakoli, 2011; Ellis, 2009; Foster & Skehan, 1996), familiarity with the task (e.g. Sample & Michel, 2014; Bui & Huang, 2018; Michel, 2017), and so on. The second group can mainly be associated with formal instruction applied in classroom setting to develop fluency of speech. Although it has been assumed that fluency cannot be taught and emerge naturally in context (Chamber, 1997), pedagogical interventions in the classroom prove otherwise.

In order to fill the gap between the theoretical and practical aspects, Nation (2008) introduced four strands of instruction. The four strands were originally concentrated on curriculum and pedagogy with L2 language learning context; however, the principles they are based on also apply to chances for lexical development. The four strands are meaning-focused input, meaning-focused output, language-focused learning, and finally fluency development. To reach success in teaching and learning, each of these strands needs to be practiced within approximately same amount of time in the classroom. With respect to vocabulary, Table ***2.5*** summarizes the four strands of instruction and the way that they can best be implemented.

Table 2.5 Nation's (2008) four strands of vocabulary instruction (Dóczi & Kormos 2016)

Type of instruction	Meaning-focused input	Meaning-focused output	Language-focused learning	Vocabulary fluency development
Aim	Practicing listening and reading	Practicing speaking and writing	rich vocabulary instruction; boosting word consciousness; fostering vocabulary strategy uses	strengthening the existing links between lexical items; creating new associations through reorganizing existing knowledge
How to achieve	graded readers; narrow reading; movies; sitcom, series on TV	integrated output tasks, writing & presenting short texts with new vocabulary	focus on depth of word knowledge through form-focused instruction; vocabulary strategy instruction; training in dictionary use	recycling and rehearsing familiar vocabulary and topics

With respect to vocabulary, the first two strands mainly focus on incidental vocabulary learning with large amount of input and output, while the third entails intentional learning in which various features of language are learnt by applying different learning strategies. The fourth one, focusing on fluency development, concentrates on maximizing the speed and automaticity of recognition and production of lexical items for all four language skills. In fluency development activities, learners are not asked to learn new words, but to practice and make best use of what they have already learnt. Nation (2007) argues that "if the activity involves unknown vocabulary, it is not a fluency activity" (p.8). Consequently, the fourth strand targets the retrieval of meaning processing (Yang, 2014).

On that account, fluency development strand has two main aims. The first is strengthening the links and connections among lexical items already acquired and the second is creating and adding new connections, which can be gained through restructuring the existing knowledge. In other words, these two aims target one of the dimensions of mental lexicon; that is, the depth of lexical knowledge in the mental lexicon within a lexicon-based perspective. More particularly, this strand doesn’t aim to create form-meaning connections (developing the breadth of the lexicon), but instead to strengthen the connections among various items in the lexicon, contributing to the structure and organization of knowledge stored in the lexicon, and how information is structured. As mentioned earlier, mental lexicon is referred to as a network of associations, “a web-like structure of interconnected links” (Sökmen, 1997, p. 241). The lexical content stored in it, which in turn needs to be retrieved in speech production process, is not presented as “mere aggregation of independent words” (Stubbs, 2001), but instead the nature is accumulation of interrelating network of relationships among words.

A way to reach fluency strand goals is through retrieving the knowledge already stored in the lexical network (Nation, 2001), which is facilitated by inter-item association among concept nodes, whose semantic representations overlap. Retrieval benefits the activation of semantically related nodes in the network. Baddeley (1999) points out that “each retrieval strengthens the link between form and meaning”. In addition, it can be linked to the increased centrality of the node in the network. Consequently, it can be summarized that in order to make changes to the organization of the mental lexicon and the way the information is accessed to check further changes on the fluency of speech, the strength of the connections among the items stored and also the centrality of the nodes in the network can be manipulated.

Concentrating on pedagogy, it has been argued that a way to achieve the goals is through “recycling and rehearsing familiar vocabulary and topics” (Dóczi & Kormos, 2016a). That is, teachers need to get learners engaged to work and practice with the same input they have already acquired in different ways. Focusing on speech fluency, a way can be practicing tasks orally by reproducing or paraphrasing the language already known (Laufer & Nation, 2001). This way entails production of the lexical items rather than enhancing and improving the recognition. Productive retrieval is one of the most important general cognitive processes that can lead to words being remembered and retrieved faster.

A review of literature clearly states the undeniable role practice plays in fluency development at the formal context of the classroom. The following section involves a discussion of an innovative approach, incorporating theories of mental lexicon, providing L2 language learners with a type of practice mitigating the degree of dysfluency in accordance with models of speech processing.

2.6.1. Semantic mapping; A novel approach to fluency development

In order to incorporate the factors introduced in vocabulary fluency development strand along with theories of bilingual lexicon in the classroom to help language learners with the fluency of speech, semantic mapping technique was introduced. This technique has been originally introduced as a way of elaborating semantic networks (Sökmen, 1997). While semantic networks are made up of sets of nodes and connections of different degrees of strength (Collins & Loftus, 1975), they can be elaborated and promoted by mentally connecting new items with the ones already formed and known (Boers & Lindstromberg, 2008), which is the basic prerequisite of learning (Baddeley,1990). Accordingly, this technique has been primarily applied to teach new vocabulary by incorporating the new knowledge to the old (Badr & Abu-Ayyash, 2019; Khoii & Sharififar, 2013; Kogok & Ahamed, 2017; Margosein, Pascarella, & Plfaum, 1982; Mohammed & Malo, 2020; Morin & Goebel, 2001; Saragih, 2019; Schmitt & Schmitt, 1995; Stoller & Grabe, 1993; Svenconis & Kerst, 1994; Zahedi & Abdi, 2012), and has its basis on the realm of memory research, arguing that integrating a new item to a rich network of old-established interwoven associations helps to facilitate retrieval (Baddeley, 1990).

Semantic mapping serves as a visual representation of the associative network of relationships among lexical items, taking "advantage of the way the brain organizes information" (Pittelman, Heimlich, Berglund, & French, 1991, P.5). It is a classification strategy and a type of graphic organizer, resulting in a network of ideas and concepts interlinked together, while providing an opportunity for students to have a visual display of related ideas of a concept.

This technique has its theoretical foundation on the notion of spreading activation within the semantic system (Collins & Loftus, 1975). It is an approach which is believed to access semantic networks and increase the strength of the links by tapping and activating already stored lexical knowledge, building up on the schemata already presented in memory structure, leading to higher level of activation in the semantic network, and finally leading to faster retrieval of lexical items (Boyle & Coelho, 1995). This technique establishes links between items, in order to better store them in memory and facilitate retrieval (Sökmen, 1997).

From the pedagogical perspective, semantic mapping has been applied in classroom context both as a teaching tool and a classroom language instruction (Nation, 2001; Sánchez, 2004), with a major point of strength that “it helps students to construct a model for organizing and integrating information that can be applied to a wide variety of situations" (Al-Baili, 1988, p.2). It has been argued that semantic mapping technique facilitates drawing “a meaningful connection between what the learner knows and what the learner is learning” (Nation & Gu, 2007, p.90), as it organizes semantically similar lexical items into clusters (Lowe, 1991). Research on the impact of exercises enhancing formation and strengthening of the links on semantic network demonstrates positive results of long-term retention of the vocabulary (Hague, 1987; Machalias, 1991), since based on the depth of processing hypothesis the cognitive energy that one exerts on manipulation and thinking about a lexical item helps to recall and use it rather more easily subsequently (Craik & Lockhart, 1972).

Semantic mapping involves the teacher and learners working as a group towards creating a map, which consists of a series of single concepts, placed in circles, squares, or rectangles, along with lines of various length connecting them together, as illustrated in Figure ***2.8***.

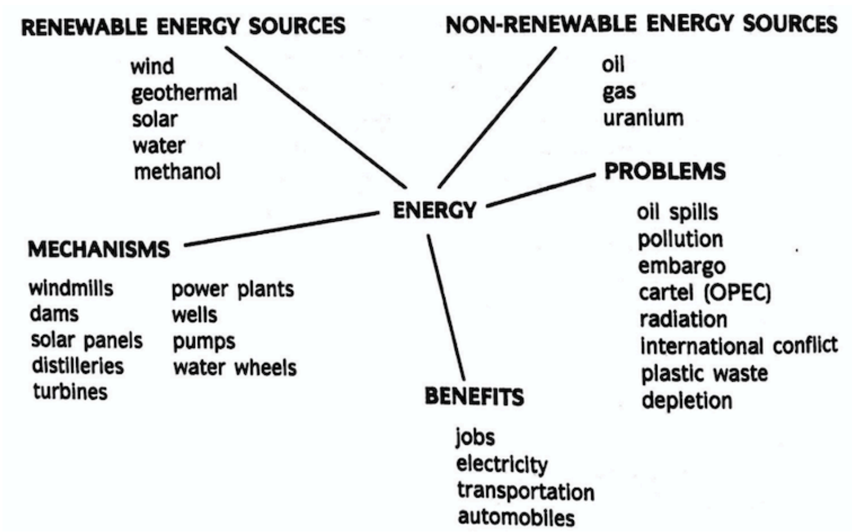


Figure 2.8 Semantic map on Energy (Stoller & Grabe, 1993)

The process can simply be initiated by recalling an earlier read story, a general topic like *travelling*, a current event, a film watched, or subject of the unit of study. Learners are encouraged to produce lexical items and place them in the map while categorizing the items according to the semantic relations among them. Meanwhile, the teacher’s main responsibility is motivating the learners in two processes; firstly, encouraging them to take part in the process, which can be accomplished by providing hints when they run out of ideas, such as translation equivalents, paraphrases, or formal clues such as the initials, and secondly, creating opportunities for the learners to explain, justify, and increase the links between the lexical items in the map since it encourages repetition which paves the way to creative use of the lexical items in association with others, and provides an opportunity for the teacher to rephrase their production regarding grammatical or collocational features of the lexical items (Nation, 2001). To gather up, it is assumed that it is not only the visual representation of the lexical items and the relations among them which leads to the effectiveness of this technique in vocabulary learning, but also it is the discussion and debate in the process of building up the map, which is argued to aid productive vocabulary learning (Pittelman, Levin, & Johnson, 1985; Stahl & Vancil, 1986). That is, active engagement in the process of creating the concepts and the associative semantic relatedness among them plays a significant role (Blachowicz et al., 2006).

Stahl and Vancil (1986) in their study investigating the effectiveness of semantic mapping in learning vocabulary on 45 students in three different groups of semantic mapping, discussion only, and semantic mapping and discussion found out that there was a significant difference between both groups with discussion and the one with semantic mapping only. The results reveal that the active engagement of the learners in the process of creating meaning leads to effective vocabulary learning, though no significant correlation was reported between the classroom contributions in the process of discussion and vocabulary learning, which is in alignment with the results of a study by Pittelman et al. (1985). The underlying reason can be due to the process of active thinking and formulating possible associations leading to maintain desirable results. In addition, the results can be discussed on the grounds of development in lexical knowledge; that is, in the process of discussing the meaning through, participants could have developed knowledge of lexis.

One of the few studies on the use of semantic mapping as a classroom language instruction tool in assessing the cognitive organization was carried out by Sánchez (2004). This study investigated if teaching vocabularies belonging to a certain semantic field of the concept *shine* via semantic mapping, would lead to changes in the cognitive organization, as a result of incorporating new knowledge to that which has already been acquired. In this study, changes in the cognitive organization were measured through four variables of lexical test, proximity (relatedness ratings), distance (minimum distance between the nodes in a network) and similarity (similarity between networks) (Sánchez, 2002; 2004). The results strongly support the use of semantic mapping as a technique to restructure the organization of knowledge, which in turn lead into long term retention of new words. Although this research provides valuable insight into the relationship between semantic mapping and organization of the lexical knowledge, it misses the concept of automaticity of use of the lexical items in the natural and spontaneous setting, as it mainly concentrates on tasks which were directly taken form textbooks.

The only study on this realm assessing the efficiency of concept mapping on the spontaneous use of vocabulary and fluency of speech was conducted by Ghonsooly and Hoseinpour (2009). This research was based on the assumption that concept mapping helps participants to increase their vocabulary knowledge as it provides opportunities to present knowledge in an organized way, which in turn has an impact on L2 fluency of speech. Eighty participants in the intermediate level of language proficiency took part in the study, and the ones in the experimental group received training with concept mapping for twenty-two sessions. Participants from both groups took part in speaking tests both prior and after the training, and the speech was analyzed in terms of speech rate, articulation rate, mean length of fluent runs, number of silent and filled pauses, mean length of pauses, and number of disfluencies. The results of analysis revealed that semantic mapping had a significant impact on all measures of fluency except for mean length of runs, mean length of pauses and number of filled pauses per minute. It was argued that application of concept mapping as a technique to familiarize students with more lexical items helped to activate large number of vocabularies, which were passively stored in the lexicon.

Consistent with the aims of the current research, semantic mapping was not applied as a teaching technique tool for new lexical items, rather it was a technique of instruction to fulfil the two aims of fluency development strand. That is, it was applied as a way of strengthening the connections among already known lexical items in the process of productive retrieval, and also creating new connections among items within and between the lexical and conceptual levels by processing the information according to semantics, as processing lexical items according to meaning leaves stronger memory traces in comparison to other types of processing such as the ones based on sounds (Craik & Lockhart, 1972). This technique is adopted as a way to productive retrieval, that is recycling the lexical items by repeating them and retrieving them in an effortful manner which “promotes activation of more elaborative information, relative to less effortful retrieval, hence establishing more retrieval routes and increasing later retention” (Kang, 2010, p. 1009).

2.7. Summary

The nature and developments in L2 fluency of speech has been the focus of wide range of research (e.g. Derwing et al., 2009; Hilton, 2008, 2009; Kahng, 2014, 2018, 2020; Koponen & Riggenbach, 2000; Kormos & Dénes, 2004; Segalowitz & Freed, 2004), and it has been proven that linguistically encoding one's intentions in a second language can be one of the most difficult and complex activities to be carried out in fluent and automatic manner (Kormos, 2006a). In this chapter, bilingual speech processing mechanism was addressed thoroughly along with various points where difficulties in processing might bring about problems in fluency of speech. It was proposed that not only the processes involved, but also the knowledge stores, which "do not share processing functions" (Kormos, 2006a, p. 7) can also have a significant impact on L2 fluency of speech (Singleton, 2009). However, their impact on reaching automatic control over the process of speech production and fluency have been overlooked. Accordingly, mental lexicon is introduced as one of the fundamental components in language processing, containing language specific pieces of knowledge about lexical items, which are regarded as one of the main obstacles in the process of fluent production of speech (Hilton, 2008, p. 162).

The current study takes a lexicon-based perspective of the bilingual mental lexicon, focusing on the strength and number of connections lexical items can have with other associated and related ones in the lexicon, proposing that fluency can be the outcome of a more semantically organized mental lexicon. That is, it is assumed that a more structured mental lexicon with many strong connections and paths among its items would consequently increase the speed of producing utterances in L2. In order to attain this, making changes to the strength of the connections among the items stored in the lexicon and also the centrality of the nodes in the network was suggested, and it is aimed to find out if application of semantic mapping, as a way of strengthening the links between the semantic representations and word forms in the mental lexicon and also elaboration of the semantic networks, can help students gain improvements in L2 fluency and also if any reorganization occurs as a result in the structure of the lexicon.

Consequently, the current study aims to weave the lines of fluent bilingual speech production and the structure of the bilingual mental lexicon, together, which might be explained with the help of strength theory of automaticity, focusing on strengthening the connections within the hierarchical system of language to maintain smoother and faster production of speech, that is more fluent outcome, and answer the following research questions;

RQ1: To what extent does semantic mapping affect organization of bilingual mental lexicon as measured through word association?

RQ2: To what extent does semantic mapping affect L2 speech fluency measured through empirically established means of gauging L2 fluency?

RQ3: Is there a relationship between organization of bilingual mental lexicon and L2 speech fluency?

3. Chapter 3: Methodology

3.1. Overview

As mentioned in previous chapter, the aim of the current study was to investigate the impact of a certain type of vocabulary instruction technique on ELF adult learners' L2 fluency of speech and the structure of the bilingual mental lexicon. To be more precise, it aimed to find out whether training EFL students with semantic mapping could contribute to improved and enhanced measures of L2 utterance fluency, and also if this technique tended to result into a more structured and organized network knowledge of specific strata of the lexicon. To this end, the study addressed the following research questions (RQs):

RQ1: To what extent does semantic mapping affect organization of bilingual mental lexicon as measured through productive word association?

RQ2: To what extent does semantic mapping affect L2 speech fluency measured through empirically established means of gauging L2 fluency?

RQ3: Is there a relationship between organization of bilingual mental lexicon and L2 speech fluency?

The methodology and methods, which were employed to find out answers to the above research questions, will be presented below in detail. To investigate the impact of semantic mapping on EFL learners' utterance fluency and network knowledge, the research strategy adopted in the current study and justifications for application of experimental design are provided (see Section 3.2). Following this, the participants of the study and the rational for inviting this specific group are introduced (see Section 3.3). Next, the design of the study, that is the overall strategy chosen integrating various components of the study to effectively answer the research questions, is presented thoroughly (see Section 3.4). Subsequently, the instruments and the testing materials are presented (see Section 3.5), which is followed by a detailed discussion of the procedure followed (see Section 3.6). Data analysis and pre-analysis procedures are addressed in analysing the data collected from the narratives and word association test are discussed (see Section 3.7). Next, ethical considerations of classroom-based experimental research are considered (see Section 0). Finally, the chapter was concluded with a brief explanation of the pilot phase of the study and the changes applied in the main phase (see Section 3.9).

3.2. Research strategy

The current study employed an experimental design, since it tried to find out the impact of the certain type of vocabulary instruction technique as the independent variable on the participants' utterance fluency, as well as the organizational properties of the mental lexicon, which were both considered as the dependent variables of the study.

3.3. Participants

The participants were adult EFL learners from Iran, ranging in age between 18 to 25. The participants were selected from language learners studying English as a foreign language at a private language school in Iran, teaching English to L1 Farsi speakers as a foreign language. None of the participants knew any other language. This language school offered classes to both males and females, ranging in level of language proficiency from beginner to advance. Furthermore, in order to provide increased instructional interaction opportunities among teachers and learners, the number of students in each class did not exceed ten. Language learners in this language school took two session classes of one hour and half for ten weeks for each term of studies.

Participants were in intermediate level of language proficiency as measured by standard institute's placement test. The reason why this level of language proficiency was opted for was two folded. First and foremost, the intermediate level of language proficiency was chosen, since it is believed that fluency improves and enhances as long as learners are involved in activities that are within their prior experience and have already had familiarity with the topic and discourse (Nation, 1996). Accordingly, beginners were not in the suitable group of language proficiency, since they might not have had previous experience with the topic and discourse, as they had newly started.

Secondly, intermediate level language learners can benefit from the context and tasks usually involved in deeper processing, while shallow strategies such as rote learning and memorization can be more effective for beginners, since such activities contain less material, which may distract a novice (Cohen & Aohek, 1981). Semantic mapping, as the technique applied in this study, is considered to require greater cognitive energy and deeper levels of processing, since it manipulates the relationship among words in deeper levels of cognition, as it does not deal with the analysis of the stimulus in the acoustic or visual levels but deals with the analysis of meaning at a deeper and more elaborative level. Therefore, it is more suitable for this level of language proficiency. In addition, advance learners of language have already developed that proficiency in the language and most of them might have already reached a high level of fluency. Accordingly, practicing such type of activities may not have as influential impact on them and tracking changes in fluency as a result of treatment could prove to be more difficult in comparison to the intermediate level, who have been assumed to be more sensitive to effective treatment and demonstrating the impacts of the treatment would be more easily done.

The total number of participants is thirty-three. They are students from four different classes who took part in the study and form the required two groups, with sixteen in the control group and seventeen in the experimental group.

3.4. Design

The current study utilized the between-subject multiple pre-test, immediate post-test, and delayed post-test control group design, in order to find out if training participants with already known lexical items had any significant impact on their utterance fluency and organization of the mental lexicon. The general design of the study is presented in the following diagram (Figure ***3.1***);

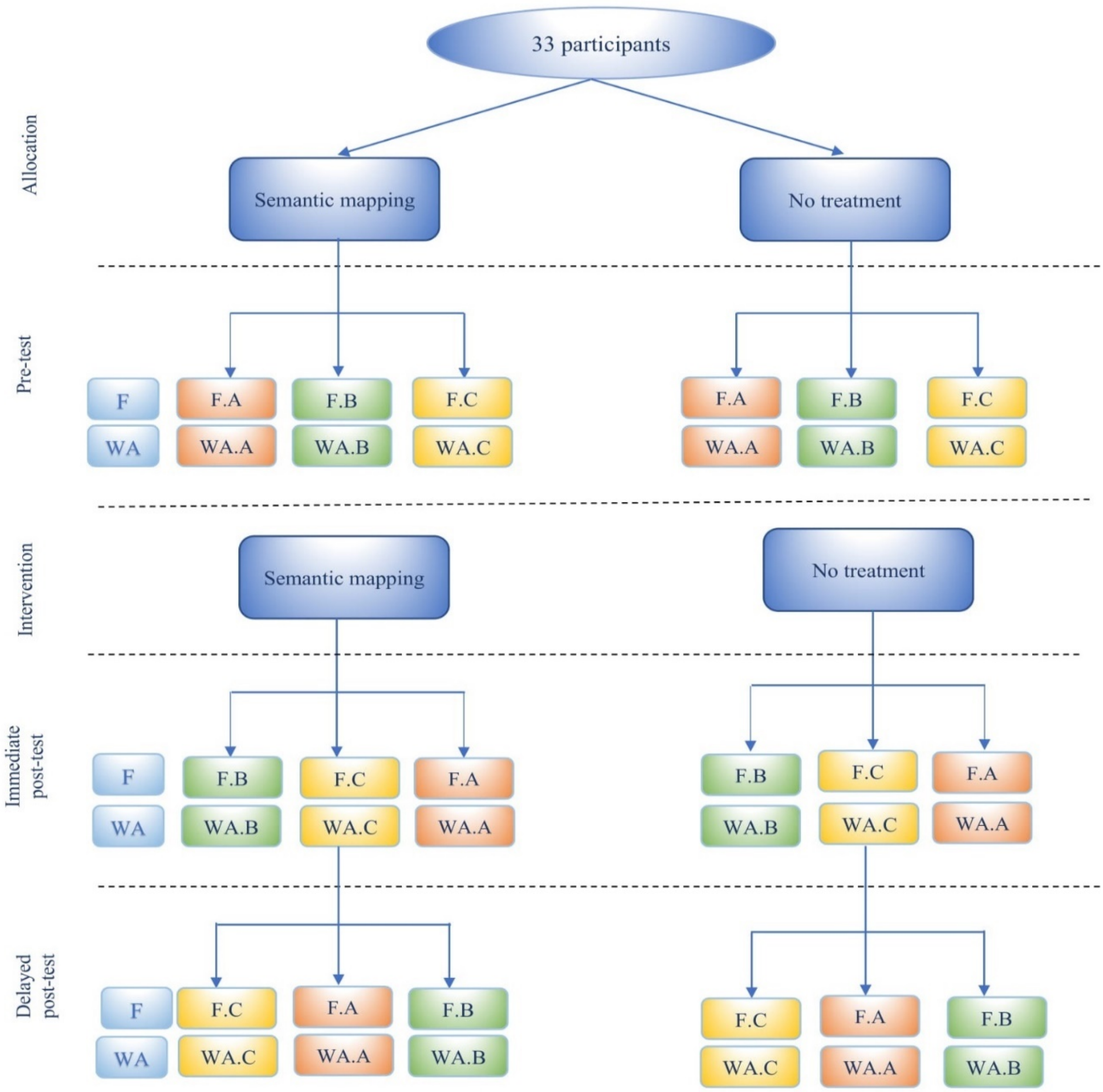


Figure 3.1 Design of the study

Note: F: Fluency test
WA: Word association test

The very first step in the process of designing this study was to opt for a method of sampling the participants. The convenience sampling strategy, as a type of non-probability sampling, was used in the process of selection of the participants. The main reason lied in the ease of access to the cohort. The participants were chosen because they were readily available and accessible to the researcher at the time. Although this type of sampling does not provide equal chances for the individuals to be chosen as the subjects of the study, it has many advantages, such as facilitation of data collection in a shorter period of time, relatively faster and more inexpensive way of accessing the sample, and it also saves time and money.

In order to choose a homogeneous group of participants in both groups of control and experimental, a C-test was administered in randomly selected classes. The test, developed by Daller and Phelan (2006), was adopted in order to make sure that the participants were at the same level of L2 proficiency. After administration of the placement test, a homogenized intermediate group of volunteer participants based on their performance was selected.

As it was nearly the end of the term of the classes, it was decided to wait for the beginning of the new term to randomly assign participants to different classes. Thirty-three individuals from the resulted homogenized group were selected and assigned into two different groups of sixteen and seventeen. The rest of students who were not going to take part in the research, either due to their score on proficiency test or their lack of interest as being research participants, were grouped together and placed in another classes, as the supervisor of the language school had already agreed for so doing.

The selected sample, which was the whole cohort, was randomly allocated into two different groups of one experimental and one control. Simple randomization was applied to allocate participants into different groups. This process was totally random, thus it helped to maintain the internal validity of the study, that was to establish causal interpretability (Mertens, 2010). By the use of this process of allocation, each participant had an equal chance of being placed in any of the groups, which in turn helped the researcher to ensure that the final results and the differences observed between the groups could be attributed to the process of treatment. An online research randomizer was made use of, which provided a quick way of generating random numbers to participants, in order to assign the participants into different groups. The advantage was that it was pretty fast and unpredictable, and also on average known and unknown variables were divided among groups on balance.

Consequently, in the present study, the participants, being randomly allocated to different conditions, were assigned to one experimental and one control groups:

- ✓ Semantic mapping group as the experimental group
- ✓ No treatment group as the control group

The experimental and control groups consisted of thirty-three participants in four different classes (eight participants in three classes and nine participants in one other class). All participants in two groups were asked to take part in both speaking tests and the productive word association tests. They took part in a single type of speaking task to measure their utterance fluency and also productive word association test of vocabulary to assess the organization of mental lexicon, more specifically the structure of network knowledge on specific stratum of lexicon. As this study benefited from multiple pre-tests and post-tests, the order of presentation of the tests might have greatly influenced the accuracy of the final results. That is, participants' performance might have changed due to practice, fatigue, boredom, or order effect, as they had opportunities to become familiar with the tests and took part in one more than once. Therefore, in order to avoid this, the tests were counterbalanced; that is, participants in each group were divided into subgroups, each receiving a version of the task on pre-test, another on immediate post-test, and finally the third version of the same task on delayed post-test. For so doing, three matched versions of each test were developed. In order to counterbalance the three versions of each test, Latin square counterbalancing was applied (Table 3.1).

Table 3.1 3*3 Latin square counterbalancing

Pre-test	Immediate Post-test	Delayed Post-test
A	B	C
B	C	A
C	A	B

In this procedure, each version of a test appeared precisely once in pre-test, immediate post-test, and after two-week time interval in delayed post-tests, and the tests were distributed as Table ***3.2*** shows;

Table 3.2 Timescale of the study in sessions

Groups	2nd session	3rd session	4th to 9th sessions	10th session	11th session	16th session	17th session
	Pre-tests		Intervention	Immediate post-tests		Delayed post-tests	
Experimental group	Fluency pre-test	Word association pre-test	Semantic Mapping	Immediate post-test on fluency	Immediate post-test on word association	Delayed post-test on fluency	Delayed post-test on word association
Control group	Fluency pre-test	Word association pre-test	No treatment	Immediate post-test on fluency	Immediate post-test on word association	Delayed post-test on fluency	Delayed post-test on word association

3.5. Instruments

The present study consisted of various instruments and tests; language background questionnaire, language proficiency test, speech fluency test measuring fluency of speech, and free productive word association test assessing the network knowledge and structuredness of the bilingual mental lexicon. Initially, participants in both experimental and control groups took part in language proficiency baseline. Next, they were asked to complete a language background questionnaire. Finally, three versions of picture narrative tasks and productive word association tests were administered in three phases of pre-tests, immediate post-tests, and delayed post-tests. Each of these instruments is described below.

3.5.1. background questionnaire

A paper-based questionnaire was designed in order to collect demographic information about the participants (see Appendix A). The main reason for gathering such data was that it would be easier to compare the obtained results of the study with other different research in this domain. The questionnaire would include information on participants' age, gender, occupation, hearing or speaking disabilities, educational background, English language proficiency level, purpose of studying English, hours spend learning English per day, years spent studying English, any studying or living abroad experience. It included fifteen questions and was given to the participants on the very first session of the class. They spent five to ten minutes at the end of the session to complete the questionnaire and the researcher collected them before they left the class.

3.5.2. Baseline proficiency test

In order to measure the participants' L2 proficiency and make sure if they were all at the same level of language proficiency, a C-test, developed by Daller and Phelan (2006), was employed in the current study (see Appendix B). This specific type of test was chosen among different alternatives to assess the level of language proficiency of the participants as C-test have proven to have a significant relationship with language proficiency tests such as TOEFL (e.g. Babaii & Ansary, 2001; Bachman, 1985; Hastings, 2002; Khodadady, 2014; Klein-Braley, 1996); hence, the results could reliably predict whether participants were at the same level of language proficiency. In addition, C-test could be completed and scored in a short period of time, which best fit the time constraints of the researcher's experimental paradigm.

The C-test is comprised of a series of short texts, in which every second half of every second word is deleted. Since the main intention was assessing the participants' general language proficiency, not only content words but also structure words were deleted. The test included six passages. Each of the passages had twenty gaps, which made a total of one hundred twenty gaps to be filled. The texts in the C-tests were originally taken from the website of number 10 Downing Street (www.number-10.gov.uk). They were about mainly dealing with general and cultural aspects of the UK. The participants were provided with twenty-five minutes to go through all the pages and fill the gaps with appropriate words. They were told that each gap, that was each for a single word, had one score and only the words, which were spelt correctly, would be considered as correct ones. In order to mark the tests, the missing words had to be spelt correctly, and only the ones contained in the original texts were accepted as correct.

3.5.3. Speaking task

For fulfilling the purpose of this study, speaking tasks were required to serve as pre-test, immediate post-test, and delayed post-tests, which would inspire the participants to talk for approximately three to four minutes uninterruptedly.

A single form of task, picture narrative, was chosen as the speaking test in all three sessions of testing. The reason for opting a single type of speaking task lies on the fact that different tasks can lead to different measures of fluency, since producing contents in varying degrees of structure place different cognitive loads (Derwing et al., 2004; Foster & Skehan, 1996; Skehan & Foster, 1999; Tavakoli & Foster, 2011; Tavakoli & Skehan, 2005). Therefore, in order to gain reliable results picture narrative tasks, as the only type, were adopted.

Picture narration is defined as "stories based on a sequenced set of picture prompts, which are given to participants in order to elicit language performance" (Tavakoli & Skehan, 2005, p. 248). The choice of this type of task rather than others lied on several reasons. First and foremost, picture narrative is the most frequently used task in different studies on measuring fluency; hence, its application facilitates comparability of results (Ahmadian & Tavakoli, 2011; Ahmadian, Tavakoli, & Dastjerdi, 2015; De Jong & Vercellotti, 2015; Derwing et al., 2004; Skehan, Foster, & Shum, 2016; Tavakoli, 2016; Thai & Boers, 2016). Second, for the purpose of this research a monologic task was preferred rather than a dialogic one; since due to its nature, monologic task is an easier and more convenient instrument for data collection. In addition, the speech sample collected is an uninterrupted production of reasonable length, which facilitates the process of analysis. Another positive point of narrative task is that performance of the participants is not affected by the interaction variables, such as taking turns, clarification request, interruptions, backchannels, etc. Thirdly, it was possible for the researcher to impose certain constraints and limits on participants regarding the content of the talk and lexical choices by the use of picture narrative. Consequently, "the degree of control and predictability of the outcome" in monologic mode are truly beneficial for research purposes (Tavakoli & Wright, 2020, p. 56).

As the characteristics of the narratives can influence task performance, specific task demands were identified and taken into account in the process of choosing the picture tasks. The degree of task structure was the most significant factor to be taken into consideration, since it has a direct effect on the formulator stage of speech processing system, and therefore mostly affecting fluency of speech (Skehan, 2009). According to Robinson (2001) and Skehan (2009), more complex tasks, which are looser in their structure, can lead to lower fluency of speech, while in contrary, tighter structure of tasks leads to decreased task difficulty, which enhances fluency.

According to Tavakoli and Skehan (2005), degree of structuredness depends on three factors including the logical relation among the elements of the story, timeline, and also a conventional beginning, middle, and end.

In order to comply with the mentioned requirements, picture narrative tasks were adopted from the study by De Jong and Vercellotti (2015). In their study on the impact of six similar picture narration prompts on fluency, complexity, accuracy, and lexical diversity of 25 upper-intermediate adult students, they investigated whether similar prompts elicit similar performance. All of the prompts in this study had a tight sequential structure, similar storyline complexity, and similar main characters and props. According to the results, the prompts were similar in both complexity and accuracy measures, while there were differences in fluency and lexical diversity, which was mainly discussed on the grounds of problems in lexical retrieval. Out of six picture narratives used in their study, three of them which did not result into significant changes on various fluency measures were adopted. The six-frame picture prompts in their study were selected from three sources; Heaton (1966), Mayer (1967), and Mayer and Mayer (1971). The chosen ones were *Bicycle*, *Race*, and *Frog*, which resulted into similar scores in all measures of complexity, fluency, and accuracy.

3.5.4. Free productive word association tests

The vocabulary test used in this study was free productive word association test. This specific type of test was opted for on the grounds that it is considered as an indication of organization of mental lexicon (e.g., Dóczi & Kormos, 2016; Henriksen, 1999; Meara, 1996; 2010), which was one the aims of this study to address. Word association test is compatible with the models of mental lexicon that uses the metaphor ‘network’ to describe the connections and relationships among words in the lexicon. In addition, this type of test is relatively fast and easy-to-administer way of assessing network knowledge (e.g., Cremer et al., 2011; Dóczi & Kormos, 2016; Meara, 2010; Schmitt, Ng, & Garras, 2011).

3.5.4.1. Developing the word association test

In order to develop the word association tests, several steps were required to be taken (see Section 4.3), which are discussed thoroughly in the next chapter (chapter 4) and only a summary of it is presented in this section. Initially, the first and most important step was choosing the most appropriate cue words, as the cue words can be considered as the determining factors in the type of responses that participants provide (Fitzpatrick, 2006). The cue words were chosen from students' textbooks from two specific units. The reason for choosing from the textbooks was the purpose of the study, which was building up on the current connections and not teaching new ones. The process of selection of words was random on the first step; that is, all words related to the two topics were selected. Afterwards, sixty words were chosen for the next step based on several criteria. First, phrasal verbs and compounds were omitted and not included. Secondly, the remaining lexical items were checked not to have multiple meanings. In addition, the lexical items were chosen from different frequency levels and from various grammatical word classes. The final step was examining the list of remaining words to ascertain that they were known to the participants (Fitzpatrick, 2006).

In order to ascertain whether the cue words chosen were known by the participants, a receptive Yes/No test of vocabulary was conducted (see Section 4.3.1.2). Participants who took part in the test did not participate in the main phase of the study, but they were in the same level of language proficiency as the main participants. The words, marked as known by all participants, were chosen as the cue words for the word association test.

The thirty selected words from the Yes/No vocabulary test, which were from various frequency levels, different word classes, and resulting into more associative links in comparison to others, were divided into three group of ten (see Section 4.3.2). Each group of ten words was decided to be used in one phase of the current study. As the three tests of pre-test, immediate post-test, and delayed post-tests were required to be of the same level of difficulty, same number of words from each word class, and words from similar frequency levels were assigned into each group. The words were put in random order via online randomizer.

3.5.4.2. Retrospective interviews

As it will be discussed further in more depth (see Section 4.2.2), one of the main methodological challenges in the way of interpretation of the results of word association tests, is understanding the link between the cue word and the responses provided. In order to solve this problem, it was decided to conduct short retrospective interviews with the participants to confirm their motivation for providing their responses. This method was applied in accordance with Fitzpatrick (2006) who suggested conducting retrospective interviews as a complementary stage in the process of data collection in word association tasks. The main purpose of conducting this type of interview was facilitating the process of assigning responses into coding categories with higher reliability and objectivity, as it helps in reducing the subjective judgment of the coder for responses which were more difficult to find the link to the cue word. Hence, no response would be wasted either by being assigned to a wrong category or marked as an unqualified response, bearing no association to the cue word.

3.5.5. Semantic mapping technique

The technique applied in the current study during the intervention sessions was semantic mapping technique. As it has already been discussed (see Section 2.6.1), semantic mapping points to "brainstorming associations which a word has and then diagramming the results" (Sökmen, 1997, p. 156).

Semantic mapping was opted as the main technique in the current study since it promotes deeper level of semantic processing and richer levels of encoding, rather than shallow analysis of visual or acoustic properties, which contributes to better learning of lexical items (Baddeley, 1990), which are considered to be the greatest obstacle in the way of producing fluent speech (Hilton, 2008). In addition, this technique serves as a kind of template organizing and structuring the knowledge and the connections among lexical items, which paves the way to building up on network knowledge for faster retrieval and longer-term retention of the items (e.g., Hague, 1987; Sökmen, 1997).

In the process of semantic mapping during the training sessions, the maps were developed and drawn on the white board by the researcher while produced in a collaborative manner between the students and the researcher (Nation & Newton, 1997). The underlying reasons for the collaborative manner of production was the fact that initially there might be times that participants were running out of ideas and the teacher can provide further support. Additionally, this keeps students on the track and does not let them deviate from the main topic of discussion.

Hence, the researcher encouraged the participants to produce associated words that could fit into the semantic map by giving them suggestions, hints, examples, or paraphrasing the words. Participants were asked to manipulate the lexical items, relate them to others and also their own previous experiences through a process of discussion, and finally justify their choices. In this way the connections are believed to be reinforced.

Two different semantic maps were developed by the researcher prior to the training sessions. The maps were on two different topics, *traveling* and *sports*. The maps were developed online via *mindomo.com* webpage by the researcher and printed out in colors for the final stage of practice (see Appendix C).

3.6. Procedure

The current study was an experimental one consisting of four key stages presented in Figure **3.2**;

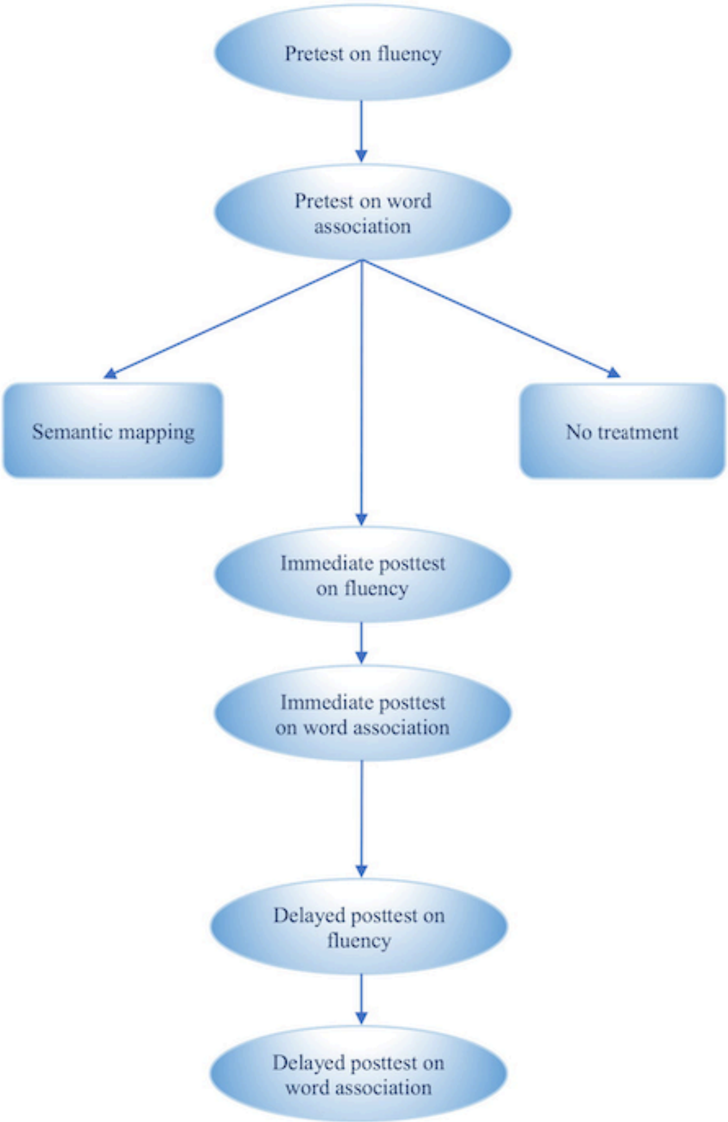


Figure 3.2 Experimental procedure

3.6.1. C-test

The very first and essential step in the current study was administrating the baseline proficiency test in order to ascertain that the participants were all at the same level of language proficiency (see Appendix B). Although all students had already taken part in the placement test of the language school and assigned into specific terms accordingly, a C-test was administered to confirm that no big differences in their level of language proficiency exist.

The C-test administered was developed by Daller and Phelan (2006). It was administered in the classroom environment at the very end of the term prior to the beginning of the next term of study. Students from five random classes, who were at the same language level according to the language school's system and volunteered to take part in the study, participated in the test. They received instructions on how to do the test.

In the testing session, the researcher provided a detailed instruction of the process, asking subjects to read through the texts carefully and filled the provided gaps with a word. The spelling of the word was considered to be very important and they were told that if they spelled the word incorrectly, they would not be scored for that word. The researcher collected and scored all the responses for further analysis.

From a group of forty-one students in total, thirty-three of them were selected and assigned randomly into different groups and classes. The very first session of the new term, which was the first session of the study, participants were asked to complete the background information questionnaire. They were given 10 minutes' time at the end of the class to complete the papers and return them to the researcher. It is worthy to note that all the sessions from the beginning to the end of the term was taught by the researcher to all four classes in order to maintain consistency across the sessions.

3.6.1. Pre-tests

There were two different types of pre-tests in the process of the study; a pre-test on L2 fluency of speech, which was followed by a productive word association test, each administered on the second and third sessions of the term, respectively.

3.6.1.1. Pre-test on fluency

The pre-test on L2 fluency of speech was administered on the second session of the term. Each class, for both control and experimental groups, were provided with same speaking tasks. There were three different versions of picture narrative task, that each participant talked about one version at each test time on fluency (pre-test, immediate post-test, and delayed post-test). The various versions were considered to be on the same level of difficulty and complexity as they were adopted from the study by De Jong and Vercellotti (2015) (see Appendix D).

The test took place two sessions before the commencement of the intervention period, and it was conducted by the researcher, who was also the teacher of the classes. The reason for administering the test by the same researcher was to reduce the negative effect of meeting somebody new as the test taker. Participant had the chance of knowing the researcher as their teacher which could have helped them to feel more comfortable and less anxious during the test times.

The pre-test on fluency was not administered during the regular class time. The main reason was that each participant was going to be tested individually and it might have taken long time to test all students; in addition, the test was required to be administered in a place with minimum background noise. Finally, as the researcher was the same teacher and test taker, it was not possible to spend one session on testing only. Thus, participants were invited to meet the researcher prior to their classes in a quiet room. Each participant was given a time slot, which was set in advance.

Each participant was welcomed warmly and thanked for taking part in the study in the first place. They were asked to take a seat, and then the researcher provided them with a detailed instruction on how the process was going to be (see Appendix E). Afterwards, they were presented with one six-frame picture, a question card, a pen, and a piece of paper. They had three minutes' time to prepare themselves and plan what they wanted to say and how to say it. This time was provided to help students generate enough semantic content to fill three to four minutes that they were supposed to talk uninterruptedly. The strategic planning time ranges in different studies from one minute (Mehnert, 1998; Wigglesworth, 1997) to one hour (Ellis, 1987), while in the present study the time provided was three minutes (De Jong & Vercellotti, 2015; Tavakoli, 2009; Tavakoli & Foster, 2011). Providing participants with time to involve them in thinking about the content and the language they needed for task performance had proven to be beneficial to their fluency of speech. Gilabert (2007) on his study on the impact of strategic planning on learners' fluency of forty-eight first and second-year university students with a lower intermediate proficiency level in English found out that planning results in greater fluency score. Same results have been obtained from many other studies on this same realm of research (Sangarun, 2005; P. Skehan & Foster, 2005; Tavakoli & Skehan, 2005; Yuan & Ellis, 2003).

While planning, they could take notes of the points they would have liked to talk about, but they were asked not to write complete sentences as it would have wasted their time. In addition, the question cards, presented along with the pictures, could help them in their narration and provide them with hints and clues on what to talk about in case they were out of ideas (see Appendix F). They had already been informed that the notes and question cards would be taken away prior to the actual oral production, while they could have the pictures in front of them throughout the narration time. The main reason for leaving the pictures with the participants during the talk was reducing the demand on short-term memory and avoiding memorization of the pictures (Fulcher, 2014).

After planning time, they were encouraged to start immediately. One-on-one recordings were conducted in a quiet room while they were telling the stories, in order to reduce the background noise to the minimum. They were advised that the story narration comprises three to four minutes, and they were notified on the time left twice by a card, once when there was just 1-minute left, and once when there was only thirty seconds left, in order to speed and wrap up. The researcher explained that they wouldn't have had a second chance to speak, and they were not re-recorded. After the test, the first one minutes of the recordings were imported into PRAAT software and transcribed by the researcher (see Section 3.7.1.2.2).

The same procedure was applied to all fluency testing sessions, pre-test and post-tests (immediate and delayed). The only difference was the pictures they were asked to talk about. The stories they were asked to talk about were different from one another, while the general concept was on what they had already been exposed to during intervention sessions, since the researcher was interested to find out if application of semantic mapping as an approach working towards the elaboration of network knowledge and strengthening the links between semantic and lexical representations, finally could lead to improvements in measures of fluency and changes to the structure of the lexicon.

3.6.1.2. Pre-test on word association

The word association pre-test was administered in both groups of experimental and control during the usual classroom sessions. Three different versions of the test were designed in order to reduce the practice effect, and on each testing occasion (pre-test, immediate post-test, delayed post-test) one of the versions was administered. The tests were of the same level of difficulty and included words that students had previously met (see Section 4.3.2).

The pre-test on lexical knowledge took place in the third session of the term, a session before the intervention commencement. The test was administered in the classroom context and took nine minutes only. The researcher handed out the answer sheet booklets to the participants. There were eleven pages in this booklet (see Appendix G). The very first page was an instruction page on how to do the test and an example to make it more obvious. The rest of the pages provided participants with spaces for their responses.

Ten cue words were presented via PowerPoint through a projector. Each slide of the PowerPoint consisted of a single cue word, which disappeared after twenty seconds. Various studies, asking for multiple responses, set different times for eliciting the associated responses based on the purpose of their studies. For example, Henriksen (2008) set the time for fifteen seconds for each cue words while asking for two responses. For the purpose of the current study, twenty seconds time interval was provided between presentation of the cue words as multiple responses, more than two or three, was favorable for the purpose of examining the strength of intralingual and interlingual connections. Meanwhile, participants were asked to write down any word which came to their mind once reading the word on the PowerPoint as instructed. They were invited to write as many responses as they could think of in the time provided (20s). The main reason for asking for more than a single response was leading the participants to produce denser network representation as a result of which it was possible to capture the distributional properties of the lexicon (e.g., Aitchison, 2012; De Deyne et al., 2012). In this way, it is feasible to make predictions about ways that the mental lexicon is organized (De Deyne et al., 2016).

In case no word came to their mind, they could just leave the page blank. As soon as the second word appeared on the power point slide, they had to flip the page. They were asked not to flip back to the previous page to add more words or change their mind after they had written their responses. In addition, they were told not to worry about the spelling mistakes as long as the word could convey the meaning.

The last stage of data collection on word association test is the short retrospective supplementary interview phase, asking participants to clarify the link between the cue and the response provided. As asking all participants about all the responses was a time-consuming procedure, it was decided to scan the responses immediately after collecting them and highlighting the ones that the researcher could see no direct or meaningful link. Besides, the other reason for not conducting the interview to all participants was the fact that the responses produced were not under their conscious control; hence, it was possible that they would provide justifications for some responses other than what they had on mind when providing responses or they were not able to provide any justification for their responses at all as it was produced entirely automatically. Consequently, those specific participants were approached and asked to explain what they had on mind when providing that response. Notes were taken for later reference.

3.6.2. Intervention sessions

The first intervention session started right after the pre-tests on fluency and vocabulary. The training sessions took six consecutive sessions, which started on the fourth session and finished on the ninth. Each session was one hour and thirty minutes in total, while forty-five minutes of each was assigned to the intervention. During the training sessions, no new and unfamiliar words were presented in any of the groups, since the aim of this research was not teaching new lexical items but helping students to retrieve the ones they had already acquired, develop lexical knowledge by elaborating their network knowledge, and building stronger connections and links between semantic representations, and finally being able to retrieve and produce words faster to improve their fluency.

As for the experimental group, the first session was devoted to the introduction of semantic mapping technique. So, initially participants were introduced to semantic maps and the process of semantic mapping while producing a small semantic map as a sample. In addition, they were provided with semantic map completed model to gain a better understanding of what the outcome might have looked like.

Only ten minutes was assigned to the process of familiarizing the students with the technique. The reason for this short time was that semantic mapping technique was not totally new to the participants. One of the supplementary resources of the American English File book series, which was the main textbook of this language school in General English Language Proficiency Courses, was Mind mapping to practice the new vocabulary at the end of each unit. Although students had never been asked to produce semantic maps and had only been asked to complete them with the new vocabulary they had learned, they had some familiarity with the general idea.

The sample map used for this part was on the vocabulary related to *appearance*. The steps taken were a little different from the main procedure of semantic mapping, as the main aim was only acquainting the participants with the general procedure. The steps taken were as follows:

1. Asking the participants to describe the teacher's appearance by posing the questions; 'what is the first thing you notice about me?' / 'what do I look like?';
2. Writing down all the concepts and ideas which were related to the concept of appearance on the side of the board. This was mostly a brainstorming stage;
3. Using different strategies, such as posing more questions, paraphrasing, providing them with the initials of the words or examples, etc. to elicit as many related words as possible;
4. Putting the words in the semantic map by the help of students while categorizing and linking them;
5. Providing them with the completed version of the map taken from the American English File book 1 while reminding them that the map could have been further developed in case of more time available (Figure *3.3*).

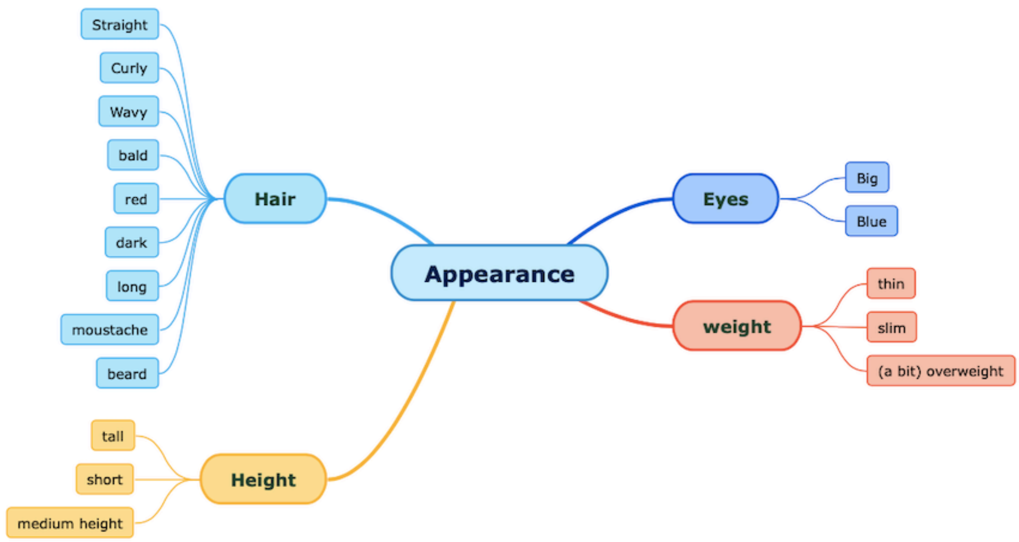


Figure 3.3 Model Semantic map

After the introduction stage, participants went through the process of semantic mapping technique. There were several stages in developing a semantic map, which were very close to the introduction phase and also the general process of semantic mapping, as discussed earlier (see Section 2.6.1).

In the current study, the first step was presenting a clear context in the form of a short reading text to the participants. The texts were used solely as a means of presenting the general topic, some related vocabulary, the associative relationships, and links among words, which could help the participants to activate and retrieve the lexical items they had already learnt in this domain (see Appendix H).

Based on the text, the teacher posed several questions and students were supposed to answer solely by looking at the pictures, reading the title of the text and guessing what was going on. Based on their answers, the teacher introduced the general topic by drawing a circle in the middle of the board and writing the main topic in it. Afterwards, students were invited to provide as many semantically and thematically related words as they knew while the teacher was writing them on the “parking lot” of concepts on the side of the board (Davies, 2011, p. 285), and not in the map. Similar to the introduction stage, this step was carried out by the help of the teacher providing various type of clues to further support this process and feed the participants with more ideas and concepts (step 3 in introduction phase). After brainstorming some related words, participants were asked to help in categorizing the suggested ones into semantically related categories. The main reason for this early categorization and drawing connections among lexical items was avoiding confusion, as it might have got difficult to see the relationship between large number of words at the end of the process and it would have been easier to begin by more general and bigger categories and further categorize them later.

Afterwards, the words were placed into the map under the agreed main topics and subtopics with the help of students. Later, the participants were asked to read the text silently while underlying the words they believed were related. Finally, they discussed what they had found from the text and tried to place them in the map with teacher. It is noteworthy to mention that the teacher made used of extensive class discussion all through the process of semantic mapping. Participants were asked to justify and explain the relation between lexical items they were suggesting while mapping, or in some cases, the teacher asked them some questions from a list she pre-prepared. For example, while mapping the words related to the subtopic *transportation* in the process of semantic mapping of the main topic *travelling*, participants were asked questions like ‘what type of transportation do you prefer? And what kind do you use more often? Have you ever heard a story of an airplane crash? When was the last time you travelled by train? Do you feel safe when you use subway? Have you ever gotten a parking ticket? How about a speeding ticket? etc.'. The goal for asking this type of questions was different from the clue questions asked earlier to elicit words. These questions were asked to invite students to discuss and explain the choices they made and to see the relationship, similarities, and differences among different words (Figure ***3.4***).

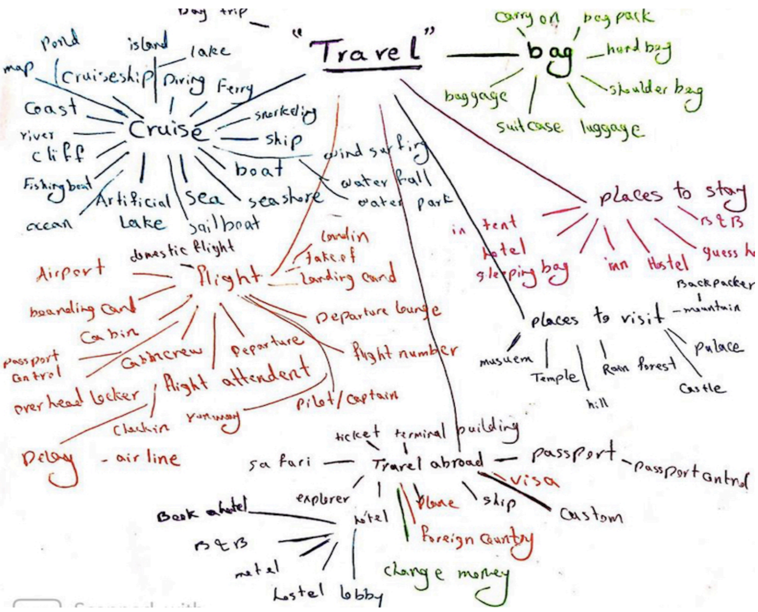


Figure 3.4 Semantic mapping on 'travelling'

One important point during the process of semantic mapping was to keep students on the track and not letting them to move the discussion on a way which was not advantageous for them. The teacher and students cooperated in building up the map; therefore, the associations were indicators of suggestions from the teacher, the students, and also the text. It was believed that using semantic maps in classroom would lead to having easier access to denser network of words and associations, introduced in lexical networks (Stoller & Grabe, 1993), which in turn could contribute to faster vocabulary activation and production.

The following intervention sessions were the same as the first, with different list of words. Same topic was discussed for the following two sessions. The only difference was an extra stage of reviewing the words from the previous sessions in the beginning of the class, and also including some of the words randomly in the main phase of semantic mapping. The reviewing stage which started on the second session of intervention period, was an oral stage before initiating next sematic map. The teacher posed a word on the board and asked students to suggest as many items as they could. In addition, at the end of the third session of the intervention, the model map created and printed by the teacher was presented to the participants. The printed map was used as a sample of many lexical items which could have been provided in the domain practiced. Also, after showing the map, the teacher asked volunteers to justify some of the links among the items presented as a final stage of reviewing. The third, fourth and fifth sessions of semantic mapping was the same while the second topic was introduced and practiced.

While the experimental group received training on the application of semantic mapping to practice the lexical items, the control group received no treatment. The material used in the control group was the same as the one presented in the experimental group; though, they didn't receive any training and practice with semantic mapping. The teacher began by asking students to look at the title and the pictures of the reading text and answer a few posed questions. Next, participants were invited to read the text as naturally as possible for comprehension. While reading, they were asked to underline the unknown words if coming across any. Next, the meaning of unknown or less familiar words were discussed. Finally, at the end of reading task and getting to know the topic, they were required to answer five comprehension questions with the aim of eliciting lexical items practiced. After they finished, they checked the answers in the class. The procedure remained the same on the following sessions.

3.6.3. Immediate and Delayed Post-tests

The tenth and eleventh sessions of the term were dedicated to immediate post-test on fluency and word association, respectively. Also, delayed post-test on fluency and vocabulary were administered on the sixteenth and seventeenth sessions. The same procedure as the pre-tests was to be iterated. The only difference was the tasks used. Participants who had taken a version of the test were going to take another version as presented in the Table ***3.3***;

Table 3.3 Order of the tests for pre-test, immediate post-test, and delayed post-test

	Pre-test			**Immediate post-test**			**Delayed post-test**		
Word association test	A	B	C	B	C	A	C	A	B
Fluency test	A	B	C	B	C	A	C	A	B

3.7. Analysis

For the purpose of this study, a purely quantitative approach was employed for data collection and analysis. In order to answer the research questions posed, an explanation of how the design meets the methodological requirements is presented in this section, which is followed by a detailed discussion of the process of coding, the measures adopted, and analysis of the dependent variables of fluency and organization of the mental lexicon presented and discussed in depth.

Two different instruments were used to measure and examine the impact of vocabulary practice training on ELF adult learners' utterance fluency of speech and organization of the lexicon: fluency tests and productive word association test of vocabulary. The research questions addressed were;

RQ1: To what extent does semantic mapping affect organization of bilingual mental lexicon as measured through free productive word association test?

RQ2: To what extent does semantic mapping affect L2 speech fluency measured through empirically established means of gauging L2 fluency?

RQ3: Is there a relationship between organization of bilingual mental lexicon and L2 speech fluency?

Research question 1: the impact of semantic mapping technique on the structure and organization of the lexicon,

The aim of the first research question was to investigate the impact of semantic mapping on the network knowledge; that is, to find out if application of semantic mapping technique in the experimental group has any impact on the organization and structure of the EFL learners' mental lexicon, in comparison to the control group. In so doing, the data collected during the pre-test (collected on the 3rd session) was initially compared to the one from the immediate post-test (collected on 11th session) and on a second phase to the delayed post-test (collected on 17th session).

In order to answer this research question, a mixed between - within ANOVA was employed to analyze the performance of the participants (see Section 5.2). Accordingly, there exist two independent variables of within and between subjects. Time is the independent variable relating the performance of the participants within each group, either the experimental or the control group separately, at pre-test, immediate post-test and finally at the delayed post-test, and explores if there were any changes in word association scores over the three times of testing. The second independent variable is group, which is regarded as the between subject variable, comparing the experimental and control groups in terms of their gains in scores of pre-test to immediate post-test and again gains from pre-test to delayed post-test. In order to answer this research question and measure the network knowledge as the dependent variable, overall word association score was used.

Research question 2: The impact of semantic mapping technique on L2 utterance fluency

In addition, the second research question aims to explore the impact of the same technique on EFL learners' L2 utterance fluency. In order to do this, the oral data, which was collected on the pre-test (collected on 2^{nd} session) session, immediate post-test (collected on 10^{th} session) and the delayed post-test (collected on 16^{th} session) were analyzed.

In order to answer the second research question, the data was analyzed using series of Wilcoxon and Mann-Whitney *U* tests (see Section 5.3). The same within and between-subjects variables of time and group were assessed. The variable time investigates whether there were any differences in various dimensions of utterance fluency over the three points of the testing time (pre-test, immediate post-test, delayed post-test). In addition, differences and gains regarding participants performances between the two conditions was assessed, in terms of differences in gains from pre-test to immediate post-test and also pre-test to delayed post-test.

Table 3.4 Independent variables for word association test and L2 utterance fluency

	Type of variable	Number of levels
Time	Within subject	3 levels (pre-test, immediate post-test, and delayed post-test performances)
Group	Between subject	2 levels (experimental group and the control group)

Research question 3: The relationship between oral fluency and network knowledge

The last research question investigates the extent to which the overall word association scores correlate with various measures of L2 utterance fluency of EFL learners. In order to answer this last question, the statistical correlation between the experimental and control groups on the scores of overall word association test and four measures of utterance fluency, each collected on pre-test, immediate post-test, and delayed post-test, were assessed (see Section 5.4).

3.7.1. Pre-analysis procedures

This section discusses the pre-analysis procedures of data obtained from the word association tasks and also the picture narratives. In both cases, it is explained how the data was first recorded, which is followed by a section on coding and operationalizing the measures. Finally, the inter-coder reliability for both coding systems is discussed.

3.7.1.1. Pre-analysis procedures of word association test

One of the most challenging steps to take in application of the productive word association tests is how to categorize and measure the responses collected, which is heavily dependent on the aim of the research and what the researcher wishes to elicit. The following section is on how the lexical associated responses collected from L2 language learners were treated, categorized, and analyzed.

3.7.1.1.1. Treatment of the responses

In order to analyze the data collected by the application of word association test, several steps had to be taken in advance. In the very first place, all participants' handwritten responses were transcribed into an Excel file. While transcribing the responses, spelling mistakes were corrected, as the participants had been advised not to think about spelling as far as the meaning could have been conveyed, and the intention was clear.

While transcribing the responses, it was noted that some of the responses provided were no single words, instead few multi-words, a combination of two, three words, or even a phrase, were presented. Thus, in order to maintain objectivity and consistency in categorization process, the lexical category of each of the phrases, the heads, was decided to be used, as they bear the core meaning of the phrase (Sag & Wasow, 1999). Accordingly, multi-word responses were shortened to head words. For example, the response ‘very fresh air’ to the cue word ‘jungle’ was shortened to ‘air’, also a response like ‘lose important things’ to the cue word ‘cheat’ was shortened to ‘lose’. In cases that the response was a compound, it was the stem which determined the semantic category.

The third adjustment was lemmatizing the response in order to deal with both derivational and inflectionals. One of the main reasons for choosing to lemmatize the lexical items and not to continue with single word forms is the way lexical items are processed. Aitchison (2012) argues that lexical items are stored and retrieved in the basic form of a lemma and affixes are added during online processing, with the exception of irregular forms which are stored as single items.

Different studies took different approaches in lemmatizing the list of words. Wolter (2002), adapting Meara and Fitzpatrick (2000), lemmatize the responses in level 2 and 3 affixes of the classification system proposed by Bauer and Nation (1993). This system of lemmatization considers lexical items with the same base and inflectional suffixes (plural, third person singular present tense, past tense, past participle, -ing, comparative, superlative, possessive) and derivational affixes (-able, -er, -ish, -less, -ly, -ness, -th, -y, non-, un-) to be members of the same word family. For example, responses of ‘murderer’ and ‘murdering’ to the cue word ‘kill’ would be reduced to ‘murderer’ while ‘murderous’ would not. Fitzpatrick et al. (2013) and Playfoot et al. (2016) did not lemmatize the derivational affixes but the inflectional ones, and they opt for only level 2 of Bauer and Nation (1993), that was words having same base and inflections were considered to be in the same word family. Accordingly, as a response to the stimulus word *win*, associated responses such as *loses*, *losing* were considered to belong to the same word family, while the response *loser* belonged to the level 3 and not regarded as being in the same word family, and treated as a different response. The current study lemmatized the responses following Fitzpatrick et al (2013) and Playfoot et al (2016).

After transcription of the responses, the responses needed to be coded for further analysis.

3.7.1.1.2. Coding of word association data

In order to code the responses collected from the productive word association test, various procedures were taken into account. Initially, all the responses for each single cue word collected from participants in both groups was compiled in a single Excel file prior to the actual coding, resulting into thirty different spreadsheets. In this way, it was possible to make sure that the coder, in this case the researcher, was not affected by "the respondent's previous behavior patterns, or by the popularity of a particular response across the sample" (Fitzpatrick at al., 2013, p. 17). The duplicate responses were decided to be removed in the first phase of coding in consistency with the mentioned reason; however, they were added for the second part of coding to trace the chains or two step associated responses.

After having all responses to each cue in a file, the following step was assigning the responses to pre-determined categories. Hence prior to coding, a spreadsheet was prepared with fifteen different categories. Each of the columns was assigned to one single code and one for the total score, as presented in Figure *3.5*.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
1	CYCLE	No Association				loosely associated			Closely associated			Canonicality		Frequency		Total
2		Repetition	Example	Proper name	Ragbag	Chaining	Two step As:	Form relatec	Position base	Semantically	Position/Sem	Canonical	Non-canonic	High Frequer	Low Frequency	
3	Bicycle									•		•		•		3
4	Fell down									•			•		•	5
5	Park									•		•		•		3
6	Hang out					•										1
7	Biology				•											0
8	Sport									•			•		•	5
9	Biking									•			•		•	5
10	Ride										•	•			•	4

Figure 3.5 Word association coding spreadsheet

The coding scheme in this study was adapted from Henriksen (2008), while changes were applied to meet the goals of this specific study. The details of the coding scheme will be presented and explained thoroughly in the next chapter (see Section 4.4). In addition, figure 4.2 provides an overview of the coding scheme applied in the current study. The coding was organized in three main groups and ten subgroup of response types;

- ✓ Not associated responses which is composed of repetition, proper names, examples, and ragbags
- ✓ Loosely associated responses which is composed of chains, form-related, and two step associated responses.

✓ Closely associated responses which is composed of semantic based, position-based, and semantic/position-based responses.

The very first main category was assigned to 'No association' main group, which was further classified into four subcategories of *repetition, examples, proper names, and ragbags* (columns B, C, D, E in the excel file). The four subcategory headings with examples are summarized in Table ***3.5***;

Table 3.5 Subcategories used to classify 'no association' group of responses in word association test

Category	Subcategory	Definition	Cue	Response
No association	Repetition	Cue and response are identical words	Lake	Lake
	Example	Response is an example of the cue (in the form of a proper name)	Lake	Khazar
	Proper name	Reponses are names individual person, places, or organization	Amateur	Golzar
	Ragbag	Responses have no relationship to the cue word	Jungle	She

The next main group is 'loosely associated' responses, which is composed of three subcategories of *chaining*, *two-step association,* and *form-related* responses (columns F, G, H), presented in Table ***3.6***;

Table 3.6 Subcategories used to classify 'loosely associated' group of responses in word association test

Category	Subcategory	Definition	Cue	Response
Loosely associated	Chaining	Responses are related to the previous response rather than the cue word	Cycle	Hang out (via *park*)
	Two-step association	Responses are associated with the cue through a formal link (phonological/orthographic) with another word	Lake	Early (via *Late*)
	Form-related	Reponses are associated to the cue word due to phonological/orthographic similarity	Hurt	Heard
		Responses are the same as the cue with addition, deletion, or changing of either inflectional or derivational affixes	Enjoy	Enjoying

The last group of responses are closely associated ones. This group is classified into *semantically related, position-based*, and *semantic/position-based* responses. In order to assign responses precisely and accurately to each of these categories, the following subcategories are adopted from Fitzpatrick et al. (2013); though, no further discussion of these subcategories are required (Table *3.7*). Although the true nature of some associations might be obscured by not including the detailed identification and description of the subcategories into account, this level of discussion and analysis might be more fruitful concentrating on the degree of advancement of the lexicon instead. What we can do, however, is look in more detail at the more general subcategories in order to identify and differentiate between various levels of structuredness of the lexicon.

Table 3.7 Subcategories used to classify 'closely associated' group of responses in word association test (Fitzpatrick, et al., 2013)

Category	Subcategory	Definition	Cue	Response
Semantically related	Synonyms	Cue and response are synonymous in some situations	Establish Fraction	Build portion
	Lexical set	Cue and response share a hyponym, or one word in the pair is an example of the other; includes antonyms	Bean Bean	Vegetable Pea
	Other conceptual	Cue and response are related in meaning, but are not synonyms or in the same lexical set	Sin Nurse	Prayer Illness
Position-based	Cue-response collocation	Cue is followed by the response in common usage; includes compound nouns	Rock Fence	Roll Post
	Response-cue collocation	Cue is preceded by the response in common usage; includes compound nouns	Fence Plug	Electric Spark
	Cue-response and response-cue collocation	Cue could precede or follow the response in a common phrase	Rock	Hard
Semantic/position based	Lexical set and cue-response collocation		Bread Gold	Cheese Silver
	Lexical set and response-cue collocation		Nurse Cheese	Doctor Bread
	Synonym and cue-response collocation		Torch	Light
	Synonym and response-cue collocation		Shove	push

While assigning the responses to categories, the data collected through retrospective interviews was made use of to arrive at a precise categorization of responses.

After coding all the responses, a particular score was then awarded to each single valid and qualified response, by consulting the table of scores presented below (Table ***3.8***). Consequently, after assigning the responses into the categories, individual response profiles were created for each participant for further analysis.

Table 3.8 Scores awarded to different categories of response types

Response type		**Examples from L2 with the stimulus word 'Jungle'**	**Score**
No association	Repetition	Jungle	0
	Example	Golestan Jungle	0
	Proper name	Amazon	0
	Ragbag	She	0
Loosely associated	Chaining	(Jungle...Lion...) King	1
	Two step association	(Joggle) Ball	1
	Form-related	Jungles	1
Closely associated	High frequency non-canonical, but semantically/position -based related	Bird	2
	High frequency canonical semantically/position -based related	Tree	3
	Low frequency canonical semantically/position -based related	Forest	4
	Low frequency non-canonical, but semantically/position -based related	Safari	5

3.7.1.1.3. Inter-coder reliability of word association test

Once the coding is finalized by the researcher as the first coder and total scores have been assigned, a sample of roughly 10% of the data was given to a second coder to check for the reliability of the scores.

As it has already been mentioned (see Section 3.7.1.1), all responses to each single cue word, collected during word association tests from the participants, were compiled in single Excel files; therefore, there were thirty different spreadsheets each of which was assigned to responses collected during pre-test, immediate post-test, and delayed post-test to each cue word. The second coder was presented with three spreadsheets with responses already lemmatized and cleaned; that is, the responses were treated as explained (see Section 3.7.1.1.1); however, the duplicates were not removed as it was not possible to find the chains or two-step associated responses. Besides, the same coding scheme (figure 4.2) and instructions on how to assign different responses to different categories were presented to the second coder.

Once the coding had been completed by the second coder, the number of responses in each of the categories were compared, revealing that 89.67% of response items had been assigned to the same category as in the first coding by the researcher. A further 10.33% of the classifications were agreed after a short discussion and close reference to the definitions. The main reason for the difference was the responses in the closely associated category. For example, the first rater coded the response *expensive* to the cue word *palace* to semantic-based responses while the second coder decided to choose semantic/position-based response category. In order to come to an agreement, it was decided to refer to collocation display of British National Corpus, in case of any disagreement on closely associated group of responses.

3.7.1.2. Pre-analysis procedures of the fluency measures

In order to collect data to measure dimensions of fluency of speech, several pre-analysis steps are required to be taken, and procedures are needed to be taken into account. The very first step was audio recording the elicited speech (3.7.1.2.1), followed by the detailed description of the way the oral data was annotated and coded using a computer software, discussed in section 3.7.1.2.2. Finally, the results of the internal consistency reliability test are reported in section 3.7.1.2.3.

3.7.1.2.1. Audio recording

In order to analyze and evaluate fluency of speech, all sessions of pre-test, immediate post-test, and delayed post-tests were audio recorded via two different devices, both with GarageBand application on MacBook laptop and Marantz MP3 recorder. The reason for using two different recording devices simultaneously was the problems faced during pilot phase of the study. In the pilot phase, some of the data was missed either due to not being recorded or not being saved after the recording was finished. The recording procedure was straightforward as there was only a single participant in a quiet room each time to be recorded; hence, no complications were faced.

3.7.1.2.2. Coding of narratives - PRAAT

In order to analyze the data, the software PRAAT 6.0.39 (Boersma & Weenink, 2018) was used. PRAAT is a free computer software, which provides a great opportunity to visualize and analyze the speech segments in great detail. The analysis procedure is summed up in following steps.

In order to maintain higher degree of reliability in the coding process, the extracted recordings were first coded automatically using De Jong and Wempe (2009) PRAAT script with a cut-off point of 0.25 seconds for the silent pauses and sounding parts, and then coded by hand manually. The main reasons for double coding and rechecking the automatic ones manually was two factors. Firstly, although the data was collected in a quiet room, there was still some level of background noise from the students in the corridors or other classes, which made the detection of silences difficult in an automatic manner. In addition, it was necessary to code the data considering the particular fluency measures believed to reflect cognitive fluency. The automatic programming script was limited regarding the measures it can calculate. As an example, it is possible to calculate the measure ‘syllable run’, while calculating ‘syllable duration’, which requires a pure measure of phonation time, minus both filled and unfilled pauses, was not feasible. The reason is that the PRAAT script does not identify the length and number of filled pauses and measure them as phonation time. Accordingly, in order to ascertain that the data had been coded accurately and reliably, it was further coded manually.

In order to code and analyze the data, the very first step taken was converting the audio data collected into a format which was compatible with PRAAT. GarageBand recorded the audios in MP3 format while it allowed the recordings to be exported to a WAV file, which could be read by PRAAT. Then, each recording was opened in PRAAT individually and the first sixty seconds of the talk was extracted from each of the recorded speech. Pauses, either filled or unfilled, at the very beginning of the recording, before the first syllable, and after the last syllable were discarded. Accordingly, the recording for analysis in PRAAT all began with lexical items (e.g., 'One') or fillers (e.g., 'well').

The next step was annotating the files by studying the spectrogram. The process of annotation was carried out by careful listening to the short stretches of the audio file along with inspecting the spectrogram. Annotating the files in PRAAT is done in a TextGrid file, which is finally saved separately from the sound file. TextGrid annotations in the current study were composed of nine tiers, each of which marked specific points within the recorded sound file.

Initially, the automatic PRAAT script was run, segmenting the recording to articulation phases (*sounding*) and pauses (*silent*). Then, the recording was listened again to make any necessary changes, the most important of which was identifying filled pauses. In this tier, not only significant adjustments were applied but also distinctions were made between different types of pauses. In this process, the breath sounds, and total silences were both identified and coded as unvoiced pauses, while hesitation particles such as [əː], [əːm] and nonverbal fillers like "uh," "ah," "um," and "mmm" were categorized as voiced pauses. Hence, the second tier was assigned to unvoiced pauses (*u*), voiced pauses (*v*), and sounding boundaries (*sounding*). It is noteworthy to mention that although all silences higher or lower than 250ms total were coded, only the ones which totaled 250ms or longer were only calculated as pauses in the final analysis.

The third tier was assigned to transcription of the audio files orthographically. Therefore, tier 3 was assigned to the lexical items produced in *sounding* segments and the pauses. The main reason for transcribing the data was double checking the number of syllables reported by the automatic script and ascertaining that filled pauses or background noise was not counted mistakenly as syllables.

The fourth tier was devoted to pause location; that is, if a pause occurs within the clause boundaries (*e*) or between the boundaries (*m*), besides the number of syllables produced. In this tier, pause intervals were marked as either voiced, unvoiced, or composite, while their location was specified and annotated. This was accomplished by careful listening of the audio files and inspecting the orthographic transcription in tier three.

Next tier was assigned to the number of syllables, which was counted by listening to the audio file, inspecting the spectrogram, and finally by consulting the transcriptions. The reason for listening to the audio file, counting the number of pitches on the spectrogram along with consulting the transcriptions was the many discrepancies observed between the expected and pronounced number of syllables. This case is more highlighted in the case of Iranian speakers as a result of the syllable structure of Persian language (CV, CVC, CVCC) in comparison to English. Persian speakers have difficulty pronouncing consonant clusters in English, which is due to the fact that in Farsi a word cannot begin with two consonants. Hence, "initial consonant clusters in English words are broken up by vowel epenthesis" (Shademan, 2002, p.1). Accordingly, there might be an extra syllable in comparison to English number of syllables for the same word. For example, the word *star* pronounced as */'sta:/* with a single syllable in English, is pronounced as */'əsta:/* in Farsi, as a two-syllable word. De Jong ad Perfetti (2011) posited that "where there was doubt about the number of syllables pronounced (e.g., 'every' can be pronounced as /ɛvri/ or /ɛvəri/), the original recording was consulted" (p. 545). Accordingly, it was advised that the best option to address the differences was consulting the original recording (De Jong & Perfetti, 2011). Hence, the number of raw syllables in each run was manually entered into the TextGrid.

The number of counted syllables in tier 4 was equal to the raw syllables with inclusion of the repairs, reformulations, and repetitions, while the next tier was assigned to the pruned number of syllables after reduction of the syllables assigned to repairs. Accordingly, tier 6 accommodated number of syllables of repairs, while tier 7 defined the boundaries of repairs. Finally, all pauses were annotated in two different tiers of voice (tier 8) and unvoiced (tier 9) simultaneously, for the ease of automatic running of the scripts.

Consequently, the final TextGrid included manually transcribed data with information on number and duration of voiced and unvoiced pauses, number of raw and pruned syllables as well as repairs in nine tiers as shown in the Figure ***3.6***.

Finally, the automatic script, developed by Handley (2017), was adapted to elicit the desired fluency measures automatically. Although the original script was designed incorporating almost all measures of utterance fluency, changes were required before running it to meet the requirements of the current study.

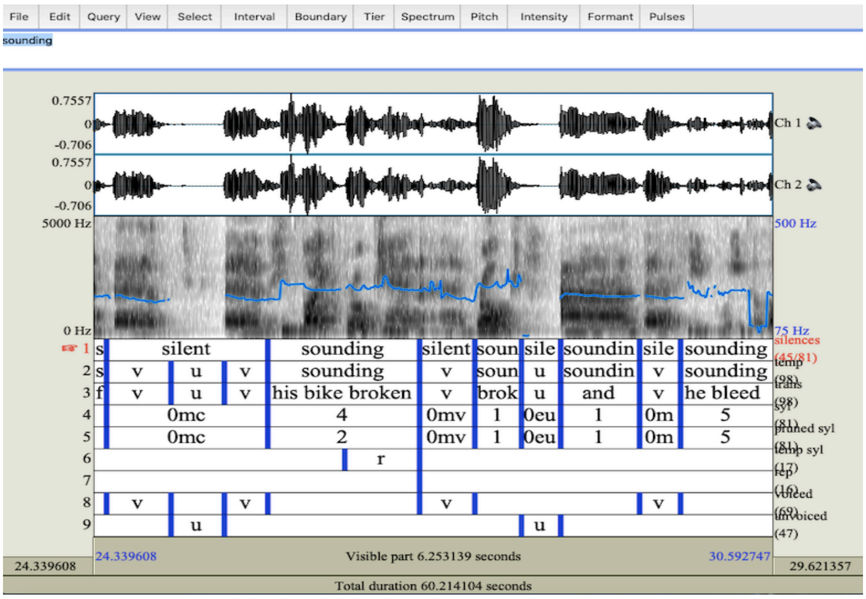


Figure 3.6 Annotated speech in PRAAT

3.7.1.2.3. Inter-coder reliability of narratives

Once the speech data was coded, 10% of the sample was re-examined by an independent trained researcher using PRAAT. The second rater coded ten samples of speech, which meant roughly ten minutes of the total duration of the speech samples collected. As the data was first coded automatically and then adjusted manually by the first rater, in this case the researcher, high correlation between the raters was expected. The results of Spearman's correlation revealed a high reliability score (r = .98) for syllable run, phonation run (r = .93), measure of syllable duration (r = .95), and finally silent mean duration of pauses (r = .93).

Table 3.9 Interrater reliability for four dimensions of utterance fluency

	Syllable run (rater 2)	*Phonation run* (rater 2)	*Syllable duration* (rater 2)	*Silent pause duration* (rater 2)
Syllable run (rater 1)	.98	-	-	-
Phonation run (rater 1)	-	.93	-	-
Syllable duration (rater 1)	-	-	.95	-
Silent pause duration (rater 1)	-	-	-	.93

3.7.2. Dependent variables

In order to answer the posed research questions, two groups of dependent variables are used in the current study. The first, which is the measure of overall word association score, was applied to assess organization of mental lexicon, that is more specifically, the structuredness of the network knowledge. In addition, a number of dependent variables were used to measure various dimensions of L2 utterance fluency. An overview of the measures is presented in Table ***3.10***.

Table 3.10 Overview of the dependent variables

Dimension	Measure	Calculation
Organization of the mental lexicon	Overall word association score	Response type score multiplied by the measure of semantic strength (semantically related responses divided by the number of types multiplied by 100)
Fluency	Syllable run	Total number of syllables produced divided by total number of silent pauses >250ms
	Phonation run	Total time spent speaking divided by total number of silent pauses>250ms
	Syllable duration	Total speaking time (phonation time minus all pauses >250) divided by total number of syllables
	Mean duration of silent pauses	Total duration of silent pauses divided by the total number of silent pauses

In this section each variable and the way it was calculated will be presented thoroughly. In addition, the main reasons for their inclusion in the current study will be justified.

3.7.2.1. Organization of the mental lexicon – Overall word association score

Final point to be taken into consideration regarding analyzing the data collected by use of free productive word association test was allocation of a single overall word association score (OWAS) representing the overall network quality of the mental lexicon, so that developments in the lexical network can be envisaged.

In order to assign a score to the productive word association tests representing overall network quality of the mental lexicon, the following steps, adopted from Henriksen (2008), were taken. All steps from 1 to 7 is adapted from the original scoring scheme; however, step 8 is different due to the purpose of the current study, which in turn changes the final results on step 9 as well;

1. For each participant, eliminate duplicates in responses to each of the stimulus words (After lemmatization of the responses, duplicates appeared in the list of responses for some cue word. For example, the associated responses *loses* and *losing* to the cue word *win* were counted as one single response and one was deleted as a duplicate, since they were lemmatized and reduced to *lose*)
2. Calculate total score for each individual by looking at the scoring scheme
3. Calculate the number of valid item responses, which are defined as the ones which are either strongly or loosely associated to the cue word (all response types except for the unqualified ones)
4. Calculate Response Type Score (R.T.S) by dividing total score each of the participants received by the number of valid items provided

R.T.S = Total score/number of valid items

5. For each participant eliminate duplicates in the whole set of responses they gave, and calculate unique types score across stimuli
6. Calculate number of unique types by counting number of unique responses provided after deletion of the duplicates from the whole data set
7. Calculate number of semantically related responses (semantically related responses are the four categories of highly frequent canonicals, highly frequent non-canonicals, low frequency canonicals, and low frequency noncanonical response, that is all responses except two groups of not associated and loosely associated ones)
8. Calculating Semantic Strength (SS) which was dividing semantically related responses for each participant on each phase of word association test over the unique types (step 7 over step 5)

S.S = semantically related responses/number of types * 100

9. Multiple Response type score by Semantic Strength (SS) to come to Overall Word Association Score (OWAS)

$$OWAS = R.T.S * S.S$$

So, each participant was assigned a response type score and semantic strength score, which then checked for semantic density of the network by multiplying them together. Although application of this scoring procedure was a great way to compare between levels of productive network knowledge, it is not a good representation of the details of the response characteristics as they were obscured. It was very difficult to distinguish between two participants both achieving a same score, as one might come to a few low-frequency non-canonical associations having a higher score, but the other to many high-frequency non-canonicals and gain the same score as the previous participant. In order to solve this problem, it was decided not only to judge and decide on participants network knowledge based on the overall score of the word association test, but also analyze type of the associations supplied in more depth by providing descriptive statistics on the nature and type of associations provided (see Section 5.2.5).

3.7.2.2. Various dimensions of fluency

In light of the fact that fluency is a fairly difficult concept to define as a result of its multifacetedly, its measurement is therefore, not an exception. Narrowing down the exhaustive list of potential candidates with the aim of building up a comprehensive picture of the changes in fluency, four different measures were chosen. The selected measures, reflecting core L2 fluency features and construct, are "mean run length defined either as syllable or phonation run between silent pauses, and two other measures were mean duration of syllables and silent pauses" (Segalowitz et al., 2018, p. 113). Although not all these measures belong to one of the single categories of speed, breakdown, or repair (Skehan, 2003), they are believed to be among the most reliable indicators of cognitive fluency, as it will be discussed in this section.

These measures belong to both specific and composite measures of oral fluency. Initially, in order to link fluency measures to particular stages of speech production mechanism, two pure and specific measures touching upon only one aspect of fluency were chosen (syllable duration and mean duration of silent pauses). The biggest advantage of such measures can be argued to be avoiding the multicollinearity problem associated with fluency measures. Besides, it is possible to attribute them to particular cognitive processes involved in speech production (Lambert, Kormos, & Minn, 2017). While the specific measures result in a fine-grained analysis, two composite measures were also opted for, namely syllable run and phonation run. For example, syllable run which is one of the most widely applied measures correlating highly with both perceived and cognitive fluency, encompasses all three aspects of fluency (Lambert, et al., 2017; Witton-Davies, 2014). Including such measures provide opportunities for concentrating on different aspects of fluency, capturing a more global and holistic picture of oral fluency.

3.7.2.2.1. Holistic measures of mean length of run

For the purpose of the current study, two holistic measures of mean length of run are assessed in terms of both syllable run and phonation run.

Syllable run has been agreed to be one of the best indicators of fluency (e.g., Ginther et al., 2010; Kahng, 2014; Segalowitz & Freed, 2004; Towell et al., 1996), and has been proposed as a measure of automatization (Towell, et al., 1996). It can be considered as a "density measure that appears to represent both syntactic well-formedness and vocabulary" (Ginther at al., 2010). This measure has been operationalized in different ways in literature in an attempt to represent the density of content between pauses. Most commonly, it is measured in terms of average number of syllables produced between pauses (e.g. Derwing et al., 2004; Kahng, 2014; Segalowitz, et al., 2018), while in some other studies, it has also been defined as the number of words (e.g. Hilton, 2004) or phonemes (e.g. Cucchiarini et al., 2000) in a run between pauses.

In the current study, syllable run was calculated by dividing the total number of syllables uttered in the speech sample by the total number of silent pauses.

The second holistic measure applied in the current study, phonation run, is an indication of length of phonation burst between interruptions of two silent pauses. This measure is one of the core measures of utterance fluency, representing the fluidity of talk (Segalowitz, et al., 2018). Although phonation run is another measure of mean length of run similar to syllable run, it is not a measure of density of an utterance. Instead, this measure scrutinizes the length of uninterrupted phonation, which makes it equivalent to the number of pauses per minute (De Jong, 2018).

A point worth to be taken into account is the role of pausing and how it is defined, which has a direct impact on both measures. In this case, a short cut-off point of pause length can contribute to higher number of fluent runs while setting the threshold higher tends to reduce the number of runs. For the purpose of this study, the most reliable cut-off point can be 250ms. Setting this minimum threshold leads to measures of fluency which are highly representative of the knowledge of lexis. De Jong and Bosker (2013) discussed this issue more fully. In addition, setting this cut-off point helps to make the results of the study comparable to other studies in literature (e.g., De Jong et al., 2012; Goldman-Eisler, 1968; Kahng, 2014, 2020; Lennon, 1990; Towell et al., 1996).

The main argument in here is that both measures of syllable run and phonation run are affected to a great extent by how often a speaker needs to pause, considering that fewer pauses can contribute to longer stretches of speech. However, syllable run can also be a representation of articulation rate at the same time. In the case of syllable run, “the faster the speech, the more syllables in any stretch of speech” (De Jong, 2018, p. 241). Accordingly, it is a measure combining articulation rate, as a pure measure of speed, with silent pauses per minute. In order to assess if the increase in syllable run is not attributed to faster speech only, the duration of talk between pauses have been examined. Accordingly, the length of phonation burst between interruptions of silent pauses was decided to be taken into account.

In the current study, phonation run was calculated by dividing the phonation time by the total number of silent pauses.

3.7.2.2.2. Pure measures of fluency

In order to track the changes in the fluency of speech, two other measures of syllable duration (measure of speed) and mean duration of silent pauses (measure of breakdown fluency) were also taken into account and assessed. These measures have the main advantage of revealing more about the underlying cognitive processes in the process of formulation and production of speech (Tavakoli, Nakatsuhara, & Hunter, 2020). The reasoning behind the choice of these specific measures is explained below.

Speed fluency

The current study measures speed fluency in terms of syllable duration (inversion of the articulation rate). The main reason of inclusion of this measure is the fact that syllable duration can be considered as the core and pure measure of speech, best capturing the speed of utterance production, by excluding pausing (>250ms) and the duration of time as time divisor (De Jong et al., 2013; 2015; Segalowitz et al., 2018).

Syllable duration has attracted lots of attention recently as a measure of dysfluency, capturing a single aspect of fluency features at a time. It manifests the size of the syllables making up the flow of speech and reflects its fluidity (Segalowitz et al., 2018). In addition, this measure has the advantage of reflecting L2 specific fluency (De Jong et al., 2015; Kahng, 2020). Hence, it can be applied as a pure measure picturing the underlying cognitive processes, controlling for individual differences and various speaking styles. The findings regarding this measure strongly supports its high correlation with both linguistic knowledge (vocabulary knowledge) and linguistic processing skills (lexical retrieval) (e.g. De Jong et al., 2013; 2015; Ginther et al., 2010; Towell et al., 1996).

However, an important point to take into consideration is the existence of elongated syllables in calculation of phonation time as the dividend in the measurement of syllable duration. Elongated syllables provide a gauge for the extent to which speakers buy time to produce an utterance instead of voiced pausing. To the extent of my knowledge, Lambert et al. (2017) was the only researcher who identified filled pauses as "non-lexical fillers such as *er* and *uhm* and elongations of sounds (drawls)" (p. 177). In the current study, elongated syllables were included in the calculation of the phonation time since the cut-off point to identify them is not clear in literature.

In the current study, syllable duration was calculated by dividing the phonation time (minus all pauses>250ms) by the total number of raw syllables.

Breakdown fluency

Breakdown fluency can be measured in a variety of possible ways as there exists a pool of promising candidates, such as phonation time, silent/filled pause ratio, silent/filled pause rate, mean silent/filled pause duration.

The first important factor to decide upon in relation to measuring breakdown fluency concerns the character and quality of pausing; that is, whether silent or filled pauses are to be studied. For the purpose of this study, it was decided to opt for silent pauses. The choice of silent pauses over filled pauses lies on the grounds that silent pauses are applied mostly to indicate that the speaker deals with speech processing difficulties, while filled pauses serve different purposes. Filled pauses either have communicative and pragmatic purposes and are applied to delay message transfer and hold the floor (Maclay & Osgood, 1959) or they are elements "whereby the speaker, momentarily unable or unwilling to produce the required word or phrase, gives audible evidence that he is engaged in speech-productive labor" (Goffman, 1981, p. 293). Accordingly, although both types of pauses might be originated due to same problems in speech processing system (Kormos, 2006a), as filled pauses might be produced to serve purposes other than signaling problems formulating and producing utterances, silent ones were opted for, as they significantly contribute to determining L2 break down fluency.

As it has already been mentioned, the range of breakdown fluency choices, which are frequently examined in fluency research, is wide. Despite this wide range, there is no agreement on breakdown measures which are the best representatives of the underlying cognitive processes. Following Segalowitz et al. (2018), mean duration of silent pauses is a core fluency feature in relation to the breakdown fluency, which reflects "a common underlying fluency construct" of speech fluidity with other measures discussed in this section (p. 100). This measure has been found to have correlation with the lexical knowledge (Hilton, 2008) and the related processing skill of lexical retrieval (De Jong, et al. 2013; Kahng, 2020). Hence, the length of silent pauses, as the pure feature of breakdown fluency, is measured for the purpose of the current study.

In the current study, mean duration of silent pauses was calculated by dividing the total duration of silent pauses by the total number of silent pauses (>250ms).

In addition to mean duration of silent pauses, it can be argued that frequency of silent pauses per minute is another good indicator of underlying cognitive processes as it strongly correlates with linguistic knowledge (De Jong & Bosker, 2013; De Jong et al., 2013; Hilton, 2014) and is most likely to be a representative of L2-specfic breakdown fluency (De Jong et al., 2015). This measure has already been included in the study in terms of the inversion of phonation run, as a composite measure of mean length of run.

3.8. Ethical considerations

The very first step in conducting any educational research is consideration of a range of ethical issues (Fraenkel, Wallen, & Hyun, 2012). In order to make sure whether this research was ethically acceptable, numerous concerns were taken into consideration prior, during, and after conduction it. The very first step in identification of any ethical issues was discussing the research topic and proposal with the supervisor and TAP member (thesis advisory panel). The next step was completing the ethical audit form and submitting it to the Department of Education at University of York to obtain their permission to conduct the experiment (see Appendix I). The form can provide thorough information for the Education Ethics Committee about the participants in the study, if you are working with any of the vulnerable groups such as children or people with disabilities, or any kind of intervention to the normal procedure. As this educational research did not involve any of the above, the University of York granted approval to conduct the research.

Two different types of informed consent forms had been developed as well: one for the head teacher at the language school the data was collected from (see Appendix J) and the other for the adult students taking part in the research as participants (see Appendix K). The procedures, what is expected from the participants, data collection instruments, the withdrawal procedure, its anonymity and confidentiality were explained in the consent forms. Consent form for the head teacher was emailed to him before travelling to Iran to do the data collection, as his permission was the first to take. Afterwards, consent forms were given to the participants of the study and they were informed of the goals and aims of the research. Besides, participants were informed on how they can withdraw from the study until two weeks after the data collection was finished and if they decided to withdraw, there would be no penalty. Finally, they were assured that the data collected would remain confidential, and after two weeks all data will be anonymous.

3.9. Pilot

Running a pilot study, as a kind of feasibility and scoping opportunity prior to conducting the main study, is vital, and is considered an essential part for completion of any successful study design (Teijlingen & Hundley, 2002). The main goal of conducting a pilot phase prior to the main study was to determine the suitability, practicality, and feasibility of the intervention procedure, activities, and tests for the language proficiency level in question and identify and resolve any issues coming up while implementing them.

There were several issues that needed careful consideration in the piloting phase in order to determine whether the experiment would run well before administering it, since if anything had gone wrong meanwhile running the experiment, a lot of time and energy would be wasted, and the obtained results wouldn't be an indicator of what the study aimed to gain. This phase of the study involved applying the semantic mapping technique, as well as assessing the materials and tests of fluency and vocabulary. The purpose of piloting was fivefold as outlined below:

- ✓ To choose a proficiency baseline test and determine the practicality of either administering elicited imitation test or C-test;
- ✓ To determine the feasibility of employing the semantic mapping technique in the classroom and its suitability for the intermediate Iranian learners of English;
- ✓ To develop appropriate cue words for the word association test;
- ✓ To test the logistics of carrying out both picture narrative and word association tasks, including the appropriateness of the materials for the participants, along with any issues associated with the procedure of administering them; and
- ✓ To determine whether application of semantic mapping technique results in improved measures of utterance fluency in a small-scale study;

The pilot phase of the study was conducted in the spring term prior to the main study. As the number of sessions needed to conduct the pilot was ten sessions, the researcher decided to teach the classes herself for the whole duration of the term. Eight participants from an intact class were recruited on voluntarily basis for the pilot. The pilot study was conducted over a five-week period of time during the regular classroom hours. The term began on Saturday 24th February 2018 and finished on Saturday 2nd May and the pilot study was conducted from March the 10th to May 11th. Table ***3.11*** presents the pilot study timetable and plan.

Table 3.11 Plan for the pilot phase

Session	Activity	Number of participants
5th session	Elicited imitation test	8
6th session	C-test	8
7th session	Fluency test (Pre-test)	8
8th session	Vocabulary test (Pre-test)	8
9th – 12th sessions	Semantic mapping training	8
13th session	Fluency test (Post-test)	6
14th session	Vocabulary test (Post-test)	6

The results of the pilot phase indicate that semantic mapping technique was appropriate for the learners in intermediate level of language proficiency and that timing, although tight, allowed for enough practice and discussion time within the language school's timetable. Verbal feedback from the participants indicated that they enjoyed the process of practicing vocabulary via the application of semantic maps. It was evident that the participants were highly motivated and involved in the process, and they strived to contribute to the discussions and suggesting words they had already known. Accordingly, it could be judged that the application of the semantic mapping technique was thoroughly feasible and practical.

In addition, the narrative tasks adopted to measure the fluency of speech from the study conducted by De Jong and Vercellotti (2015) revealed to be appropriate for the intermediate level of language proficiency and could elicit the anticipated and desired sample of speech. The pictures were comprehensible to the participants and they were not difficult to follow and interpret in terms of the main line of stories. Finally, the stories were interesting enough for the participants to motivate them to talk. The result of the analysis and findings pointed out clear gains for the participants in three measures of pruned speech rate, syllable duration, and syllable run; however, no significant changes was found in measures of frequency of pauses, mean duration of pauses, and number of repairs.

A number of additional observations were also made that were relevant for the final design and procedure of the current study:

First, the fifth and sixth sessions of the term were dedicated to the piloting of the baseline proficiency tests, and finally it was decided to assess participants' level of language proficiency using C-test by Daller and Phelan (2006) rather than the elicited imitation test by Tracy-Ventura, McManus, Norris, and Ortega (2014). Although elicited imitation test is a good choice for measuring the participants' oral proficiency, as it checks the participants speaking ability in a very reliable and valid way (Chaudron, Craig, Prior, Matthew, & Kozok, 2005; Graham, Lonsdale, Kennington, Johnson, & McGhee, 2008), concern was raised regarding its feasibility and practicality in the classroom context. Although elicited imitation test checked the oral language proficiency of the subject, it was not economical in terms of time. It took long to administer the test to the participants as it needed to be conducted to one participant at a time and took longer time to transcribe and score the results. Furthermore, the scoring procedure was an endeavor to reach objectivity while it was still subjective, especially in comparison to C-test, which was totally objective.

Second, in order to assess the quality of the network knowledge; that is, the structuredness of the lexicon, word association tests were required to be piloted. The piloting of this phase of the study was comprised of two different stages of piloting the preliminary list of cue words, followed by piloting the administration of the test.

Initially, the selected cue words for the word association test were required to be piloted in order to ascertain that appropriate list had been selected. The choice of this type of test was grounded on the fact that the responses provided to the cue words reflect the strongest, or most salient, or most automatic links (e.g., Fitzpatrick, 2006; Fitzpatrick & Izura, 2011; Playfoot et al., 2016); hence, the process of selection of cue words was of significance importance. The preliminary list of cue words in the pilot phase was selected from two units of *travelling* and *crime* from students' previous chapters of their course books. The selection of these two topics was based on the picture narrative stories for the fluency tests. In order to make sure that participants in the intermediate level of proficiency knew all words used in testing sessions, a Yes/No test as a measure of receptive vocabulary knowledge was designed. Accordingly, ninety-two words from various frequency ranges and grammatical classes were selected from the two topics (the process is discussed in detail in chapter 4.3). In addition, in order to control the guessing element, eleven 'pseudowords', which were phonologically and orthographically possible words, were added to the list.

A group of thirty intermediate students, who did not take part in the main phase of the study, were asked to participate in the word association test. They were asked to put a cross mark next to the words they believed they knew the meaning and skip the ones they didn't know or were not sure about. The results clearly showed that the words in *crime* category were more difficult for the participants to recognize. Out of forty-six words on *crime*, only five of them were checked by all thirty participants. Accordingly, the results revealed that participants did not know many of the words related to the *crime* topic. Since the resulted words were going to serve as cue words in the word association test, the researcher decided to omit the *crime* topic as it appeared to be difficult and many of the words in this group were left unmarked. Finally, the topic *crime* was replaced by *sports,* and a new list of cue lexical items was developed and tested. The process of the design of the test is discussed thoroughly in the next chapter (see Chapter 4).

Next, the process of administration of the word association test was piloted to assess the practicality and feasibility of the process. The tests were administered on the eighth and fourteenth sessions of the term. Two comparable word association tests were designed, each of which consisted of ten prompt words. The cue words were presented on the board, and the teacher had a stopwatch to take the time since they had twenty seconds for each word to write down as many associated words, but the process proved to be difficult and problematic. By using a stopwatch and writing the cue words one after the other on the board, the researcher sometimes spent more than twenty seconds on each word and students sometimes had more time for some words and less for others. Therefore, for the main study it was decided to use timed PowerPoint slides and a projector.

Finally, while piloting the picture narratives, it became clear that additional recording measures were required to avoid loss of data through human error. Particularly, some recordings were lost as the researcher had accidentally stopped the recording without saving the file or on another occasion, the battery of the laptop recording the voices died and some files were lost. In order to make sure that no data would be lost in the main phase of the study, it was decided to record the voices using both an MP3 recorder and a laptop.

3.10. Summary

In this chapter, an outline of the design and procedures of the current study were explained thoroughly. This chapter was divided into ten sections. The first section presented a general overview of the current study, restating the research questions, which was followed by a description of a detail account of the research strategy, nomothetic approach, a variable-based approach to answering the proposed research questions. In the third section, comprehensive information on participants was provided. The participants in this study were a group of thirty-three adult language learners in Iran. They were in intermediate level of language proficiency, speaking Farsi as their native language and learning English as the foreign language.

Afterwards, the design of the study, between-subject multiple pre-test, immediate post-test, delayed post-test control group design was discussed in detail. In the process of explaining the design of the study, the method of sampling of the participants, convenient sampling strategy, was presented. Following this, the procedure of the allocation of the participants into different groups of experimental and control was presented. Next, as the study benefited from multiple pre-test and post-tests, a counterbalancing strategy was employed in which one out of three versions of the picture narrative test and word association test was administered each time in pre-tests, immediate post-tests, and delayed post-tests. The main purpose for this systematic variation in presentation of the tests was enhancing the interval validity by controlling for the confounding effect of the order of presentation of the material.

In the fifth section, the various number of instruments applied in the current study were presented and explained thoroughly; language background questionnaire, baseline language proficiency test, speech fluency test, and word association test assessing the network knowledge and structuredness of the bilingual mental lexicon. Subsequently, a step-by-step discussion of the experimental procedure of the study was provided, consisting of four key stages of pre-test on fluency, pre-test on word association, intervention, post-test on fluency and the word association, and finally delayed post-test on fluency and word association.

This was followed by a clarification of the general data analysis procedure and pre-analysis of the picture narrative and word association test. In this section, it was explained how the data was recorded, coded, and operationalized. Finally, the inter-coder reliability for both coding systems is discussed.

Following this, the two groups of dependent variables were presented along with the rationale for the choices made and explanations of how each measure was calculated; that is the overall word association score, assessing the organization of the mental lexicon, and four dependent variables of syllable run, phonation run, syllable duration, and mean length of silent pauses, used to measure various dimensions of L2 utterance fluency. Finally, the chapter ends presenting discussing the ethical considerations and brief explanation of the pilot phase of the study, in light of which details of the current study was set out.

4. Chapter 4: Word association research

4.1. Overview

As it has already been discussed in chapter 3, the principal purpose of the current study is to explore the extent to which semantic mapping technique has an impact on the organization of mental lexicon, and also on L2 oral fluency of speech. To achieve this aim, the following research questions have been posed;

RQ1: To what extent does semantic mapping affect organization of bilingual mental lexicon as measured through free productive word association?

RQ2: To what extent does semantic mapping affect L2 speech fluency measured through empirically established means of gauging L2 fluency?

RQ3: Is there a relationship between organization of bilingual mental lexicon and L2 speech fluency?

Accordingly, one of the purposes of this experimental study was to assess the impact of semantic mapping technique on the organization of the mental lexicon. The technique applied in this study to examine the structure of the bilingual mental lexicon is word association task, as it has been proven that "word associations provide a privileged route" into understating the organization of the lexicon, allowing its structural features to be studied closely (De Deyne & Storms, 2015, p. 412). Hence, in order to reach reliable and accurate results, paving the way to better understanding of the changes in the organizational structure of the mental lexicon and answering the first research question, word association tests are required to be designed.

This chapter is divided into three main sections. The first part discusses word association behavior literature in addition to the main sources of concern which needs to be addressed while designing word association tasks (see Section 4.2). The following section discusses, in detail, the steps taken to develop the productive word association test used in this study (see Section 4.3), and finally presents the coding scheme, adapted from Henriksen (2008), and changes applied to categorize type of responses and measure the overall productive network knowledge score (see Section 4.4).

4.2. Literature review: Word association behavior

“Word association research is deceptively simple” (Fitzpatrick & Thwaites, 2020, p. 237). In its basic form, it involves informants suggesting words associated with cues, without thinking, analyzing, or censoring their responses. Hence, the responses can be considered to be the first words coming to the mind right after hearing or reading the cue words. Due to its relative ease and speed of administration, word association research has attracted lots of attention in the field as it has been argued to reveal valuable information about the organization, structure, processing, and development of the mental lexicon, and big number of research and hypothesis have been formed on this ground.

Word association has been applied since late nineteenth century as a main technique of exploring the conceptual and lexical connections among items in the two levels of knowledge representation of mental lexicon. In early 1960 and 1970, word association behavior was widely applied in the studies of L1 mental lexicon (e.g. Lippman, 1971; McNeill, 1966; Nelson, 1977), and based on the theoretical background and research methodological formed, foundation for L2 word association research was laid down. Accordingly, a major shift happened to studies on bilingualism, and numerous studies concentrate on structure and processing of the mental lexicon in non-native speakers (Meara, 1992; Wilks & Meara, 2002; Wilks, Meara, & Wolter, 2005). Fitzpatrick (2012) posed that this technique leads into valuable data in the field of both L1 and L2 lexical research due to two main reasons. First and foremost, the data collected by the application of word association task is produced totally naturally, not under artificial experimental conditions. Secondly, the nature of the task is quite spontaneous as the first words coming to mind are asked for; hence, it can provide information about the most salient and strongest links among words, and consequently reveals valuable information about the mental lexicon.

However, as a result of the apparent simplicity of word association technique, a wealth of potential analytic approaches has been emerged, which in turn has led to inconsistencies into the gained results in terms of identifying behavioral patterns. Nevertheless, ‘hope’ is a powerful force (Schmitt, 1998a, 2010; Wolter, 2002; Zareva & Wolter, 2012). Schmitt (2010) proposed that researchers are “still waiting for a breakthrough in methodology which can unlock undoubted potential" (p. 248). Accordingly, research in this domain keep refining various protocols and analysis still hoping for consistent and meaningful patterns.

In an attempt to come to an understanding of why word association research fails to provide informative and consistent findings on the structure of the bilingual mental lexicon and finally generate solutions to address such obstacles, three overriding areas of concern have been suggested to be taken into consideration on the way of pursuing the potentials of word association tasks (Fitzpatrick, 2006). The three domains are the stimulus words selected for elicitation of the associated responses, the link between the cue word and the responses and the degree to which it is clear and correctly interpreted, and finally categorization of the responses. Each of these areas of concern are discussed thoroughly in the following sections to provide an overview for the general trends taken into designing the word association task in the current study.

4.2.1. Choice of the stimulus words

Stimulus words or cue words in word association tests are the words presented to informants and associated responses are asked for. The choice of these words is of utmost importance as they strongly influence or even determine the responses provided by the respondents (Fitzpatrick, 2006). For example, cue words from different frequency levels can elicit different responses. More frequent words tend to prompt more predictable responses than words with lower frequency (e.g. De Groot, 1989; Meara, 1983). Also cue words from different word classes tend to elicit certain types of associations, as nouns prompting nouns, adjectives eliciting nouns, and verbs prompting verbs (e.g. De Deyne & Storms, 2008; Deese, 1962; Entwisle, 1966; Nissen & Henriksen, 2006; Sökmen, 1993).

The first important point in the process of selection of the cue words can be considered to be the source from which the cues are chosen from. As both Fitzpatrick (2006) and Meara (2009) have noted, many researchers turned into using different well-established native speaker norm lists of cue words in their studies. One of the most widely applied word lists is Kent-Rosanoff (1910). Although such a list has the advantage of offering high degree of comparability between studies, its application can be attacked on the grounds that the cues in the list are all very highly frequent items as a result of which specific group of responses can only be elicited. In addition, the cues are all either nouns or verbs and no other grammatical category can be spotted in the words

(Meara, 1983; Schmitt, 1998a; Wolter, 2001). Fitzpatrick (2006) echoed this notion adding that the sample of responses elicited via this list cannot be considered as a representative sample based on which judgments to be made. Setting the advantages against the drawbacks and points of weaknesses, it appears that drawing on such lists should be considered with caution. A potential solution to the problems posed, by the application of such word lists, is developing and compiling lists of cue words depending on the nature of the study being conducted. In this way, it is possible to take various factors into consideration while choosing the cues, such as different levels of frequency and grammatical classes.

Another factor which contributes greatly to the validity of application of word association tests depends on the number of cue words included in a study. The number of cues varies hugely from one study to the other, mostly depending on the purpose of the research. Studies using Kent-Rosanoff (1910) most often include all 100 words, as the responses were well-established and ready for the analysis. Fitzpatrick (2007) also prepared a list of 100 cue words from the Academic Word List to investigate the homogeneity and consistency of adult native speakers to the stimulus words. In another study of her, she selects 60 words from the same Academic Word List to compare native and non-native responses using her novel proposed categorization of the responses. However, Kruse, Pankhurst, and Smith (1987) used only 12 of the cues and Ruke-Dravina (1971) only four. The only point of caution is application of high frequency words as cues, due to the fact that they have very strong tendency to constantly elicit primary and even secondary associated responses (e.g., *white* associated with *black*), and deciding and extrapolating behavioral patterns on the basis of such responses, especially when the number of cues are little, should be drawn with care (Fitzpatrick, 2006).

The last point in this section on selection of cue words deals with the number of associated responses asked for and elicited to each cue word, which has led to long disagreement among the researchers. It has been claimed that multiple responses provide better opportunities for investigating the structural and organizational properties of the lexicon (De Deyne, Leopold, Navarro, Perfors, & Storms, 2012; Schmitt, 1998), while many researcher tend to ask for single responses to avoid contaminating the reliability of associative data (Fitzpatrick, 2007; Fitzpatrick & Izura, 2011; Namei, 2004). For instance, in a study by Schmitt (1998), following Schmitt and Meara (1997), respondents were only asked for the first three responses springing in their mind, while Fitzpatrick and Munby (2013) asked for up to 12 responses. The main problem in such type of studies, asking for multiple responses, might be the chaining associations, which instead of providing association for the cue word, they come up to associations for the previous response.

Consequently, considering the mentioned issues, influencing the findings and results of word association studies, compilation of the list of stimulus words should be done with extensive care and in principled manner (Fitzpatrick, 2006).

4.2.2. Understanding the links between cue words and responses

One of the main methodological problems in interpreting the results of word association tests is understanding and interpreting the connection between the cue word and the associated responses provided by the informants, as there are cases that the link might need to be clear. The responses which are difficult to interpret in the context of second language might occur due to various reasons, such as incomplete understanding of the cue word, providing associated responses through a cognate/false cognate in L1, or random retrieval of any response in order to fill the gap (Meara, 1983). In order to remove this obstacle, various solutions have been put forward. Wolter (2001) checked the appropriateness of the responses by asking two native speakers to judge. Some other researchers also suggest categorizing some responses as 'unclear cases' (Yokokawa, Yabuuchi, Kadota, Nakanishi, & Noro, 2002).

Fitzpatrick (2006) suggested that the most appropriate solution, in order to minimize the degree of subjectivity in interpreting the responses, can be retrospective interviews regarding the responses given, and asking the informants to explain the link they come up to. Although the process is time consuming, it might be the only way to make sure that responses are neither assigned into wrong categories, nor to 'unclear cases' (Fitzpatrick, 2006). Although retrospective interviews sound like a promising solution, its excessive use might pose challenges at the same time. The problem is the assumption that the responses provided by informants are produced under conscious control though it is possible that the processes are automatic, and the justifications provided after the word association task would reveal little about the link in mind.

4.2.3. Categorization of the responses

The third point of concern in word association research, which has also proven to be among most challenging decisions to be made in the application of word association tasks, is how to categorize and measure the associated responses collected. In order to remove this obstacle, which might result into inconsistencies and difficulties in interpretation of the data, two major approaches have been applied: stereotyping/norm-list and categorization system approaches.

The first approach examines the responses by making decisions on how similar they are to a list of norms/stereotypes, collected and compiled from the responses provided by native speakers to the cue words. This approach stemmed from the assumption that native speakers' responses are homogenous and systematic; accordingly, the norm lists can be used as a means of comparing the responses collected with them and find out how native like the responses are.

This approach was traditionally used in L1 as a way to distinguish the points of deviation from the norms. Traditionally, it was applied in studying mental health, psychology, and personality traits. The first list of norms published was by Kent & Rosanoff (1910). This list was widely applied in many L2 linguistic and lexical studies (over 920 citations reported by Google Scholar). In spite of its popularity, two main factors have been neglected. Firstly, the purpose for which this list of norms was designed. Kent & Rosanoff cue words was chosen for the purpose of diagnosing mental health problems and psychiatric screening by Sommer (1910), via the application of word association technique. The aim of this word list was totally different from the purpose of most linguistic studies on, either L1 or L2, mental lexicon. More importantly, the process of selection of the cue words was not based on any systematic attempt and justification, but it was done in an unprincipled manner. The only justification for using the selected cue words was that half of the words were translation equivalents from Sommer's German original, and they tried to avoid using ones which were "liable to call up personal experiences" (1910, p. 4). Finally, the responses, which have been used unquestionably, were collected from non-natives and high school children, and ironically called normal standard.

The other source of problem with the use of word lists originates from the use of the response derived from them, which are regarded as norms. It has been argued that these norm lists promote the belief that there exists a "normal standard", homogenous and consistent way for native speakers to respond to words in word association tasks (Fitzpatrick, 2007, p. 321). This assumption is misleading as the responses are mostly reflection of sustained preferences, both on an individual and a generational level, in the way of responding word association cues (Fitzpatrick, 2006, 2007, 2009 ; Fitzpatrick et al., 2015).

Consequently, many researchers have decided to collect their own list of norms on the basis of the purpose of their studies and design lists which can clearly and reliably be a representation of the characteristics of the target population (Fitzpatrick et al., 2015; Hirsh & Tree., 2001; Miller & Chapman., 1983). Developing such a norm list supports the results and interpretations made on its basis, as the differences observed between the responses can be assigned to the differences across the populations. Therefore, although using already designed and published list of norms can be tempting, they should be applied with caution and many variables are required to be taken into consideration to assure that their application fits the requirements of the research. In general, it is worth noting that all such studies using norm list missed the fact that type of associations is a critical issue to be taken into consideration as well (Fitzpatrick & Munby, 2013).

Concentrating on type of associated responses paves the way to the second approach, which is categorization of the type of associations. Recently, many studies have focused on type of associations in order to find out about the organization and structure of the mental lexicon. This approach originates from L1 lexical development and stems on the idea that L2 development of the lexicon mirrors L1 patterns. By comparing the two lexicons' growth, via analysing the shifts in the type of responses, researchers tried to measure L2 language proficiency (Fitzpatrick, 2012; Fitzpatrick & Munby, 2013).

Generally, the conventional and popular three-way categorization of response types is used in many studies, which dates back to Saussure (1916). There are three different types of categories accordingly; syntagmatic, paradigmatic, and clang. Fitzpatrick (2012) defined the three categories as follows; paradigmatic responses are the ones which can replace the cue word without making any changes to the structure of the sentence; that is, such replacement causes no syntactic violation. They can in turn be categorized into four groups of coordinates (*cat* eliciting *dog*), subordinates (*cat* eliciting *Persian*), superordinates (*cat* eliciting *animal*), and synonyms (*cat* eliciting *pussy*). Next, syntagmatic responses cannot replace the cue word but can be found along with the cue word in the same structure, like collocations (*cat* eliciting *domestic* or *meow*). Finally, clangs can neither replace the cue words, nor can be found in the same phrase, but they are orthographically or phonologically similar to the cue words (*cat* eliciting *CAD*).

Response type categorization has been applied in the domain of L2 vocabulary research for quite a long time, while recently other categories has been added and new frameworks have been developed based on the nature of the research conducted. Main reasons underlying these changes and addition of more categories is two-folded (Fitzpatrick, 2006). First, three-way categorization imposes artificial limitations and constraints on investigating type of responses, as it is difficult to define the content of categorization; hence, it is open to interpretations. Some associated responses might fit into more than one category (Fitzpatrick, 2006, 2007; Wolter, 2001); hence, it calls for subjective interpretation. The only way to get over this limitation is going back to the respondents and ask them about the reasons they have chosen specific response to be able to put it in one of the categories, and it is not an easy task to be done as it can be very time consuming and also it is not an easy to reach the respondents after the study is conducted (Meara, 1983). The second is that these three categories are very broad and there might be many other subcategories involved which might be revealing important information, and finally, the last challenge is the fact that there might be some associated responses that do not fit into any of the categories. In many studies in order to overcome this problem, different approaches have been adopted in the way of developing new schemes. Some researchers add a fourth category as "other" (e.g. Nissen & Henriksen, 2006; Orita, 2002; Wolter, 2001) and some others develop different new frameworks suiting the purpose of their research (e.g. Fitzpatrick, 2006; 2007; Fitzpatrick & Izura, 2011; Henriksen, 2008).

Accordingly, it appears that the potential solution can be developing new categorization systems and schemes, which are finer-grained. Fitzpatrick (2006) developed a detailed categorization system based on various sets on information. She used the information from the previous research on word association, while closely examining the elicited and collected responses and classified them into different categories. Consequently, in order to identify the main categories, she applied Nation's (2001) pedagogical word knowledge taxonomy (Fitzpatrick, 2006, p. 130–131). Finally, the following table (Table ***4.1***) provides a general overview of a few salient and recent word association studies, organized on the basis of the approaches each has taken to categorization and classification of the responses.

Table 4.1 A methodological overview of word association studies

Study	**Number of cues**	**Number of responses**	**Categories**
Wolter (2001)	90	1	paradigmatic, syntagmatic, clang, no response
Namei (2004)	100*2	1	paradigmatic, syntagmatic, clang,
Nissen and Henriksen (2006)	90*2	2	paradigmatic, syntagmatic, phonological, other
Fitzpatrick (2006)	60	1	Four main categories and seventeen subcategories:

			✓ meaning-based (defining synonym, specific synonym, ✓ hierarchical/lexical set, quality, context, conceptual); position-based (consecutive xy, consecutive yx, phrasal xy, phrasal yx, different word class collocation); ✓ form-based (derivational, inflectional, similar form association); ✓ erratic (false cognate, no link). ✓
Fitzpatrick (2007)	100	1	Three main categories and nine subcategories: ✓ meaning-based (defining synonym, specific synonym, lexical set, conceptual); ✓ position-based (consecutive xy, consecutive yx, other collocation); ✓ form-based (change of affix, similar in form only); erratic (no link). ✓
Fitzpatrick & Izura (2011)	190	1	equivalent meaning, non-equivalent meaning, collocation, form, form plus meaning, meaning plus collocation
Yokokawa et al. (2002)	20*2	n/a	antonym, syntagmatic, category-exemplar, exemplar-exemplar, synonym, other.
Henriksen (2008)	48*2	2	canonical and high frequency; canonical and low frequency; non-canonical and high frequency; non-canonical and low frequency; chaining; form-related; translation; repetition; empty; ragbag

Scrutinizing these studies along with many more in this realm, it is still possible to claim that although each single study asserts to use more transparent category labels and are highly fine-grained to observe differences among different categories, they still “echo the traditional paradigmatic/ syntagmatic/clang approach, despite deriving from different theoretical frames” (Fitzpatrick & Thwaites, 2020).

Another group of more recent studies have also applied various approaches in categorizing the responses by taking the aim of their research into account. For example, Yokokawa et al. (2002) investigating the semantic network building aspect of vocabulary acquisition, focused on seven categories of antonym, syntagmatic, category-exemplar, exemplar-exemplar, synonym, other, which was solely developed for the purpose of their research. Fitzpatrick and Izura (2011) have added response time measures and updated the categorization based on lexicosemantic relationship between cue and response. Also, Henriksen (2008) incorporates frequency, canonicality to the categorization of responses. This categorisation takes into account shift from form-related responses to more meaning-related, increase in the number of canonical responses, and decrease in the number of highly frequent responses and increase in the low frequent ones. Finally, a sophisticated and novel system of measurement of overall network knowledge was created.

Summary

From the discussion above on the literature on word association research, it appears that much of the difficulty and inconsistencies in this paradigm originates from the methodological aspects. Due to the variety of methodological possibilities and alternatives, it is difficult to say what constitutes good data in terms of selection of the cue words or system of analysis of responses, which might mask subtle behavioral patterns (Fitzpatrick, 2006; Meara, 2009). In order to avoid the challenges in this paradigm and come to reliable data, based on which judgments can be made on the structural properties of the bilingual mental lexicon, an attempt has been made to design a word association test for collecting and processing data that address the difficulties discussed.

4.3. Developing and validating productive word association test

Associations are regarded as "language of thought" (De Deyne & Storms, 2015), assuming to reveal the truth about the nature of the mental lexicon and offering insight into the ways it is organized and structured; hence, the process of designing a valid and accurate test is of paramount importance. Hence, this section explains various steps taken in the design of the word association task for the purpose of the current research, starting with the process of selection of the cue words and the lexical variables taken into account in this process, following by a Yes/No receptive test of vocabulary, and finally developing tests for three phases of pre-test, immediate post-test and delayed post-test of word association test.

4.3.1. Selection of cue words

The discussion above on the literature of word association research has indicated that the properties of the cue words have significant impact on the inferences which can be made on the organizational properties of the mental lexicon since the associated responses, based on which conclusions are drawn, are strongly influenced and even determined by the cue-level variables and their characteristics (Fitzpatrick, 2006). Lexical variables such as the grammatical class of the cue, level of frequency, and propensity to elicit strong responses reveal differences about network properties of connectivity, strength of the connections, and centrality. Therefore, selection of the cue words is one of the most important steps in developing the word association tests, the process of which will be explained in this section.

4.3.1.1. Controlling for the lexical variables

As the purpose of the current study was practicing the already acquired lexical items with the aim of building up on the associative connections rather than teaching new items and creating new links, the cue words were decided to be selected from textbooks that participants had already covered at the language school in previous terms.

The book series used in this language school to teach General English was *American English File Second Edition* by Christina Lathem_koenig and Clive Oxenden. Each of these books were not covered in a single term, but each one was divided into three sections and each section was delivered in a twenty-session term. Participants in the current study, who were in the intermediate level of language proficiency, were studying the third and last part of the *American English File* 3, covering units 7 to 10.

The primary list of cue words for the word association test was selected from the previous units that participants had already covered either in the same or previous books. The process of selection of the words was totally random from the vocabulary parts, reading sections, and listening transcript sections of the units. A list including one hundred and twenty lexical items was generated, sixty of which were related to *traveling* topic and other sixty were *sports* vocabulary (the words were selected from unit 11 book 2, unit 5 book 3 on *sports* and unit 5 book 3, unit 2 book 2, and unit 3 book 2 on *travelling* topic). Subsequently, from this basic list of words only the ones meeting certain criteria were singled out for the next phase of the study (see Appendix L).

First of all, phrasal words (e.g., take off, put on, warm up, etc.) and compound nouns (e.g., jet lag) were omitted from the list. As the participants in the current study were in the intermediate level of language proficiency, the possibility of such items being processed serially as two separate lexical items rather than a single one was high, which could affect the speed of activation and access of the items. For example, phrasal verbs could be accessed as two separate lexical items connected by a syntactic rule (requiring syntactic processing rather than lexical processing). Hence, both these two categories increase the load of processing, which makes it incomparable to single words. In order to have homogenized and comparable cue words in this term, all cues were single lexical items accordingly.

Secondly, polysemous and homographic words were omitted from the list; that is, single word forms which were associated with two or several related senses or the ones sharing the same orthographic/phonological forms though have different meaning origins. It needs to be noted that no all polysemous words were removed from the list but rather polysemous words, which were defined as those for which participants had been taught more than one meaning, were omitted. As the purpose of the study was manipulating the structure of specific strata of the lexicon on local level rather than the global one, it was attempted to keep the lexical items which only had a single sense particularly related to the topics, to avoid further expanding the connections globally over the lexicon and assessing a particular stratum in question. In addition, some participants might not know the other senses of the lexical item and would be at a disadvantage at the word association test in comparison to the ones who know all meaning senses, even the ones not related to the topic. Hence, they might receive higher scores though those associations were not the main aim of the task. For example, the word *tie,* which was selected on the first place from the textbooks, was omitted afterwards, as it was not clear if the participants provide associations for one of the senses as 'attach or fasten with string or similar cord' or the intended meaning which was 'achieve the same score or ranking as another competitor or team' as during the administration of word association test, there was no notion which one of the grammatical categories they had to write associations for.

Next, the frequency level of lexical items was taken into account; that is, the number of times a lexical item appears at a particular corpus of texts. Fitzpatrick (2006) suggested that the cue words are better to be compiled from various levels of frequency, while the ones from very highly frequent levels to be avoided. The reason is that highly frequent cue words tend to have well-established connections, eliciting highly predictable and systematic responses (Fitzpatrick, 2007). However, for the case of the current study, as multiple responses to each cue word were permitted, such primary responses to the highly frequent lexical items wouldn't dominate the distribution of the responses in the test level, but instead the high variety of responses provided to the higher frequent items would reflect the number and strength of connectivity to the items in the core of the network. Accordingly, in order to gain a better understanding of the structure of the lexicon, higher number of associations provided to high frequency items was desirable as much as the heterogeneity of the response types as a result of multiple responses. In addition, since the participants in this study are at the intermediate level of language proficiency, most of the lexical items known to them are among this level of frequency and avoiding them would be very difficult.

While it was decided to keep the highly frequent cue words, very low frequency items were not included in the list of cues in this study, as decrease in the level of frequency has a direct significant impact on response availability and low distribution of associations on multi-response word association tasks (Cramer, 1968; De Groot, 1989). Accordingly, in order to choose the cues from different frequency levels, while not relying heavily on very low frequency ones, the list of words which had already been developed by the researcher from the textbooks, was compared to an English corpus of words.

The level of frequency of the lexical items in the list of words was checked against Nation's (2012) BNC/COCA2000, while setting the high frequency threshold on 2000 most frequent (Nation, 2001) and the low frequency +9000 (Schmitt & Schmitt, 2014). The superiority of this corpus is a good balance between the written (40%) and spoken (60%) components in comparison to the other corpora mostly comprising of written material (e.g. West's (1953) General Service List is 100% written, Nation's (2006) BNC2000 is 90% written and only 10% spoken). In addition, it is argued that this corpus is the most appropriate and useful for the language learners from the perspectives of corpus linguistics, learners, and teachers (Dang, 2017; Dang & Webb, 2016).

Subsequently, the next point, which was taken into account in the process of the selection of the cues, was the grammatical class of the cues. It has been argued that the grammatical class of the cue word tends to be reflected in response and can affect response behavior (e.g. Deese, 1962; Fitzpatrick, 2006; Fitzpatrick & Izura, 2011; Nissen & Henriksen, 2006). Nissen and Henriksen (2006) found out that "nouns seem to be predominantly organized in paradigmatic relations, whereas verbs and adjectives are characterized by syntagmatic relations" (p. 46). Accordingly, it appears that use of cue words from various word classes prompt different kinds of associations; hence, the distribution of the cue words in terms of the grammatical class was controlled for, and cues from the three main classes of nouns, verbs, and adjectives were selected for the purpose of this study.

Finally, after controlling for all above variables, from the first list of one hundred twenty random words, seventy-eight of them remained for the next phase of the study, which was a Yes/No receptive test of vocabulary.

4.3.1.2. Developing Yes/No receptive vocabulary test

The next stage towards the purpose of designing the productive word association task was to ascertain that all of the cue words were known by the participants. As the purpose of the word association task was assessing the productive knowledge of the vocabulary and no new lexical items was planned to be taught during the intervention sessions, the cue words needed to be known on the receptive level of knowledge. For this purpose, a Yes/No receptive test of vocabulary was designed and administered. This type of test is a popular test measuring the receptive knowledge of L2 vocabulary. It is a type of meaning-recall and receptive retrieval test, asking for a simple judgment on knowing the meaning of the L2 word.

This test included items both from the remaining seventy-eight words from the previous stage along with pseudowords/nonwords. The idea of including pseudowords into Yes/No tests of vocabulary was first introduced by Anderson and Freebody (1981). Pseudowords are phonologically and orthographically possible words, which are included in the test in order to control guessing. They are considered to be important in the process of Yes/No tests, as this type of test is mainly self-reporting, and students have a strong tendency to overestimate their knowledge of words at eleven percent, which increase as the words in question are among the lower frequency levels (Anderson & Freebody, 1981).Therefore, the correct responses to the words, hits, is a representation of the vocabulary knowledge, whist choosing pseudowords as correct responses, false alarms, shows guessing.

The test included one hundred and two items in total, seventy-eight of which were words from the previous stage and twenty-four pseudowords. Each pseudoword was created using *ARC Nonword Database* (Rastle, Harrington, & Coltheart, 2002) (www.cogsci.mq.edu.au/research/resources/nwdb.html). This website contains 358,534 monosyllabic pseudowords, 48,534 pseudohomophones, and 310,000 non-pseudohomophonic nonwords, and can generate legal pseudowords automatically, based on different number of criteria.

In different studies, the proportion of the numbers of words and pseudowords varies to a great extent. In a study conducted by Mochida and Harrington (2006), the yes/no test included one hundred fifty items, which included ninety words and sixty pseudowords. Different researchers decided about it on the basis of the nature of research they were carrying out. Psuedowords in this study consisted only 20% of the total number of words as the main aim of this test was selection of appropriate cue words for the word association tests, and the reason for inclusion of pseudowords was to detect false alarms (Read, 2000). That is the reliability of the test and consequently the cue words chosen would be under question if the participants claimed to know the ones which did not exist at all.

The lexical items in this test included twenty-seven nouns selected, twenty-six verbs and twenty-three adjectives. In addition to considering the word class of the words, the number of syllables were also taken into consideration. The nouns and verbs were mostly chosen from one to two syllabic words, while adjectives were up to four syllables.

The test was administered to a group of thirty intermediate level language learners (according to the language school placement test), who did not take part in the main phase of the study. The test elicited information on whether the cue words chosen were known receptively by the group of participants in this level of language proficiency. Accordingly, no judgment towards the final results was going to be made on the basis of the results of this test, but it was used as a tool to choose the cue words that most participants in intermediate level had knowledge of for the next step of the research process. The test was administered during the normal teaching hours, in the beginning of the classes. The time allocated was nine minutes in total, as Mochida and Harrington (2006) suggested that 500 milliseconds should be enough for each item in the test. The test was designed in two rows, in two pages. Each row contains twenty-five items, which were randomly assigned to the rows via online randomizer (https://www.random.org/lists/) (see Appendix M).

4.3.2. The final list of cue words

Out of seventy-eight words in the Yes/No receptive test of vocabulary, forty-eight of them received complete score of thirty out of thirty; that is, forty-eight of the lexical items in the test were known by all the participants. The results met the expectations with respect to high number of known items, since the lexical items in the test were chosen from units of their textbooks that the participants had already covered and had knowledge of. However, it was decided not to use all of the forty-eight items for the final version of the word association test. The main reason for keeping the number of cues small was practical limitations. As the word association test was a productive one with chances to provide multi-responses, the process of the data coding, analysis and conducting the retrospective interviews would have imposed significant time demand in case of high number of cue words. Accordingly, for the purpose of developing word association tests, only thirty words out of the list were used, that is ten cue words for each phase of pre-test, immediate post-test, and delayed post-test.

Hence, as the number of words selected by the participants were more than required for the next phase, the ones with higher number of association probabilities were selected. In order to check the number of associations, the *University of South Florida Free Association Norms* was chosen.

This database is the one of the largest free association databases with more than 6000 participants, generating three-quarters of one million responses to 5019 cue words (Nelson, McEvoy, & Schreiber, 2004).

Accordingly, among forty-seven words, the ones with lower potentials of promoting retrieval of many related words (e.g., *fascinating, reasonable, and relaxing*) were excluded from the list. In addition, in case of having lexical items with a strong pair in the list, such as *boring* and *exciting* (antonyms) or as *jungle* and *forest* (synonyms), it was decided to keep one of the words resulting in higher number of associations and remove the other. Consequently, from the remaining list, thirty cue words for the purpose of productive word association tests were selected randomly from the remaining list by the use of online randomizer.

The final step was dividing this list into three groups for the three test times of pre-test, immediate post-test, and delayed post-test of productive word association. In order to have tests of equal level of difficulty, three factors were required to be taken into consideration. Firstly, each of the tests had same number of ten cue words. Secondly, words from various levels of frequency ranging from high to mid, with priority on higher frequency, were included in all three versions. Therefore, in each test version, there were seven cues of the first 2000 most frequent words and three of mid frequency level. Besides, the grammatical class of the cues in each test version was consistent. The cue words were from three categories of nouns, verbs, and adjectives, which were distributed in three versions of the word association test in equal number. In each test, there were four verbs, three adjectives, and two nouns. As the number of cues was ten and there were three grammatical categories, the tenth word was added randomly to each group. Consequently, the final version of the selected cues for each phase of word association test was randomized for the presentation in their order via the application of online randomizer (https://www.random.org/lists/) (see Appendix N).

4.4. Productive word association coding scheme

The next stage after selection of the cue words and designing the test concerns deciding on a system to categorize and measure the response data supplied by the participants on the free word association task, with the aim of investigating the organization of words and concepts in mind.

As it has already been mentioned (see Section 4.2.3), there are two main approaches to word association measurement; firstly, assessing responses in terms of number of the links in the mental lexicon while comparing them to stereotypes, and secondly analyzing the lexico-syntactic pattern of associative responses; that is, categorizing the responses according to the linguistic relationship between the cue and the responses provided. Accordingly, many previous studies applied the use of weighted or non-weighted scoring procedure comparing participants' responses to list of norms (e.g., Randall, 1980; Kurse et al., 1987; Wolter, 2002), while many others have analyzed types of associations provided by classifying the responses into fine-grained categories (e.g., Deese, 1965; Fitzpatrick, 2006, 2007; Fitzpatrick & Izura, 2011; Navrascics, 2007). Deciding on which to adopt is heavily dependent on the purpose of the study and type of information the researcher needs to elicit about the lexicon.

In line with the aim of the current study, investigating the network knowledge and the structure of the mental lexicon, the theoretical properties of the lexicon are taken into account in the way of choosing an appropriate system of measurement. To this end, a sophisticated measurement system, incorporating the connectivity of the items in the lexicon and the strength of the connections, both of which leading to determining the centrality of the items in the network, is adapted for the purpose of this study from Henriksen (2008).

This measurement system incorporates the number of connections among lexical items in the network along with the weight of the connections, which is determined by the canonicality, type, and frequency of their occurrence in the network. Along with the coding scheme, the scoring rubric is also a combination of both response-item and response-type analysis, which has also been adapted from Henriksen (2008). The original version of the scoring procedure was a creative and novel one in that, not only it analyzed the responses in terms of the norm list created, but also type of responses was taken into consideration; however, some necessary changes were applied.

4.4.1. Overview of Henriksen's coding scheme

The developed coding scheme is based on several distinctions and is composed of ten mutually exclusive response types in four different levels of a hierarchy, representing various response behaviors.

On the very first level, distinction is made between two groups of responses; responses providing sort of association to the cue word and the ones which are not related at all. Responses which are either direct translation of the cue word (*jungle* eliciting جنگل), mere repetition of it without providing any changes to the form (*jungle* eliciting *jungle*), ragbags and random words (*jungle* eliciting *computer*), or empty spaces are categorized as unqualified responses as they reveal lack of ability to provide any type of association to the cue. The other group of responses are the qualified ones supplying some type of association to the cue word. This group is further categorized in the next level; formally linked to the cue word (form-related responses), semantically linked to the cue word, and finally semantically linked to the previous response rather than the cue (chains).

Initially, form-related responses are the ones which mirror formal specifications of the cue word. That is, they are phonologically related associations/clangs, which are responses that are phonologically or orthographically linked with the cue word and are believed to bear weak type of associations that low-level proficiency L2 learners can mostly provide, representing mental lexicon at fairly early stages (*jungle* eliciting *jumble*). Next, responses supplying semantic connection to the cue word can be either syntagmatically, paradigmatically, or idiosyncratically linked to the stimulus word; however, in order to avoid problems associated with providing distinctions among these three types of response categorization, Henriksen (2008) decided not to further categorize the responses and include them all in one single category of *semantically linked to the stimulus response* (*jungle* eliciting *forest*). Finally, the third type of associations are chains or semantically related responses to the previous response. This group of responses are associated and linked to the previous response and do not bear any type of link and connection to the cue word (*jungle* eliciting *blood* which is associated with *lion* as the first response).

The third level in the hierarchy of the response types is only further categorizes the semantically related responses and identifies whether the semantically related responses are canonicals or non-canonicals. That is, it explores if such responses are prototypical responses which are provided by most of the L1 norming group or not. The main reason this level in the hierarchy is assigned to the canonicality of the responses is argued to be their central role in organization of the mental lexicon. Canonical responses act as bridges between different parts of the lexical network;

therefore, there is a strong tendency to provide more canonical responses with an increase in the level of language proficiency as mental lexicon develops (Namei, 2002, 2004).

Consequently, the last level of classification is assigned to the frequency level of the semantically associated responses, as word frequency is a critical and necessary basis for measuring learners' vocabulary knowledge and knowing words in various frequency levels can be regarded as a predictor of lexical competence and proficiency (Catalán & Fitzpatrick, 2013). Henriksen (2008) further categorized the responses into high frequency and low frequency words with the cut-off point of the first 5000 words being high frequency and the rest low frequency. The final figure for various kinds of response types identified by Henriksen (2008) is presented in the Figure ***4.1***.

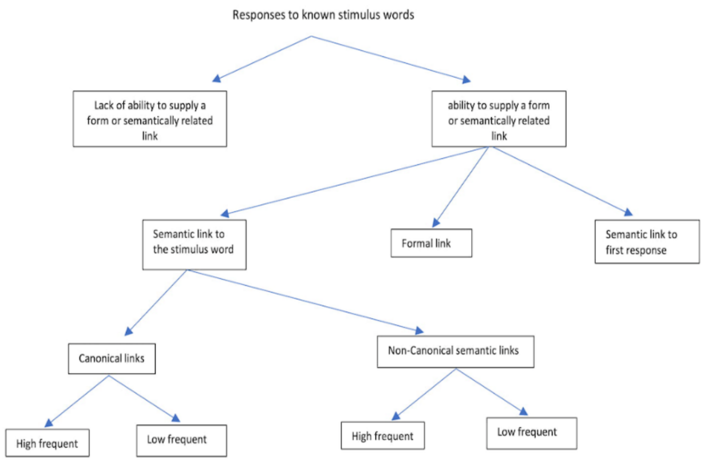


Figure 4.1 Response types identifies in the word association test by Henriksen (2008)

4.4.2. Adapted version of coding scheme

In order to explore any changes in the network knowledge and organization of the mental lexicon shifts in the connectivity and strength of the connections; that is, number and pattern of the responses, were expected. As mentioned earlier, the coding and scoring scheme to track the shifts were adapted from Henriksen (2008), although some changes had to be applied to fit the purpose of the current study. The changes applied were as follows;

- Type and categorization of associated responses,

- Unqualified response
  - ⇒ Repetition
  - ⇒ Proper names
  - ⇒ Examples
  - ⇒ Ragbag
- Loosely associated responses
  - ⇒ Chains
  - ⇒ Form-related
  - ⇒ Two step association
- Closely associated responses
  - ⇒ Semantic-based
  - ⇒ Position-based
  - ⇒ Semantic/position based

- Canonicality of the responses and the process of developing normative list,
- Level of frequency of the responses and adaptation of a frequency threshold.

4.4.2.1. Type and categorization of associations

Further to Henriksen (2008), in this study the responses were categorized into three main types, as it has already been explained in the previous chapter (see Section 3.7.1.1.2). In the first level, distinctions were made between unqualified responses, loosely associated responses and finally the closely associated response group.

The first category is assigned to unqualified responses, bearing no association of any kind to the cue word. Unqualified responses are defined as "informants lack of ability to provide a form-related or semantically-related association" (Henriksen, 2008, p. 52). This group is composed of four types of unassociated responses. The first type of unqualified responses is a mere repetition of the cue word (the cue word *lake* eliciting response *lake*). Secondly, the associated response is in the form of a proper name (the cue word *Amateur* eliciting response *Golzar*). The next group are the responses which are examples of the cue word (the cue word *lake* eliciting response *Khazar*), and finally the last type is ragbag. Henriksen (2008) defined ragbags as responses which are impossible to be coded as they are "indecipherable, or it is difficult to decide whether a Danish and English response has been given" (p. 206). However, this study defines ragbags as the responses which are either real words or none-words, while having no conceptual or linguistics connection to the cue word (the cue word *Jungle* eliciting response *she*).

The second main group is responses which are *loosely associated* to the cue words. Loosely associated responses are the ones that propose no direct and straightforward association to the cue word. They are composed of chains, two-step associated responses, and form-related responses. Chains can be defined as "semantically related associations to the informant's first association" (Henriksen, 2008, p. 47); therefore, such responses don't bear any association to the cue (the cue word *cycle* eliciting response *hang out* via *park*). Secondly, two-step associated responses are associated with the cue words through a link with another word. This signifies the existence of an intermediate level association, such as *weak, Monday* associated via *week* (Fitzpatrick et al., 2013, Fitzpatrick, 2006; 2007). Therefore, they do not share any direct semantic link with the cue. Finally, the last category is the form-related responses that are similar in form only. This group of responses are similar in phonology or orthography to the cue word, such as the cue word *hurt* eliciting *heard*. It is noteworthy to mention that responses which are the same as the cue with addition, deletion, or changing of either inflectional or derivational affixes are also considered to be form-related (the cue word *enjoy* eliciting response *enjoying*).

Finally, the last group of associated responses were the ones providing the strongest and closest relationships between the cue words and the responses, in which three subgroups of semantic-related, position-based associations, and semantic/position-based responses can be identified (Fitzpatrick, 2006, 2009, Fitzpatrick & Izura, 2011). Although in the current study, it was decided to distinguish among different types of closely associated responses, Henriksen (2008) in her coding scheme organized all these three different types of associations into one group of *semantically linked to the cue word* and didn't distinguish among them. The reason was argued to be various ways in which each of these categories can be defined, as a result of which it would be "extremely difficult to find clear and objective criteria for characterizing a specific response as belonging to any of these three categories" (Henriksen, 2008, p. 46). Henriksen (2008) classified the position-based responses as meaning-based, while position-based responses are not determined by semantic characteristic, but by syntactic and collocation properties, and it was necessary to distinguish between these very different types of responses in the coding procedure.

In the current study, rather than considering the response type categorization intractably complicated and controversial, the novel categorization framework proposed by Fitzpatrick et al.

(2013) was adopted as a potential route of solving some of the problem and assigning responses to three categories of semantic-based, position-based, and semantic/position-based. The main reason for application of this framework was the clear criteria set for assigning responses to each of the categories (Table ***3.7***); moreover, inclusion of the semantic/position-based category of responses avoid the subjective judgments to be made in assigning responses into exclusive discrete categories as a result of which the process of coding would be carried out in a more reliable and accurate manner.

The main reason for distinguishing between response types in closely associated group was the contribution of this classification to our understanding of the structure of the lexicon. It has been argued that "words that are barely known may elicit phonologically-based associations, those that are partially known may have a strong syntactic organization, and well-known words are connected to other words mainly on a semantic basis" (Namei, 2004, p. 363). This is in line with Wolter (2001) stating that the lexical items in the center of the mental lexicon tend to elicit more meaning-based responses as they have been fully acquired while the peripheral items have mostly phonological type of associations. Accordingly, adding subgroups of closely associated response types can add to our understating of the level of knowledge of the lexical items contributing to the structure of the lexicon and the particular strata in question.

Although three different types of closely associated responses were identified, all of them were grouped together for the purpose of the quantitatively scoring procedure, while each of them was coded as a single exclusive category for further descriptive discussion of the changes in the response behaviors.

4.4.2.2. Canonicality of the responses

Next step in the process of classification of the closely associated responses was to determine whether the responses were canonicals or non-canonicals. Canonical responses are prototypical and stable associative links that are provided by many informants in word association tasks (Schmitt, 1998a; Singleton, 1999; Murphy, 2003; Wilks & Meara, 2002). This type of associations are representatives of the strongest links between the cues and responses. Henriksen (2008) decided to assess the responses in terms of the prototypicality or canonicality by collecting list of norms from native speakers. Canonical responses, in her study, were particularly defined as the most frequently provided associations by the native speakers as a response to a single stimulus word, which were either syntagmatically or paradigmatically associated with the cue. The use of native speakers' norm to probe into the canonicality of the responses was justified by arguing that mental lexicon of the native speaker is the hallmark of a very well-developed lexicon and the associations they provide can represent the conventional access routes available to a fully competent language user.

However, the assumption that native speakers' response preferences is homogeneous and can be used as a normative list have led many studies into ambiguous and unsatisfactory results, and findings of numerous studies encourage researchers to be extremely wary of the collection of L1 speakers' data as a normative list (Fitzpatrick, 2006; 2007; Nissen & Henriksen, 2006; Wilks & Meara, 2007, Zareva & Wolter, 2012). Fitzpatrick (2007) argues that "not only natives vary in the actual responses they produce; they also seem to vary in the type of associations they make" (p. 327). The assumption that there exists a coherent L1 associational behavior connected with assessing L2 association responses in terms of commonality and nativelikeness is strongly challenged.

In addition, it seems that there are systematic differences between ways natives and non-natives make associations as the nature of the mental lexicon is different. Non-native speakers develop their L2 mental lexicon fairly late, and they make use of L1 mental lexicon in the development of the L2 mental lexicon, which makes the comparison under question, and the direct impact it has on the associations cannot be neglected. In addition, both L1 and L2 speakers' mental lexicons are influenced by linguistic, contextual, sociocultural factors which heavily influence the way they create associational patterns (Pavlenko, 2009; Meara, 2009; Meara & Wolter, 2012). Therefore, it can be argued that there is not enough empirical research showing that non-natives develop native like responses and accordingly their responses are comparable (Meara, 2009; Meara & Wolter, 2012).

Creating the norm list

In the present study, it was decided to create a bespoke normative list to match the characteristics of the target population, while defining it as the "lists of responses given to cue words by particular participant groups, ranked by response frequency" (Fitzpatrick & Thwaites, 2020, p.

242). Such a list could “reliably reflect the maximum possible number of characteristics of the study population” (Fitzpatrick et al., 2013, p. 33). In this way, the list of norms was specific to the target population, which was tied up to their age, educational background, geographic context, linguistic, sociocultural, and socioeconomic factors (Pavlenko, 2009; Fitzpatrick, 2015; Fitzpatrick et al., 2013). Consequently, by collecting the stereotypes from the same population, it was possible to track the shifts across the population and explore the changes.

In order to create the list, all associative responses to each cue word collected from the participants were compiled into a single list in Excel files. Next, in order to develop objective criteria for identifying the most frequent responses to each of the cue words, the number of most frequently occurring responses was taken into account. Accordingly, it was decided that the four most frequent responses from the frequency list would be regarded as canonical responses. The prerequisite to count how many of the responses were repeated, was to lemmatize data and define what it was meant by ‘word’. In order to lemmatize the normative list of responses, the same procedure as the lemmatization of the associated responses was applied (see Section 3.7.1.1.1).

4.4.2.3. Frequency level of the responses

Level of frequency of the responses associated with the cue words was the last factor to be taken into consideration while developing the coding scheme. It has been argued that learners with more advanced mental lexicon are able to provide associated responses which are more abstract and lower in frequency (Namei, 2002, 2004; Henriksen, 2008). Accordingly, a way to track developments in the network and the degree of word knowledge can be the level of frequency of the lexical associations.

Hence, the other point which required adaptation to fit the aim of the current study was the cut-off point of the frequency level, which Henriksen (2008) set it on the first 5000 most frequent words. It was argued that 5000-word level can distinguish between the higher and lower frequency lexical items as it approximately provides a text coverage of 90% of any type of text (Nation, 2001; Warning & Nation, 2004). In addition, language users are not able to transfer L1 higher order skills into L2 with a vocabulary size less than 5000. However, operating in a spoken environment is totally different from that of the written one, due to the fact that higher percentage of vocabulary coverage might be needed, as interlocuters are required to process the spoken discourse online within time constrains and restrictions to communicate and interact efficiently, especially when fluency of speaking is the main point of consideration.

Hence, in order to come up to a more reliable threshold of the spoken word coverage required for the spoken environment, rather than setting a high cut off point of 5000-words, a new distinction between high and low frequency words was opted for, and the cut-off point was set at the first 2000 most frequent word families, which is approximately equivalent of 3000 single words. This lower cut off point was then to account for the fact that this study was dealing with fluency of speech, and there has been an agreement among researchers that approximately 2000-word families can provide enough lexical resources required to engage in everyday communication (Nation & Meara, 2002; Schmitt, 2000; Schmitt, 2010), having about 95% coverage of spoken discourse.

The next group, low frequency vocabulary, has been conceptualized in widely various ways. The boundary between high and low frequency words is arbitrary as considering a word as highly frequent or not is mostly corpus-based (Nation, 2001). Low frequency words have sometimes been conceptualized as words above 2000 level, some other times 2000+ academic vocabulary. More recently Nation (2006) have questioned the frequency levels of high and low, mentioning that approximately 8000 to 9000-word families is required for written English and up to 6000 to 7000 for spoken English. The word families beyond these frequency levels can be regarded low frequency, while leaving 4000 in between high and low. This gap is bridged by the mid frequency vocabulary category, which ranges from 2000 to 8000 for wide reading in English and 2000 to 6000-word families for spoken English environment (Schmitt, 2010), which is required to bridge the lexical gap and gaining a better command of the spoken proficiency. Learning the mid-frequency words help L2 speakers to use language without any problems due to the lack of required lexis (Nation, 2006).

Consequently, for the purpose of this study, the high frequency level threshold is set to be the 2000 highly frequent word families, followed by mid frequency being 2000 to 6000 word families, as wider range of spoken vocabulary is required to engage in everyday verbal communication. Although a cut-off point of 6000 word families has been set as a threshold to distinguish between mid and low frequency words, both levels were grouped together due to the fact that with lower levels of frequency, there is a noticeable decrease in percentage of spoken coverage which is gained per 1000 word families.

The final point to be taken into about is about setting the unit of measurement – various ways of counting the lexical items. Although it was decided to set the threshold of high frequency vocabulary to be the first 2000 most frequent word families, Spoken BNC 2014 is complied with individual single word forms rather than word families. As it is a tedious work to go down the BNC list and tally the frequency of members of word families and categorize them under headwords, it was decided to carry out the research with single word forms rather than word families. Based on a study by Adolphs and Schmitt (2003) on the analysis of two modern corpora of CANCORD and spoken component of BNC, spoken coverage of 3000-word families in CANCORD has an approximate coverage of 5000 individual word forms from BNC, which is around 95%, and accordingly, 2000 word families would have approximate coverage of 3000 single word forms, with 95% spoken coverage. Therefore, a distinction was made between high frequency word forms and low frequency word forms, with the cut-off point of 3000 single word forms (for a discussion of the choice of the corpora refer to Appendix O). Finally, based on the distinctions and revisions on the original coding scheme, the following coding scheme presented in figure 4.2 was developed.

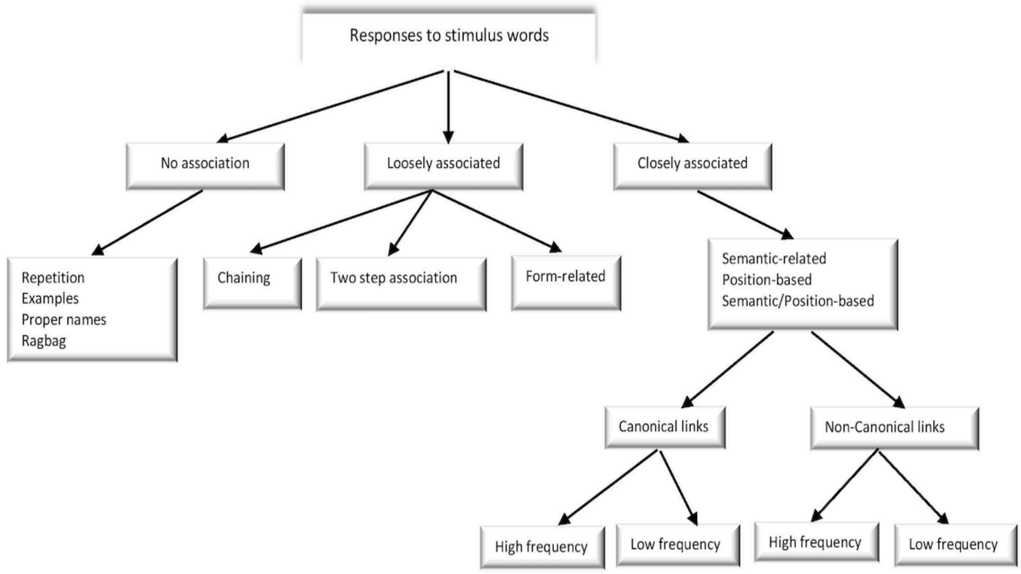


Figure 4.2 Response types identifies in the word association test

4.5. Summary

In this chapter, I have outlined the design of the word association test, procedure of developing it, and the coding scheme adapted. The chapter began with a brief introduction of word association behavior and research, which was followed by a discussion of the potential methodological obstacles resulting into inconsistencies in identifying behavioral patterns. In the light of the relevant literature in this realm and observations from the pilot study, details about the design of the task were set out, and three versions of the word association task for pre-test, immediate post-test, and delayed post-test were designed. In terms of the analysis of the data collected, it was discussed that a version of coding scheme was adapted from Henriksen (2008), while changes were applied to meet the purpose of the current study. Based on the presented coding scheme, an overall word association score was assigned as an estimate measure of each participant's productive network knowledge in specific strata of the lexicon, which was explained in the previous chapter (see Section 3.7.2.1). The next chapter will turn into the analysis of the data elicited in detail.

5. Chapter 5: Analysis

5.1. Overview

The purpose of this study was to examine the impact of practicing vocabulary by the use of semantic mapping technique on EFL learners' organization and structure of mental lexicon as well as L2 fluency of speech. That is, it aims to find out if there were any significant differences between groups of experimental and control, as the result of practicing vocabulary with the application of semantic mapping technique, in regard to the structural properties of the lexicon and L2 utterance fluency measures. Consequently, the research intended to investigate if the changes in network knowledge was correlated to L2 utterance fluency of speech. Accordingly, the following research questions were posed;

RQ1: To what extent does semantic mapping affect organization of bilingual mental lexicon as measured through free productive word association?

RQ2: To what extent does semantic mapping affect L2 speech fluency measured through empirically established means of gauging L2 fluency?

RQ3: Is there a relationship between organization of bilingual mental lexicon and L2 speech fluency?

This chapter is divided into three parts, each of which answers one of the above research questions in the order presented.

5.2. Word association

5.2.1. Overview

The rationale for assessing participants' performance on word association was to investigate if semantic mapping technique can contribute to changes in the structure and organization of the mental lexicon, particularly the network knowledge. It explores the changes in the network associated with the strongest and most immediately accessible words in the lexicon after the treatment, being described on the grounds of different categories.

Prior to the analysis of the results of free productive word association tests various steps were taken in treating, coding, and scoring the responses, as already discussed (see Section 3.7.1.1). In the very first stage, the responses were treated regarding spelling and shortening to headwords followed by lemmatizing them (see Section 3.7.1.1.1). After having the list of the responses, a coding scheme was required to systematically categorize and code the data elicited. The coding scheme was adapted from Henriksen (2008), while various changes were applied (see Section 4.4). Accordingly, the responses were categorized into three main levels. The first category is the response type (not associated, loosely associated, and strongly associated responses), which was further divided into various types of responses in each of the categories (see Section 4.4.2.1). The second category is the canonicality of the responses, which refers to the degree to which the responses are frequent in the norm list (see Section 4.4.2.2), and finally, the last category is taking the level of frequency of the responses into account, as providing response in a lower frequency band can reflect more advanced and structured mental lexicon (see Section 4.4.2.3). Consequently, based on this coding scheme, a single score was assigned to each response provided (see Table ***3.8***), following which a score representing an overall network knowledge quality was assigned to each participant (see Section 3.7.2.1).

Subsequently, in order to answer the first research question the following steps were taken. Firstly, the assumption of parametric and non-parametric tests, the data distribution, was discussed to decide on the suitable statistical procedure. Secondly, participants' performance on word association test was analyzed for each group of control and experimental separately, in each time of testing, pre-test, immediate post-test and delayed post-test. And finally, the gains were compared across control and experimental groups.

5.2.2. *Assumptions of parametric vs. non-parametric procedures*

The first step prior to analyzing the data was making decision as to choose a parametric or non-parametric statistical procedure, which is heavily dependent on satisfying certain assumptions underlying parametric statistical tests. In order to be able to use parametric tests, two main assumptions of normality and homogeneity of the variances were explored to check if they were satisfied. In general, the data was found to be normally distributed looking at histograms, Q-Q plots, and Shapiro-Wilk test results (see Appendix P). The second assumption applying to parametric tests, satisfying the homogeneity of variances, was assessed. Checking the variances of the variables between the two groups of experimental and control, Levene's test was non-significant, thus equal variances could be assumed. In addition, the F_{max} statistics was applied to check the assumption of homogeneity of variances per the repeated measure factor ($F_{max} = 1.64$), which indicated that the homogeneity of variances had not been violated.

The results in this part report that all the assumptions of parametric tests had been met, and none of them was violated. Accordingly, parametric test of mixed between - within ANOVA was employed to analyze the performance of the participants on word association test.

5.2.3. Preliminary screening of word association data for running mixed ANOVA

The very first step in the analytical procedure was setting the alpha level. The alpha level for mixed ANOVA statistical test was set at .05, as setting any lower cut off points would be too conservative. It increases the possibility of type II error, as a result of which probable significant differences might failed to be noticed (Field, 2009; Nakagawa, 2004; Rothman, 1990). The criterion adopted for identifying outliers was three standard deviation from the mean. Although the data was normally distributed, a single outlier could be spotted looking at the boxplots, though it was not an extreme outlier (participant number 8 in control group in delayed post-test phase) (see Appendix Q).

Next, the assumption of homogeneity of covariances was assessed by Box's test of equality of covariance matrices for word association scores ($p = .52$), which was not violated. In addition, Mauchly's test of sphericity was run to check if the variances of different scores among the test variables were equal in the population. The results indicated that the assumption of sphericity was met for two-way interaction for the measures of word association, $\chi^2(2) = .91, p = 25$.

5.2.4. Comparison of within and between factors in word association test

In order to check the first research question considering the impact of semantic mapping on participants' structure and organization of mental lexicons, a two-way analysis of variance was used. The mixed between – within ANOVA was conducted with 'time' as the within-subjects variable at three levels -pre-test, immediate post-test, and delayed post-test. The between-subjects factor was 'group', separating the participants who took part in semantic mapping as experimental group and those who didn't receive any training on semantic mapping as control group. By application of mixed between – within ANOVA, it was possible to find out if there was a change in word association scores over the three times of testing (main impact for time), and also it compared the experimental and control groups in terms of their gains in scores (main impact for group). Finally, it was possible to check whether the change in the word association scores over time was different for the two groups (interaction effect).

Before looking into the descriptive statistics and running ANOVA, an independent sample t-test was used to compare the differences between experimental and control groups in terms of their starting points of word association test before the training sessions. The t-test revealed no significant difference across the two groups in their measures of word association knowledge, $t(30.47) = .41$, $p = .90$. Consequently, for the measure of word association, it can be concluded that the two groups were roughly equivalent in their levels of word association knowledge at the outset of the training.

Looking at the Table ***5.1*** of descriptive statistics, it shows that practicing vocabulary had an impact on both groups, while the mean scores and degree of retention after two weeks interval differ from one another.

Table 5.1 Descriptive Statistics for word association test

Test	Control	Experimental
Pre-test	187.69(43.43)	181.65(40.56)
Immediate post-test	205.50(36.71)	255.59(26.42)
Delayed post-test	202.47(28.66)	222.47(39.53)

Note. Standard deviations are presented in parentheses.

By looking at the means and standard deviations, it is clear that the participants in the experimental group could get higher scores due to semantic mapping training (M = 255.59, SD= 26.42) in comparison to the control group (M = 205.50, SD = 43.43) in the immediate post-test phase of testing. Further, the same trend can be reported by the comparison of the groups in the delayed post-test phase. As mentioned, in the delayed post-test phase, the performance of the participants in the experimental group (M = 222.47, SD = 39.53) surpassed the control group (M = 202.47, SD = 28.66). Looking closely into the means and standard deviations, it is suggested that although participants in both experimental and control groups scored higher in the immediate post-test in comparison to the pre-test, the scores in the delayed post-test are lower than the immediate post-test, while still higher than the pre-test. This suggests that neither of the groups had retained all the improvements after the two weeks interval, while the degree is different.

Conducting mixed between-within ANOVA, and looking at the significance values, a significance main effect for time was found $F(1, 31) = 23.91$, $p = .00$, partial $\eta 2 = .43$, which is considered to be a medium effect size (see Appendix R, for a discussion of the effect size). As reported the partial eta square for time is .436 indicating that 43.6% of the between subjects variance is accounted for the time. In addition, the main effect of group was also significant, $F(1, 31) = 4.52$, $p = .04$, partial $\eta 2 = .12$, which is a small effect size. The fact that grouping of the participants interacts significantly with the time of the testing indicates that participants in the experimental group and control group responded differently to three times of testing. As it has already been mentioned, comparison of the groups revealed no significance difference at the time of pre-test; hence, for the measure of word association, it can be concluded that the two groups were roughly equivalent in their levels of word association knowledge at the outset of the training, and accordingly, the difference between the two groups can be attributed to the different treatment periods they had with vocabulary practicing activities.

Overall, it was revealed that there was a significant interaction between the groups and the time of the test, $F\ (1, 31) = 8.83$, $p = .00$, partial $\eta 2 = .222$. The Global Effect size using partial eta squared was 0.22, which is approximately a small effect. The results specify that getting practiced with semantic mapping had a different effect on organization and structure of the mental lexicon, represented in word association scores, on three times of testing in different groups of experimental and control.

To investigate the significant interaction, series of follow-up t-tests were conducted. The pairwise comparisons for the main effect of time corrected using a Bonferroni adjustment revealed that for both groups of experimental and control, there was a significant difference ($p = .00$) from pre-test to immediate post-test ($M= 184.58$, $SD = 41.43$), and also from pre-test to the delayed post-test ($M = 212.67$, $SD = 35.66$). In order to find out where the differences lie, four paired sample t-tests are conducted. After training the experimental group with semantic mapping, the results of paired sample t-test indicate that there was a significant increase in the results of the word association test score, $t(16) = 7.89$, $p = .00$, from pre-test ($M = 181.65$, $SD = 40.56$) to the immediate post-test ($M = 255.59$, $SD = 26.42$), scoring higher in immediate post-test, with a large effect size (d = 2.16). In addition, comparing the results of the pre-test to the delayed post-test (M = 212.67, SD = 35.66) in the same experimental group, still indicate a significant increase, t(16) = 3.99, p = .00, d = .79, showing that although the mean score was reduced as the participants didn't retain the high score they received after the training sessions, there was still a significant difference between the pre-test and delayed post-test in favor of the delayed post-test. The large effect size from pre-test to immediate post-test (d = 2.16) and pre-test to the delayed post-test (d = .79) suggested a favorable impact of the semantic mapping technique.

A different trend was reported on the results of the control group, as it can be seen in both the descriptive table of the means (Table ***5.1***) and the following line graph (Figure ***5.1***). Looking at the means reveal that although the mean score was increased from pre-test (M = 187.69, SD = 43.43) to the immediate post-test (M = 205.50; SD = 36.71), the difference was non-significant, t(15) = 2.09, p = .06. In addition, comparing the results of the pre-test to the delayed post-test revealed a decrease in the means from pre-test (M = 187.69, SD = 43.43) to the delayed post (M = 202.47, SD = 39.53). Finally, checking the differences from pre-test to delayed post-test reveals no significant difference t(15) = 1.29, p = .21.

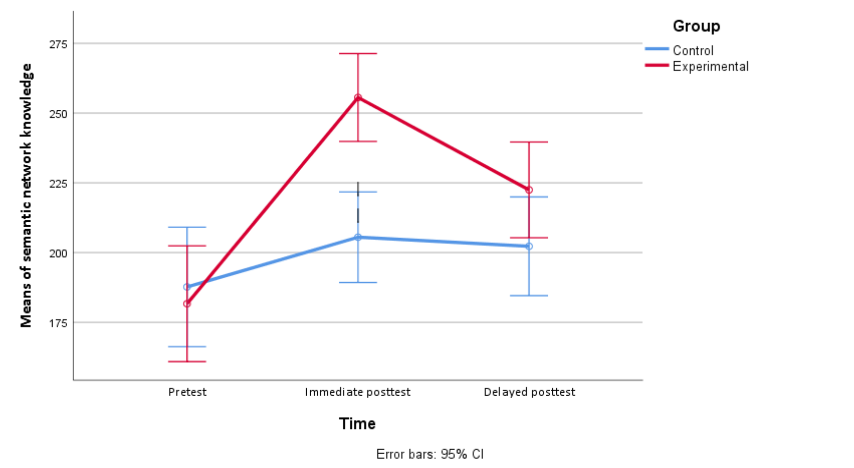


Figure 5.1 Comparison of the experimental and control groups at pre-test, immediate, and delayed post-tests of word association test

Afterwards, the pairwise comparisons for the main effect of group revealed that there was a significance difference between the two groups (p = .04). Accordingly, in order to find out in which part of the testing times the differences occurred, three independent sample t-tests were conducted. The first phase of testing, the pre-test, revealed no significant difference in the scores of participants in experimental (M = 181.65, SD = 40.56) and control group (M = 187.69, SD =

43.43), $t(30) = .41$, $p = .68$, while the results of the comparison of the scores in the immediate post-test phase showed that there was a significant difference between the experimental (M = 205.50, SD = 36.71) and control groups (M = 255.59, SD = 26.42), $t(30) = 4.51$, $p = .00$. Finally, in the last phase of delayed post-test, comparison of the control (M = 222.47, SD = 39.53) and experimental groups (M = 202.25, SD = 28.66), no significant difference was found $t(30) = 1.67$, $p = .10$.

5.2.5. Comparison of the response types in control and experimental groups

Besides the inferential statistics presented for the quantitative analysis of the results, the mean number of responses in different categories and subgroups was also calculated for both groups of experimental and control (see Appendix S). Firstly, the mean number of responses in the three main categories of not associated, loosely associated, and closely associated responses for the control group and experimental group in different test times of pre-test, immediate post-test and delayed post-test is discussed, following which various responses in subcategories are explored. Figure ***5.2*** illustrates an overview of the distribution of responses in the main categories and participants' preferences in three test times for the experimental and control groups.

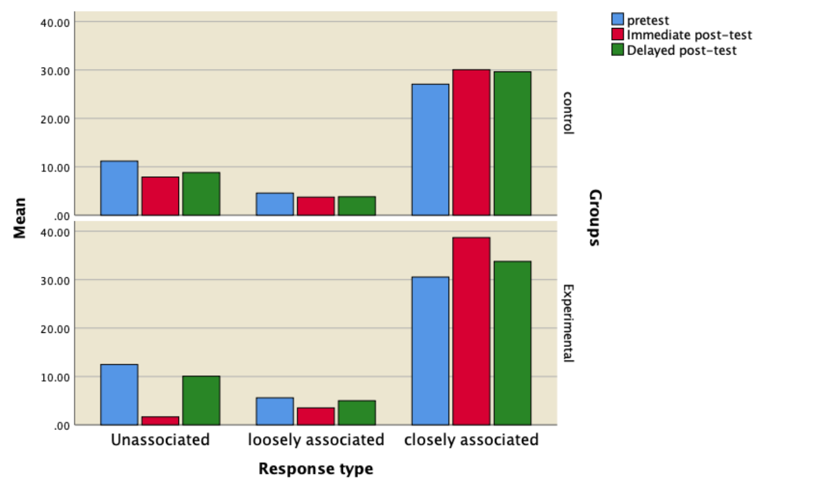


Figure 5.2 Mean number of responses in three main categories provided by participants in control and experimental groups

Looking into this figure, it is evident that both control and experimental groups show a clear preference for closely associated responses in comparison to the two other groups of not associated and loosely associative responses in all three phases of testing and managed to provide high number of responses in this group (approximately 3 times higher). As it can be seen, participants in the experimental group seemed to provide higher number of closely associated responses after the intervention in both immediate post-test and delayed post-tests; however, the mean number of responses provided by participants in the control group appears to be constant with very little changes.

The second group of response type with big number of associations falling into it, is the unassociated group of responses. Although this is the second biggest, the number of associations provided in this group is only one third of the closely associated responses in its highest point. Checking the bar graph, it shows that both groups provided approximately same number of responses in this category in the pre-test (a mean number of 11 responses), however, they didn't show equal response type preference in the immediate post-test phase. Smaller number of unqualified responses, providing no association to the cue word, was given by semantic mapping group right after the intervention, while the same is not true about the control group. However, the delayed post-test results show similar mean number of responses again for both groups.

Finally, the smallest group of responses can be attributed to the category of loosely associated responses. The mean number of loosely associated responses appear to be half number of unassociated responses for both experimental and control groups. It is in this last response type category that the smallest difference between the two groups can also be spotted. Examining the graph demonstrates that the semantic mapping group tended to produce lower number of loosely associated responses after training with semantic mapping in the immediate post-test phase, however, the changes didn't retain in the delayed post-test. In contrary, participants in the control group appear to produce consistent and static number of items in this group (with mean number of 4).

Secondly, the breakdown of the responses was also assessed in terms of the subcategories for both groups. The subcategories are examined with regard to the breakdown of the closely associated response category, the canonicality of responses, and consequently their level of frequency.

Initially, Figure ***5.3*** illustrates the subcategories of closely associated responses being divided into position-based, semantically related, and position/semantically related responses.

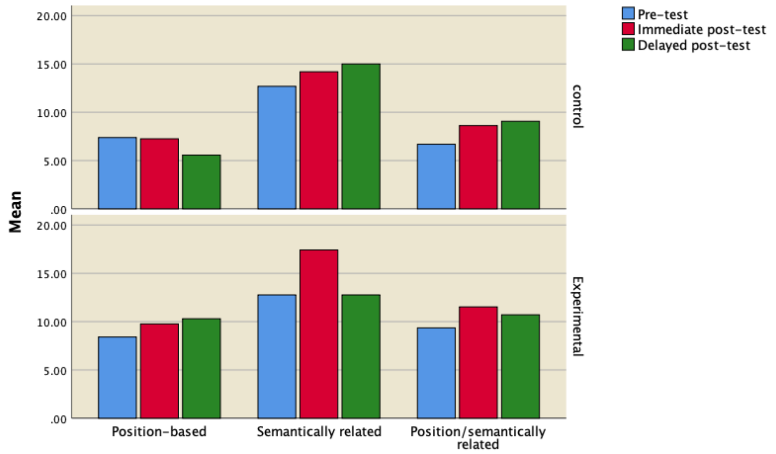


Figure 5.3 Mean number of responses in subcategory of closely associated responses provided by participants in control and experimental groups

As it can be seen, the two groups produced roughly close number of responses in the pre-test phase in all subcategories; however, participants in the semantic mapping group provided higher mean number of responses in all three subcategories after the period of intervention in the immediate post-test phase of testing. The biggest difference can be attributed to the semantically related responses, which constitute paradigmatic responses to the cue words, such as *jungle* eliciting the response *forest.* In addition, the results in the delayed post-test also demonstrate that although participants in the semantic mapping group managed to produce higher mean number of position-based responses as well as responses with a dual link of position/semantically related in comparison to the pre-training, number of semantically related responses appear to similar to the pre-test and it dropped from the immediate post-test phase (a mean of 13 responses).

Next, the following figure (Figure ***5.4***) represents changes in the number of canonical and non-canonical responses in control and experimental groups in three test times.

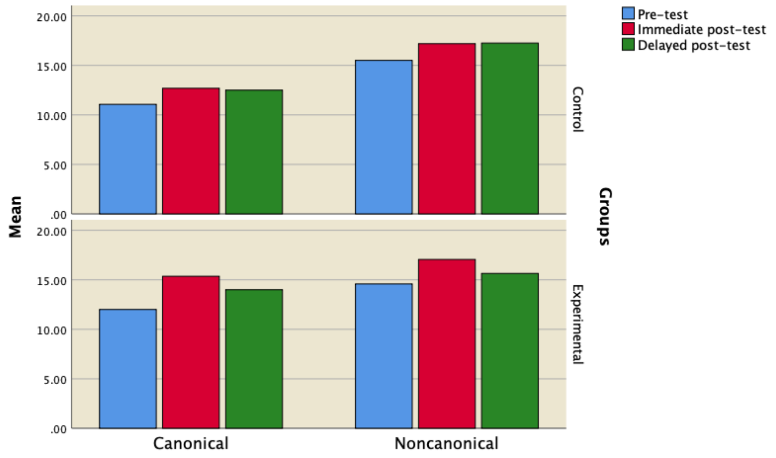


Figure 5.4 Mean number of canonical and noncanonical responses provided by participants in control and experimental groups

As it can be seen, although both groups tended to produce close number of canonical and noncanonical responses in the pre-test, changes in trend are clear in the immediate and delayed post-tests. The graph illustrates that both groups managed to produce higher number of canonical and noncanonical responses in the immediate and delayed post-tests to various degrees; however, this increase is more notable for the semantic mapping group in comparison to the control group.

Consequently, the last figure in this section (Figure *5.5*) demonstrates the mean number of high and low frequency responses for the two groups in three phases of testing.

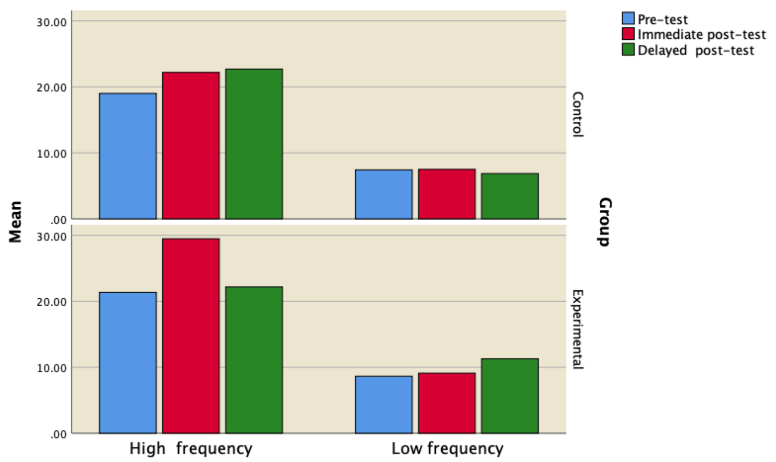


Figure 5.5 Mean number of high frequency and low frequency responses provided by participants in control and experimental groups

As it can be seen in Figure ***5.5***, the majority of responses for both control and experimental groups is associated to the higher frequency group of responses, and the number of low frequency is approximately two to three times lower. Both groups clearly gave higher number of high frequency responses in the immediate post-tests and delayed post-tests; however, changes in the control group were minimal. Considering the responses from the lower frequency level, the increase in the mean number in the experimental group after the process of semantic mapping is evident in both immediate post-test and delayed post-test; however, no change in the control group was found in this frequency level.

Consequently, the results of word association test along with the analysis of the response types provided in each group indicate variations in participants'' response behavior in the control and experimental group, which will be discussed and elaborated upon thoroughly in the next chapter (chapter 6).

5.3. Fluency measures

5.3.1. Overview

The main purpose of the second section of this chapter on fluency measures is to address the second research question regarding the impact of semantic mapping technique on L2 fluency of speech. The aim is to ascertain if the treatment helped participants to develop L2 fluency of speech as assessed by examining four measures of syllable run, phonation run, syllable duration, and finally silent pause duration per minute. The measures, selected out of the pool of fluency measures, are among the best indicators of L2 utterance fluency, reflecting the underlying cognitive processes involved in speech production. The rational for inclusion of each of these measures have been discussed thoroughly in chapter 3 (see Section 3.7.2.2).

Accordingly, in order to answer the second research question the following steps were taken. In the first place, the distribution of the data was assessed to choose between parametric and non-parametric statistical procedures, and then each of the fluency measures were analyzed separately. Each measure was firstly examined regarding its improvements and gains, pre-test to immediate post-test gains and pre-test to delayed post-test gains, between experimental and control groups. Next, the scores within each group were compared from pre-test to the immediate post-test and finally from pre-test to the delayed post-test.

5.3.2. Assumptions and choice of the appropriate statistical procedures

The very first step was assessing the outliers. In order to detect the outliers, both lower and higher interquartile were checked, while the higher interquartile range rule multiplier of three standard deviations was adopted to detect the extreme outliers (Carling, 2000; Hoaglin & Iglewicz, 1987; Wilcox, 2017). Checking the boxplots and the interquartile range for the fluency measures by the use of boxplots, for two of the measures of silent pause duration (participant number 32 and 23 in delayed post-test), and syllable run (participant number 10 in pre-test and number 19 in pre-test and immediate post-test), extreme outliers were identified. In order to treat the outliers, it was decided not to apply transformations, as it can change the relationship and difference between the variables; hence, it would be difficult to interpret the results (Field, 2009; Newton & Rudestam, 1999). Instead, each of the fluency measures, for each of which outliers were detected, were once analyzed without the inclusion of the outliers to check if there were any significant differences with and without the inclusion of higher interquartile outliers. Based on the results of the analysis excluding the outliers, there were no significant changes in the outcomes, hence, it was decided to keep the data in the analysis (see Appendix T).

Next, measures of speech fluency were assessed for the assumption of the normality to decide on parametric or non-parametric tests. Looking at the histograms and Q-Q plots (see Appendix Q), it was suggested that the data was not normally distributed, though in order to ascertain, the results of Shapiro-Wilk test was also consulted (see Appendix P). Consequently, the data was analyzed using the non-parametric tests of Wilcoxon and Mann-Whitney tests, the summary of which is presented in the following table (Table ***5.2***).

Table 5.2 Non-parametric tests utilized in the study to analyze the fluency data

Test	Between/Within-groups	Levels	Purpose
Mann-Whitney *U*	Between-group	2	Compare the experimental and control groups at the pre-test
Mann-Whitney *U*	Between-group	2	Compare the experimental and control groups at pre-test to immediate post-test gains

Mann-Whitney *U*	Between-group	2	Compare the experimental and control groups at pre-test to delayed post-test gains
Wilcoxon signed rank	Within-group	2	Analyze the experimental group's fluency measures from pre-test to the immediate post-test
Wilcoxon signed rank	Within-group	2	Analyze the experimental group's fluency measures from pre-test to the delayed post-test
Wilcoxon signed rank	Within-group	2	Analyze the control group's fluency measures from pre-test to the immediate post-test
Wilcoxon signed rank	Within-group	2	Analyze the control group's fluency measures from pre-test to the delayed post-test

It is worth mentioning that that the alpha level was set at $p = .05$ for the fluency measures analysis.

This section presents the results of gain scores of the participants in respect to different measures of fluency. The results for each measure are divided into two parts. The first part deals with the comparison of the fluency measures between experimental and control groups over each of the test phases, and the second part reports the gains from pre-test to immediate post-test and also from pre-test to delayed post-test within each of the groups.

5.3.3. Analysis of the results for Syllable Run

5.3.3.1. Comparison of the gains in syllable run between the control and experimental groups

In order to assess whether any improvements in the syllable run measure of utterance fluency had been gained after the intervention, the results of each of the groups are to be analyzed at three different points of testing, pre-test, immediate post-test, and delayed post-test. At first, the results of pre-test to immediate post-test gains are analyzed to track any improvements after the intervention period of time, and secondly pre-test to delayed post-test results are to be compared with each group. Consequently, the results in this section report if any significant changes can be found after the participants getting trained in semantic mapping in the measure of syllable run in the two phases of post-testing in comparison to the pre-test, and also the results gained in the control group in the same manner.

The mean scores and standard deviations for the two groups of experimental and control at pre-test, immediate post-test, and delayed post-test gains on measure of syllable run is provided in Table ***5.3***;

Table 5.3 Descriptive statistics syllable run

Test time	Control	Experimental
Syllable run pre-test	5.39 (1.74)	6.07 (1.63)
Syllable run immediate post-test	6.17 (1.73)	8.35 (1.86)
Syllable run delayed Post-test	5.92 (1.75)	7.15 (1.83)

Table 5.3 and also Figure 5.6 show that the measure of syllable duration was increased from pre-test (*M* =6.07, *SD* = 1.63) to the immediate post-test (*M* =8.35, *SD* = 1.86) in the experimental group. Furthermore, same trend of increase in the measure can be detected from pre-test to the delayed post-test (*M* =7.15, *SD* = 1.83) of the same group. Similar trend can be observed in the control group from pre-test (*M* =5.39, *SD* = 1.74) to immediate post-test (*M* =6.17, *SD* = 1.73), and also the delayed post-test (*M* =5.92, *SD* = 1.75).

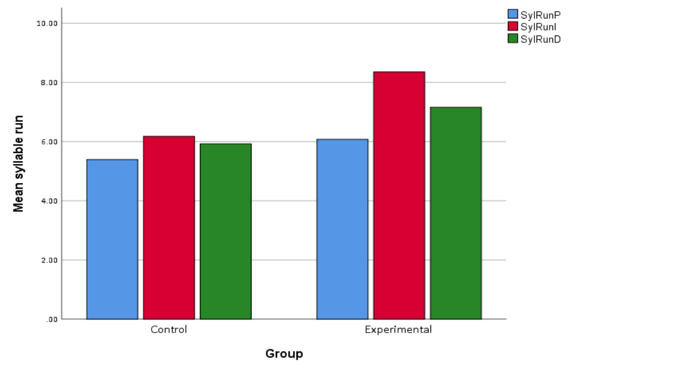


Figure 5.6 Comparison of pre-test, immediate post-test, and delayed post-test means of syllable run between control and experimental groups

Comparing the two groups regarding their level of performance at the very early stage of the study, the pre-test, the non-parametric test of Mann-Whitney *U* test revealed that there was no significance difference between experimental (*Mean Rank* = 14.69) and control (*Mean Rank* = 11.31) groups in regard to their syllable run measures, *U* = 99.00, *z* = -1.33, *p* = .19, two-tailed, while the results indicates that the syllable run measure in experimental group at the immediate post-test phase (*Mean Rank* = 22.35, *n* = 17) was significantly different from that of the control group (*Mean Rank* = 11.31, n = 16), U = 45.00, z = -3.27, p = .00, two-tailed. This can be described as "medium" (r = .56) effect size. In addition to the significant difference reported between the immediate post-tests of the two groups, comparison of the groups in the delayed post-test phase also revealed a significant difference U = 74.00, z = -2.23, p = .02, two-tailed. This can be described as "small" (r = .38) effect size.

Mann-Whitney U tests are conducted on syllable run measure to assess the gain from pre-test to immediate post-test and also from pre-test to the delayed post-test in the two groups. The results indicate that the syllable run measure for the experimental group (*Mean Rank* = 21.88) was significantly higher than those of the control group (*Mean Rank* = 11.81) in pre-test to immediate post-test gains, U = 53.00, z = -2.99, p = .00, two-tailed. This can be considered as "medium" (r = .52) effect size. In addition to that, the pre-test to delayed post-test gains revealed a significant difference for the experimental (*Mean Rank* = 20.18) and control group (*Mean Rank* = 13.63), U = 82.00, z = -1.94, p = .05, two-tailed, r = .33 on the direction of the experimental group.

5.3.3.2. Comparison of scores of syllable run within experimental and control groups

In order to analyze the data within each group, the Wilcoxon signed rank test was utilized. The Wilcoxon signed rank test indicated that participants in the experimental group were significantly different after getting trained with semantic mapping in the immediate post-test (M = 8.35, SD = 1.86) in comparison to the pre-test (M = 6.07, SD = 1.63) on measure of syllable run, T = 0, z = -3.62, N – Ties = 17, p = .00, two-tailed, with the large effect size, r = .87.

Similar to the experimental group, a significant difference was reported on scores in syllable run from pre-test (M = 5.39, SD = 1.74) to immediate post-test (M = 6.17, SD = 1.73) in the control group, T = 23, z = -2.32, N – Ties = 16, p = .02, two-tailed, with the effect size of r = .58. The results indicate that both groups benefited although the effect size was different and was bigger in favor of the experimental group.

In order to assess if any of the higher scores had been retained at two weeks after the intervention which was followed by immediate post-tests, the results of the delayed post-tests were consulted and were compared with the results of the pre-tests. Both groups reached a significant difference in the immediate post-test, while the control group participants did not retain what they had acquired as reported by the delayed post-test measure ($M = 5.92$, $SD = 1.75$) of syllable run, $T = 41$, $z = -1.39$, N – Ties = 16, $p = .16$, two-tailed, though it was not the case for the experimental group $T = 13$, $z = -3.00$, N – Ties = 17, $p = .00$, two-tailed. Relative to their pre-training phase, 15 participants in the experimental group scored higher on the measure of syllable run out of 17 in total (Sum of ranks = 140), whilst only 10 participants in the control group scored higher (Sum of Ranks = 95). This effect can be considered "large", $r = .72$.

The results suggest that both experimental and control groups benefited from the intervention sessions, while as their performance in delayed post-test can clearly indicate, the gains in the control group were only temporary and they did not retain it for a longer period of time.

5.3.4. Analysis of the results for phonation run

5.3.4.1. Comparison of the gains in phonation run between the control and experimental groups

The second utterance fluency measure to be assessed is the phonation run, which was calculated by dividing the phonation duration by the number of silent pauses. In order to find out if the participants in the two groups had any improvements and differences in this measure, the results were analyzed by Wilcoxon signed ranked test to find any differences within each group of experimental and control. The descriptive statistics are presented in Table ***5.4***;

Table 5.4 Descriptive statistics on phonation run

Test time	Control	Experimental
Phonation run pre-test	1.75 (.41)	2.04 (.52)
Phonation run immediate post-test	1.94 (.40)	2.40 (.53)
Phonation run delayed Post-test	1.94 (.52)	2.16 (.54)

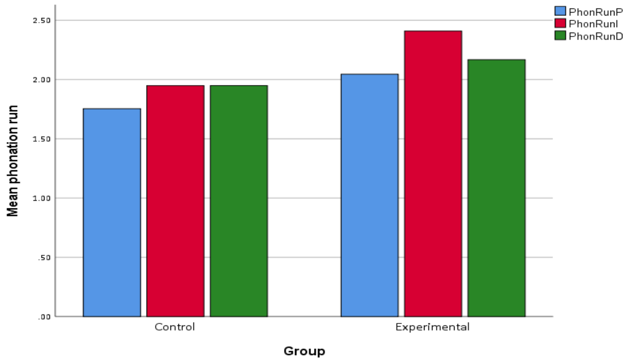


Figure 5.7 Comparison of pre-test, immediate post-test, and delayed post-test means of phonation run between control and experimental groups

Table ***5.4*** and figure 5.7 present the results of descriptive statistics of the phonation run measure. Measure of phonation run was increased from pre-test (*M* =2.04, *SD* =.54) to the immediate post-test (*M* =2.40, *SD* =.53) in the experimental group. Also, same trend of raise in phonation run can be detected from pre-test to the delayed post-test (*M* =2.16, *SD* =.54) in the same group. As in the control group, similar trend of increase can be seen from pre-test (*M* =1.75, *SD* =.41) to immediate post-test (*M* =1.94, *SD* =.40), while no increase in this measure was detected from pre-test to the delayed post-test means and the means was same as the pre-test (*M* =1.94, *SD* =.52).

To begin with, level of performance of the two groups at the pre-test was compared via Mann-Whitney *U* test. The results revealed that there was no significance difference between experimental (Mean Rank = 19. 21) and control (Mean Rank = 14.74) groups in regard to their phonation run measures, $U = 100.00$, $z = -1.29$, $p = .20$, two-tailed.

Based on the results gained from Mann-Whitney *U* test comparing the experimental and control groups from pre-test to the immediate post-test level, there was a significant difference between the performance of the two groups in the measure of phonation run, $U = 77.00$, $z = -2.12$, $p = .03$, two-tailed, $r = .37$, in the direction of the experimental group (*Mean Rank* = 20.47). Although, a significant different for the gains from pre-test to the immediate post-test was reported, the pre-test to delayed post-test results revealed no significant difference between the experimental (*Mean Rank* = 19.65) and control (*Mean Rank* = 14.19), $U = 91.00$, $z = -1.62$, $p = .11$, two-tailed.

Accordingly, it can be concluded that although the two groups experienced various degree of changes in the measure of phonation run, revealing changes in the length of phonation duration between silent pauses, from pre-test to the immediate post-test right after the intervention, they reached approximately similar gains from pre-test to delayed post-test after the time interval, indicating that the changes and gains could not have been retained in the experimental group.

5.3.4.2. Comparison of scores of phonation run within experimental and control groups

Observing the results of the Wilcoxon signed rank test, a significant difference can be traced in the performance of the participants in the experimental group from the pre-test to the immediate post-test, $T = 10$, $z = -3.14$, $p = .00$, two-tailed, whilst the control group did not reach any significant difference in this phase, $T = 31$, $z = -1.91$, $p = .06$, two-tailed. Comparing the performance of the experimental group from pre-test to immediate post-test, 13 participants received greater measure of phonation run out of 17 participants in total (*Mean Rank* = 11), while only 4 had a performance with lower measure in the immediate post-test phase (*Mean Rank* = 2.5), which accounts for a "big" effect of $r = .77$.

However, analysis of the results from pre-test to the delayed post-test via Wilcoxon signed rank indicates that there was no significant difference in neither the experimental group performance, $T = 68$, $z = -.40$, $p = .68$, nor the control group, $T = 33$, $z = -1.81$, $p = .07$, two-tailed.

5.3.5. Analysis of the results for Syllable Duration

5.3.5.1. Comparison of the gains in syllable duration between the control and experimental groups

The third measure of utterance fluency to be analyzed is syllable duration, which is an inversion of the articulation rate. The analysis takes into account both improvements and gains within each of the groups in addition to the comparison between groups of control and experimental receiving different training and practice type with words. Table ***5.5*** presents the mean and standard deviation for each of the groups in each testing phase;

Table 5.5 Descriptive statistics for syllable duration

Test time	Control	Experimental
Syllable duration pre-test	.27 (.01)	.26 (.01)
Syllable duration immediate post-test	.26 (.00)	.21 (.00)
Syllable duration delayed Post-test	.27 (.01)	.22 (.01)

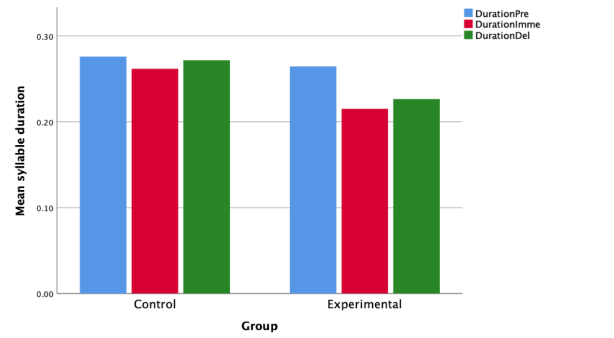


Figure 5.8 Comparison of pre-test, immediate post-test, and delayed post-test means of syllable duration between control and experimental groups

The results of the descriptive statistics for the measure of syllable duration in table 22 indicates that the mean of syllable duration measure was decreased from pretest (*M* =.26, *SD* =.01) to the immediate posttest (*M* =.21, *SD* =.00) in the experimental group. Furthermore, the mean comparison of the pretest and delayed posttest shows that the mean score of syllable duration was again decreased (*M* =.22, *SD* =.01). Although a decrease in the means in the experimental group was experienced, the results of the descriptive statistics of the control group were not alike. The mean score of the pretest (*M* =.27, *SD* =.01) was very close, though slightly higher than the immediate posttest (*M* =.26, *SD* =.00), and finally, as for the delayed posttest, identical results with the pretest was revealed (*M* =.27, *SD* =.01).

Comparing the two groups in the pretest with Mann-Whitney *U* test revealed that there was a significance difference between experimental (Mean Rank = 14.68) and control (Mean Rank = 19.47) groups in regard to their syllable duration measures, $U = 96.50$, $z = -1.42$, $p = .15$, two-tailed. Comparison of the two groups of experimental and control in regard to their gains in the measure of syllable duration in pretest to immediate posttest phase revealed that there was a significant difference between the groups, $U = 69.00$, $z = -2.41$, $p = .01$, two-tailed. In addition to that the same trend of results was found comparing the groups in pretest-delayed posttest phase, $U = 80.50$, $z = -1.99$, $p = .04$, two-tailed.

5.3.5.2. Comparison of scores of syllable duration within experimental and control groups

The results of Wilcoxon signed rank test indicates that there was a significant difference in the experimental group both from pre-test to the immediate post-test and also from the pre-test to delayed post-test, indicating that the training sessions with semantic mapping practicing the words, helped them in their utterance fluency measure of syllable duration, that is they could produce more syllables in the time spent speaking excluding all pauses. The results indicated gains from pre-test to immediate post-test, $T = 0$, $z = -3.62$, $p = .00$, two tailed. Relative to their pre-training rankings, 17 participants ranked higher in this measure after attending the intervention sessions (Sum of Ranks = 153), while none of the participants gained lower gain (Sum of the Ranks = 0), which can be considered as very large effect size of $r = .87$. Furthermore, very close results were reported for the second phase of analysis, which was gains from pre-test to delayed post-test, $T = 4.50$, $z = -3.19$, $p = .00$, two tailed, $r = .77$.

Contrary to the experimental group, the control group reached no significant differences in any of the phases of the testing. Wilcoxon signed rank test reports neither any significant changes in pre-test to immediate post-test $T = 6$, $z = -1.96$, $p = .06$, two tailed, nor from pre-test to delayed post-test $T = 7$, $z = -.62$, $p = .53$, two tailed, confirming no difference in the measure of syllable duration in this group of participants after attending six sessions of practicing the words.

5.3.6. Analysis of the results for silent pause duration

5.3.6.1. Comparison of the gains in silent pause duration between the control and experimental groups

The last measure of speech fluency to be analyzed is the silent pause duration, which is the duration of the silent pauses in speech. This section looks into the differences between the test times within control and experimental groups separately, and secondly, a comparison is made between the two groups in pre-test to immediate post-test and pre-test to delayed post-test. The following table presents the mean and standard deviation for each of the groups in each testing phase;

Table 5.6 Descriptive statistics on silent pause duration

Test time	Control	Experimental
Silent pause duration pre-test	.48 (.75)	.55 (.13)
Silent pause duration immediate post-test	.46 (.09)	.48 (.11)
Silent pause duration delayed post-test	.49 (.10)	.51 (.12)

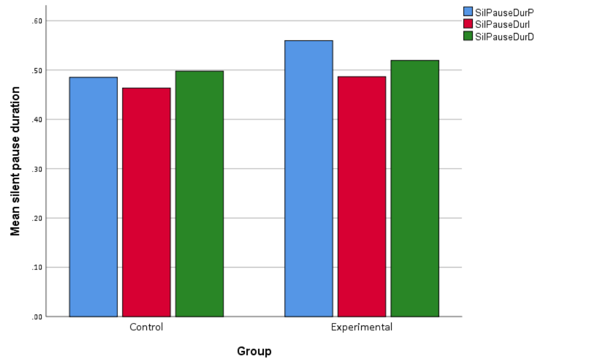


Figure 5.9 Comparison of pre-test, immediate post-test, and delayed post-test means of silent pause duration between control and experimental groups

Examining Table ***5.6*** and bar chart Figure 5.9 of the silent pause duration in both groups for each of the test times shows that the participants in the experimental group slightly decreased in the duration of the silent pauses in their speech from pre-test (*M* = .55, *SD* =.14) to immediate post-test (*M* = .48, *SD* =.11), and also the control group experienced similar decrease to a lesser degree. The results of the delayed post-test report that although the control group had a lesser duration of silent pauses in the immediate post-test, but it was not persistent, and the delayed post-test reported the same mean as the pre-test. The experimental group was also similar though they retain more and the result of the delayed post-test (*M* = .51, *SD* =.12) was still lower than the pre-test, but still it was not significantly different.

The two groups were compared on the pre-test in order to check if there are any differences between them in the measure of silent pause duration. The results revealed that there was a significance difference between experimental (Mean Rank = 19. 21) and control (Mean Rank = 14.74) groups in regard to their silent pause duration measures, $U = 75.00$, $z = -2.19$, $p = .03$, two-tailed. Next, comparing the two groups with Mann-Whitney U test suggests that there was no significant difference on pre-test to immediate post-test gains in achieving shorter silent pause duration between the control and experimental groups, $U = 86.00$, $z = -1.80$, $p = .07$, and also no significance difference was reported on pre-test to delayed post-test phase between the two groups, $U = 97.00$, $z = -1.4$, $p = .16$, two-tailed.

5.3.6.2. Comparison of scores of silent pause duration within experimental and control groups

Although looking back at the table of descriptive statistics (Table ***5.6***) show increase in the measures from pre-test to immediate post-test and also pre-test to delayed post-test, the results of the Wilcoxon signed rank test on experimental group measure of silent pause duration indicates no significant difference in neither of the phases of pre-test to immediate post-test, $T = 40$, $z = -1.72$, $p = .08$, and pre-test to delayed post-test, $T = 46$, $z = -1.44$, $p = .14$, two tailed. Relative to their pre-training, 11 participants in this group had shorter duration of silent pauses in the immediate post-test (Sum of Ranks = 113) and also the delayed post-test (Sum of Ranks = 107), whilst only 6 of them had longer duration of pauses in the immediate (Sum of Ranks = 40) and delayed (Sum of Ranks = 46). The same trend could have been found in the control group from pre-test to immediate post-test, $T = 61$, $z = -.36$, $p = .71$, and pre-test to delayed post-test, $T = 56$, $z = -.62$, $p = .53$, two tailed.

5.4. The relationship between word association test and fluency measures

In order to find out if there was any relationship between word association scores, representing the organization and structure of the mental lexicon, and measures of speech fluency, series of Spearman rank order correlations were conducted. Table 24 displays the correlation between the variables in immediate post-test phase;

Table 5.7 Correlations between word association results and fluency measures in immediate post-test

	Group			Syllable run	Phonation run	Syllable duration	Silent pause duration
Spearman's rho	Control	Word association	Correlation Coefficient	.91	.10	-.34	-.35
			Sig. (2-tailed)	.73	.70	.18	.89
	Experimental	Word association	Correlation Coefficient	.68**	.48*	-.14	-.22
			Sig. (2-tailed)	.00	.05	.57	.39

The results of Spearman's rho correlation analysis between the single measure of word association score and four measures of fluency in the immediate post-test revealed a strong correlation between free productive word association results and two out of four measures of fluency in the current study. That is, moderate positive correlation between word association test and syllable run, $r = .68$, $p = .00$ and phonation run, $r = .48$, $p = .05$, two tailed, $N = 17$, was reported on immediate post-test results. In addition, exploring the results of correlation analysis indicate a weak negative non-significant correlation between word association score and syllable duration, $r = -.14.$, $p = .57$, and mean duration of silent pause, $r = -.22$, $p = .39$.

The next step of correlation analysis was conducted on the results of word association test and measures of speech fluency in the delayed post-test phase (Table ***5.8***). The results are in accordance with the immediate post-test, reporting a moderate positive correlation between word association score and syllable run, $r = .48$, $p = .04$, and phonation run, $r = -.62$, $p = .00$. Besides, a moderate negative correlation was also reported between word association and syllable duration, as a pure measure of speed, $r = -.38$, $p = .12$; however, no correlation and association between the scores of word association and mean duration of silent pauses could have been found.

Table 5.8 Correlations between word association results and fluency measures in delayed post-test

	Group			Syllable run	Phonation run	Syllable duration	Silent pause duration
Spearman's rho	Control	Word association	Correlation Coefficient	.20	.36	-.02	-.07
			Sig. (2-tailed)	.44	.16	.93	.77
	Experimental	Word association	Correlation Coefficient	.48*	.62**	-.38	.19
			Sig. (2-tailed)	.04	.00	.12	.46

5.5. Summary

This chapter presented a quantitative analysis of the results to explore the extent of the impact of the semantic mapping technique on various measures of L2 fluency of speech as well as the organization of the bilingual mental lexicon. More specifically, after providing an overview of the whole chapter (section 5.1), the second section presented the preliminary analysis as well as results of running the repeated measure ANOVA for the word association test (section 5.2.4). This was followed by the analysis of the means of various type of responses in each phase of the testing for control and experimental groups (section 5.2.5). Subsequently, the third section dealt with the results of the analysis of the fluency measures, while each set was presented separately in a subsection (section 5.3). Finally, the results of the correlation analysis of the results of word association test and measures of fluency was presented in the last section (section 5.4). Next chapter presents a discussion of the gained results in great detail.

6. Chapter 6: Discussion of findings

6.1. Overview

This chapter discusses the main findings of this research presented in chapter 5 regarding the research questions, while incorporating the literature already reviewed. This chapter is organized into three main sections. The first section (see Section 6.2) summarizes the key findings of the current research to provide a clear picture of the trends at a glance, which is followed by two separate sections on discussion of the impact of semantic mapping technique on structure of the lexicon and network knowledge (see Section 6.3) and also separate measures of L2 fluency of speech (see Section 6.4). Finally, a summary of the chapter is presented (see Section 6.5).

6.2. An overview of the key findings

This study reports on one of the first intervention studies that sets out to examine the impact of semantic mapping on the organization of the bilingual mental lexicon and L2 utterance fluency. It is unique in its application of semantic mapping as a way to practice already known lexical items with the goal of making changes to the underlying network structure in different levels of knowledge representation, which in turn appears to have an impact on fluency of speech. In general, this study contributes to our understanding of how the network knowledge, which is a conceptualization of the depth of lexical knowledge, contributes to L2 fluency of speech.

As already stated in chapter one, the aim of the current research is threefold;

- ✓ Firstly, to explore the impact of semantic mapping, as a vocabulary training technique, on the organization of bilingual mental lexicon, by assessing changes and developments of EFL learners' network knowledge, the results of which is summarized in Table ***6.1***;

Table 6.1 Summary of the word association results between control and experimental groups

	Word association test
Effect of time	Sig
Effect of group	Sig
Interaction	Sig

Sig= a statistically significant difference at 0.05

Not sig= No significant difference

The table above shows that the results are fairly in line with the literature with regards to the productive word association test, implying that alternations in the ways students practice vocabulary by the application of semantic mapping technique can create modifications and changes in the structure of network knowledge, being manifested in the increase in the scores of productive word association tests. The results also indicate that semantic mapping technique has a significant positive effect on changing the structure and organization of the mental lexicon, more specifically the network knowledge of that particular domain, both at the immediate post-test and even after the interval of two weeks in the delayed post test phase.

- ✓ Secondly, to investigate the effect of semantic mapping, as a vocabulary training technique, on various measures of L2 utterance fluency to evaluate the underlying cognitive processes, the results of which in pre-test to immediate post-test and pre-test to delayed post-test gains in control and experimental groups are summarized in Table ***6.2***;

Table 6.2 Summary of the results from pre-test to immediate post-test and pre-test to delayed post-test within the control and experimental groups

	Syllable Run		**Phonation Run**		**Syllable duration**		**Silent pause duration**	
	Pre-test to immediate post-test	Pre-test to delayed post-test	Pre-test to immediate post-test	Pre-test to delayed post-test	Pre-test to immediate post-test	Pre-test to delayed post-test	Pre-test to immediate post-test	Pre-test to delayed post-test
Control Group	Sig	Non-sig	Non-sig	Non-sig	Non-sig	Non-sig	Non-sig	Non-sig
Experimental Group	Sig	Sig	Sig	Non-sig	Sig	Sig	Non-sig	Non-sig

Sig= A statistically significant difference at 0.05 in favor of the experimental group

Non-sig = No significant difference between the two groups

; in addition to the results of comparison across the groups of control and experimental in table 6.3;

Table 6.3 Summary of the results of between group differences regarding gains in fluency measures

	Syllable Run	Phonation Run	Syllable duration	Silent pause duration
Pretest to immediate posttest	Sig	Sig	Sig	Non-sig
Pretest to delayed posttest	Sig	Not sig	Sig	Non-sig

Sig= a statistically significant difference at 0.05 in favor of the experimental group
Non-sig = No significant difference between the two groups

Regarding fluency measures, it is evident that application of semantic mapping technique results into changes both within (Table ***6.2***) and between the groups (Table 6.3), though the degree of changes in various measures differs. Initially, within group analysis of the results revealed a positive significant effect of semantic mapping technique on three measures of syllable run, syllable duration, and phonation run from pre-test to immediate post-test; however, no significant impact was reported on the measure of silent pause duration in this same phase. Subsequently, the results of within group differences from pre-test to delayed post-test were very close for the same group with the exception of phonation run, which revealed no significant difference in the delayed phase. Following this, the results of between group differences confirmed that semantic mapping technique had a significant positive effect on pre-test to immediate post-test gains in measures of syllable run, phonation run, and syllable duration, while the gains were sustained only for two measures of syllable run and syllable duration to delayed post-test.

- ✓ and finally, to investigate if there was a relationship between in organization of the mental lexicon, measured by free productive word association scores, and each separate measure of utterance fluency, the results of which is presented in Table **6.4**;

Table 6.4 Correlation between word association and fluency measures in immediate and delayed post-test phases

	Control group		Experimental group	
	Immediate post-test phase	**Delayed post-test phase**	**Immediate post-test phase**	**Delayed post-test phase**
Syllable Run	No correlation	No correlation	Correlated	Correlated
Phonation Run	No correlation	No correlation	Correlated	Correlated
Silent pause duration	No correlation	No correlation	No correlation	No correlation
Syllable duration	No correlation	No correlation	No correlation	No correlation

In the last part (see Section 5.4), the relation between the overall word association score and fluency measures were reported. The results for the experimental group indicated that word association scores correlated strongly with the two fluency measures of syllable run and phonation run, while the correlation was weaker with syllable duration and mean duration of silent pauses. Considering the control group, no correlation between the measures were reported.

In light of the full range of data collected in the current study and drawing on the findings of the existing literature, possible explanations for the finding are presented in the following sections.

6.3. Semantic mapping and organization of the mental lexicon

This section presents a discussion of the results of the free productive word association test, analyzed from a quantitative perspective, to provide a clear understanding of the underlying processes and changes related to network knowledge. First, drawing on the quantitative analysis in section 5.2, this chapter provides an estimate of the overall productive network knowledge of each individual for a specific stratum of the lexicon, based on which a trend in favor of semantic mapping is reported. More specifically, as shown in Table ***6.1***, the findings show an increase in word association scores after the period of training in semantic mapping for the experimental group. In contrast, there was no significant impact in scores for the control group. In addition, it is worth exploring the differences within and between the groups from a more descriptive perspective, as discussion of different response types and tracing the changes can provide more useful insights into the underlying differences in network knowledge.

The results of applying semantic mapping in this study extend understanding of the value of this technique as an instructional tool for the teaching of new words (e.g. Pittelman et al., 1985; Sánchez, 2004; Stahl & Vancil, 1986), and supports the development of knowledge of known words. This might be explained by the fact that theories of L2 lexical development suggest that the acquisition of knowledge of vocabulary is incremental in nature, occurring within a network of connections and associations of the bilingual mental lexicon (Jiang, 2000, 2002, 2004; see Section 2.4.2). In other words, mastery of word knowledge is a matter of degree, from unknown to partial, and then to precise (Henriksen, 1999; Henriksen & Haastrup, 1998; Hunt & Beglar, 2005; Joe, 2010; Nation, 2001; Schmitt, 1998b, 2000, 2010; Stahl, 2009; Wolter, 2001). Thus, "vocabulary learning is not an all-or-nothing piece of learning but is rather a gradual process of one meeting with a word adding to or strengthening the small amount of knowledge gained from previous meetings" (Nation, 2001, p. 155). Drawing on the literature and as reflected in this study, it is possible to claim that various types of connections and associations appear to be generated as a result of the incremental nature of vocabulary acquisition and degree of word knowledge (Henriksen, 2008; Namei, 2002, 2004; Soderman, 1993; Wolter, 2001), which in turn could explain the variety in the number and types of connections and associations exhibited by the participants.

Evidence from in-depth analysis of the results regarding word association provides support for this explanation. First, the results of the pre-test phase are examined, discussing the structural features of the lexicon prior to the intervention phase. This is followed by tracking the changes and shifts in the organizational properties of the mental lexicon after the intervention from both quantitative discussion of the overall word association score as well as descriptive perspectives of the type of connections. The changes are discussed in terms of shifts in various types of associated responses (unassociated, loosely associated, and closely associated), in addition to alterations in the canonicality and frequency level of responses provided.

6.3.1. Structural properties of the lexicon prior to the intervention in control and experimental groups

Exploring the quantitative results in the pre-test phase revealed no significant differences in terms of the overall productive word association scores between the two groups, implying that almost all participants were at a similar stage of lexical development, with similar structural properties in terms of connectivity and the strength of the connection. This can further be confirmed by an in-depth discussion of the results in terms of the types of associations provided.

Examining the responses in the pre-intervention phase revealed that a high percentage, approximately 40%, were either not associated or only loosely associated with the cue words. Such a high number cannot be attributed to total lack of knowledge of the lexical items, as the cues were chosen from a list of known words (see Section 4.3.1.2). In addition, there is further confirmation as none of the cue words were missed out and no response given, nor was the cue simply repeated, at any level of testing, indicating that the lexical items were already known to the participants.

Accordingly, drawing on the literature in L2 lexical development (e.g., Jiang, 2000; 2004), as the types of associations are determined by the degree of word knowledge, it can be claimed that the high percentage of unqualified responses is an indication of many words that were either barely familiar or only partially known to the participants (this could be the result of such items being at the initial level of learning or due to the way in which lexical items are learnt in the Iranian language learning context, that is, through translation and the use of bilingual dictionaries, resulting in stronger L1–L2 linkage at the lexical level). Therefore, as it is not possible to retrieve the meaning of words automatically at this level of lexical knowledge and it can only be accessed through a very weak L1–L2 lexical connection (Jiang, 2000), the participants were not able to provide many responses with either a syntactically or semantically based association, but rather the responses available were unqualified and loosely associated, at the periphery of the network (Namei, 2002, 2004).

The first evidence of lexical items in the formal stage of development, which can be associated with a limited developed network of lexical representations, comes from the high number of unqualified responses, namely examples, names, and ragbags. First, the category of examples and names clearly shows that content knowledge was minimal, and the items had to be retrieved consciously although the time restriction for the task made this difficult. Therefore, the participants fell back on providing examples or names for the cues instead. In addition, consulting the retrospective short interviews (see Section 3.5.4.2) to find out more about the ragbags, it appears that most responses categorized as ragbags had neither semantic nor syntactic connections to the cue; rather, the links mostly appeared to be created in an idiosyncratic manner, i.e., established based on some random memories that were meaningful only to the individual in question, but with no linguistic basis. For example, in one case, the stimulus word *jungle* was associated with the response *my father* as the participant had recently travelled with his father to the north of Iran and camped in a jungle for a night.

Further evidence comes from the representation of the loosely associated responses. The two main types in the pre-test phase were form-related/clangs, such as *jungle* eliciting *jumble*, and two-step associated responses, entailing association with the cue through an intermediate association (e.g., *sail* and *buy* associated as result of the mediation of *sell*). Such responses could again be due to the fact that the only formal specifications (phonological/orthographic) are readily accessible as information for such items, thus resulting in the activation of phonologically associated responses. This is in accordance with the results of an earlier study on L1 and L2 mental lexicons conducted by Namei (2004), who posited that “unfamiliar or less familiar words elicit associations that are phonologically similar to the stimulus words” (p. 380). The participants in Namei’s study all produced clangs, irrespective of their L1 and their L2 level of language proficiency; however, those in lower grades supplied more phonologically associated responses than those in higher grades, suggesting that the production of clangs is significantly affected by the degree of word knowledge.

Thus, as outlined above, a high percentage of the lexical items appeared not to be fully acquired and were at the formal stage, as a result of which loose and sparse associative connections were created at the periphery of the network. However, considering that the cue words were mainly chosen from high frequency words, it was expected that stronger associative connections would also have been created in the center of the network (Henriksen, 2008; Zareva, 2016). The results clearly confirmed the predictions. Participants in both the experimental and control groups clearly supplied a proportionate number of closely associated responses: position-based, semantic-based, or semantic/position-based. The percentage of closely associated connections in the center of the network was comparable to that of the loosely associated ones at the periphery.

Approximately 60% of the responses could be assigned to closely associated groups, considered to comprise items that have developed strong connections as a result of a higher degree of lexical knowledge. As the word association test was timed and the participants were asked to provide as many responses as they could in the limited time provided, such responses could be taken as a sign of a level of lexical automaticity. Hence, such items cannot be assigned to the formal stage of development, a stage in which the activation of meaning is an effortful and conscious process due to it taking place only through translation to an L1 equivalent, and indicating a lack of direct linkage to the concept. Instead, closely associated responses can mostly be assigned to either an L1 lemma mediation stage, argued to present greater automaticity as a result of the direct link to the conceptual level of knowledge representation, or to the last stage of L2 lexical development, the L2 integration stage. In these cases, both position-based and semantically-based lexical items could be activated, as well as responses with dual links, which are “particularly strong and quick to retrieve” (Fitzpatrick et al., 2013, p. 45). However, a point worth mentioning here is that such stages and the types of association they activate are not clear-cut (Jiang, 2000), so the results must be interpreted with caution.

The discussion above has shed light on the structure of the specific strata of the lexicon prior to the intervention, revealing relatively close degrees of sparse associative networks, exhibiting weak clustering at the periphery of the lexicon and strong local clustering in the center, with a shorter path length among the items. However, inspecting the results immediately after the intervention reveals a significant difference both quantitatively and qualitatively in terms of semantic mapping.

6.3.2. Structural properties of the lexicon after intervention in control and experimental groups

The quantitative analysis shows a significant increase in the scores for the productive word association test for the semantic mapping group, which is an indication of the changes in overall network quality. The results also indicate that the participants managed to provide a higher mean number of closely associated responses, and there was a lower percentage of unassociated and loosely associated responses. Accordingly, a general overview of the results suggests a higher degree of connectivity among various items in the network, in tandem with changes in the response type and strength of the connections. This can be considered a sign of more structured and advanced bilingual lexicons, in line with previous findings (e.g., Henriksen, 2008; Wolter, 2001).

Returning to the earlier discussion of the theory of lexical development (Jiang, 2000, 2002, 2004), the findings presented can be related to the possible incremental nature of the process of vocabulary acquisition. The results highlight that the participants in the semantic mapping group gradually broadened their L2 word knowledge by attracting more syntactic and semantic features of the lexical entries. More precisely, on the basis of the L2 lexical developmental stages, in the case of increased experience and more frequent use of L2 words, the syntactic and semantic content of the L1 items can be copied into the L2 lexical items. Hence, lexical items could be used more fluently and automatically as a result of a direct link to the conceptual level of knowledge representation. This shift in the post intervention phase in the number and types of responses would seem to represent more items with a central position in the network and clusters of densely connected nodes around them, and also fewer items at the periphery as a result of the predominance of the closely associated items.

In brief, looking at the overall patterns of responses for closely associated words, it appears that the semantic-based connections among lexical items in the lexicon do appear to take precedence over other types of closely associated responses, and this phenomenon tends to create dense local connections in the center, while at the same time position-based responses, elicited as a result of moderately known words, are formed at an outer layer slightly further away from the core (Namei, 2004; Wolter, 2001). Although this distinction alone does not cause significant changes to the structure of the lexicon, it reveals differences in the degree of word knowledge and represents different strengths of connection among items as the dominant response type has changed. The results here are in line with those of Namei (2004), and argue in favor of primarily phonological organization of the L2 mental lexicon. Only under the condition of favorable learning situations, that is with sufficient exposure to the L2, is it possible to move toward a more semantic-based L2 lexicon. Hence, it can be argued that more closely associated responses can be the result of the state of L2 lexicon development with subsets of lexical items forming stronger semantic connections.

A second possible explanation, which more clearly justifies the differences between the two groups after the intervention phase, is that semantic mapping provided opportunities for the participants to get involved in active semantic encoding and processing of the lexical items, paying attention to the meaning of the words while discussing them in active manner. This implies that the semantic context provided for the lexical items by the application of semantic mapping aided the process of their integration in the network. The potential mechanism underlying this notion is the depth of processing hypothesis, which posits that the durability and elaborateness of the memory traces depend on the nature of the cognitive processes employed (Craik & Lockhart, 1972; Craik & Tulving, 1975). Precedents for this interpretation exist in the work of several researchers who have applied semantic mapping as a vocabulary teaching technique (e.g., Ghonsooly & Hoseinpour, 2009; Khoii & Sharififar, 2013; Mohammed & Malo, 2020; Sagarra & Alba, 2006; Zahedi & Abdi, 2012), and have suggested that the efficiency of semantic mapping is deeply rooted in the fact that the more semantic manipulations there are involved with a lexical item, the more elaborate routes will be available for subsequent retrieval (Craik & Lockhart, 1972; Craik & Tulving, 1975).

Nevertheless, there seems to be a substantial point of divergence between this study and other research in this field with regard to prior familiarity with the formal specifications of the lexical items. As the words used to feed the map had already been internalized based on their phonological/orthographical features, more attentional resources were available for the intensive processing of the meaning with semantic mapping, which in turn left more durable memory traces (Barcroft, 2015). This high degree of active semantic involvement on the part of the participants helped to increase the speed of retrieval, which in turn is a cornerstone in strengthening the memory traces already created in the network (Baddeley, 1990; Craik & Lockhart, 1972). Consequently, the process of generating a semantically organized structure in the form of a semantic map might have provided opportunities for the learners to develop local clusters with strong associative connections between items within each semantic field they had worked on.

The last piece of evidence in support of a more structured lexicon after semantic mapping comes from inspecting the level of frequency and canonicality of the closely associated responses supplied. The assumption was that learners would be able to produce a higher number of canonicals along with more low frequency responses as development progressed (Henriksen, 2008). First, scrutinizing the results for the canonical and non-canonical responses provided, it is clear that the mean number of both response types increased in the experimental group, and this increase was more evident in the case of non-canonical responses. As canonical responses play a central role in the organization of the lexicon and serve as bridges among various parts of the network, any increase in their number can be interpreted as an indication of those connections becoming well established in the network. Moreover, the rise in the number of non-canonical responses can be considered to be a move towards more advanced network. Second, the results demonstrate a clear effect of semantic mapping on the frequency of the responses; that is, the semantic mapping group managed to provide a higher number of low frequency responses, while a similar trend was not detected in the control group. These results tally with the expectations voiced in previous research, arguing that "the more advanced learner or the native speaker will be more likely to supply not only more canonical responses, but also a higher number of low frequent, non-canonical responses than the less advanced L2 language learner" (Henriksen, 2008, p. 35).

Finally, having discussed the results obtained immediately after the intervention, which provide clear evidence for a more structured lexicon in comparison to pre-training, we turn to the results from the delayed post-test. These showed that not all changes were preserved after a longer time span. Examining the pre-test and delayed post-test results, significant differences in favor of the delayed post-test could still be detected after the semantic mapping intervention; however, the effect of the instruction was moderated by the distance in time between the performance in the immediate post-test and the delayed test. This decrease in the overall productive word association scores led to non-significant differences between the control and experimental groups in the overall productive word association scores. However, the descriptive results on type of associations revealed that the mean score for the closely associated responses in the experimental group was still higher than in the control group, indicating a higher level of linkage in the core of the network. In terms of the non-significant difference between the overall scores of the groups, this can only be attributed to the high number of many items only sparsely connected with a few links at the periphery. The higher number of sparsely connected items at the periphery in the delayed phase of the testing could stem from the initial process through which the L2 lexical items were learnt in the first place. That is, in the Iranian teaching and learning context, it is quite common to learn vocabulary by focusing on the form of the words (morphological, phonological, or orthographic features), which can create strong and more stable form-based memory traces (Fitzpatrick & Izura, 2011). Arguably, the amount of exposure to the input was not sufficient to guarantee that the developments and changes would be maintained. Although such associations imply a reduction in the overall density of the network (Wilks & Meara, 2002), the high mean number of closely associated responses still signifies a dense network at the core.

The evidence highlights that although the impact of semantic mapping has never been scrutinized with regard to the underlying cognitive processes of the lexicon, particularly concerning the development of lexical knowledge from the lexicon-based perspective, and it has always been researched from the methodological perspective as a teaching tool, it appears that it can be used as a valuable means of making changes to the underlying network knowledge in the lexicon, which in turn can bring has implications for faster and easier access to the lexical items.

6.4. Semantic mapping and measures of utterance fluency

The second part of the research aimed to shed light on the impact of semantic mapping on various measures of L2 utterance fluency. Utterance fluency was investigated in terms of syllable run, phonation run, syllable duration, and mean duration of silent pauses. Table 6.2, Table 6.3, and Table 6.4 summarize the results of the Mann-Whitney *U* and Wilcoxon signed rank tests, and correlation analysis, respectively. Three measures—syllable run, phonation run, and syllable duration—improved the most in the experimental group after training sessions in semantic mapping. In addition, there were significant correlations with the word association scores, representing changes in the structure of the bilingual mental lexicon. In contrast, the only L2 fluency measure that neither changed significantly in the within or between group analysis nor correlated strongly with the word association scores was the mean duration of silent pauses.

In this section, findings of studies related to lexical processing skill and L2 utterance fluency measures are discussed along with suggested explanations with reference to the data of the current study. In the following, the findings on syllable run, phonation run, syllable duration, and mean duration of silent pauses are discussed in turn in relation to previous studies separately.

6.4.1. Semantic mapping and syllable run

This section presents the results of the analysis of the first holistic measure of fluent length of run, that is syllable run. The results for the semantic mapping group revealed a similar pattern in both the immediate post-test and delayed post-test, namely a significant increase in syllable run. In contrast, the significant gains attained in the immediate post-test by the control group were not sustained after the two-week interval in the delayed post-test. Analysis of the gains in syllable run also indicated a significant difference between the groups. Moreover, correlation analysis indicated a positive relationship between the productive word association test and syllable run measures in both the immediate post-test and the delayed post-test for the semantic mapping group. Thus, it seems likely that the significantly longer runs of syllables between pauses were related to underlying changes in the structure of the lexicon, represented through the scores in the free productive word association task.

As previously noted (see Section 3.7.2.2), one of the best gauges of L2 fluency is syllable run due to its strong positive correlation with speaking scores (e.g., De Jong et al., 2012; Kahng, 2014; Lennon, 1990; Segalowitz & Freed, 2004; Towell et al., 1996). This is perhaps unsurprising given that it is a measure reporting changes in the length of linguistic units uttered between pause bursts in terms of the number of syllables; hence, it is applied as a measure of density of lexical items in speech. Accordingly, of special interest here are previous studies that have examined the role of knowledge of lexical items in L2 speech fluency by assessing the measure of syllable run. Hilton (2008) found a high positive correlation between the breadth of lexical knowledge and syllable run (r = .67). Similarly, Segalowitz and Freed (2004) reported moderate correlations between syllable run and speed of lexical access (r = .38) and lexical efficiency (r = .38). Such findings were replicated in this study, finding a strong positive relationship between the results of the productive word association test and syllable run in the semantic mapping group, both immediately after training and following a two-week interval.

Therefore, the findings of this study lend further support to research that has consistently found language learners with deeper knowledge of lexical items in the lexicon have faster access to them, resulting in smoother production of the language. That is, as L2 speech production is mainly lexically driven (De Bot, 1992; Kormos, 2006), a well-organized and robust mental lexicon with strong connections between and within different levels of the network make L2 lexical items more readily accessible for retrieval and production (Koizumi & In'nami, 2013; Uchihara & Saito, 2019).

However, it should be noted that the relationship between vocabulary knowledge and syllable run has been scrutinized variously in different studies by conceptualizing lexical knowledge from different perspectives and dimensions of breadth, depth, or speed (e.g., Koizumi & In'nami, 2013). More specifically, the fact that syllable run was highly correlated with vocabulary knowledge in Hilton's (2008) study can be attributed to the relation between the breadth of the lexicon, that is the number of form-meaning links in the mental lexicon, and the measure of syllable run, reflecting the density of syllables between two silent pauses. Segalowitz and Freed (2004) explored the speed and efficacy of access to lexical items in the lexicon, contributing to understanding of the relative speed at which learners retrieve knowledge of the lexical items from the mental lexicon. This study contributes to understanding of the relation by highlighting that there is also a relationship between the depth of knowledge of lexical items from a network perspective and syllable run. The consistency between the results can be traced back to previous research showing a strong association between breadth and depth dimensions (e.g., Koizumi & In'nami, 2013). Therefore, in summary, the results of previous studies can be extended to the relationship between the conceptualization of the depth of vocabulary knowledge from a lexicon-based perspective, focusing on network knowledge and the structural properties of the mental lexicon, and syllable run as a fluency measure reflecting the density of the application of lexical items in speech between dysfluent silent pauses.

Besides the strong positive relationship between the organizational structure of the mental lexicon and syllable run, exploring the within group results, a significant increase in syllable run was found in the semantic mapping group following training in both the immediate and delayed post-tests. The results discussed in this sub-section together with the findings of the correlational analysis can be attributed to the efficiency of the application of the semantic mapping technique in helping participants develop lexical knowledge, which resulted in higher connectivity and stronger links among items. Such underlying changes in the structure of the mental lexicon resulted in more rapid activation of required lexical entries in the semantic fields as they were not placed in a random and loosely structured fashion in the word store, which could have contributed to the sustained gains in the delayed post-test. In summary, the significant increase in syllable run in the semantic mapping group after the intervention can be attributed to a decrease in lexical searching and encoding time as a result of having more items readily available throughout the spontaneous production of the utterance. Accordingly, the results of this experimental study not only reveal a strong association between organization of the mental lexicon and syllable run, in accordance with the literature, but also provide further evidence that enhancing vocabulary knowledge can actually lead to an increase in the level of density of lexical items produced between pauses in the process of speech production.

The fact that syllable run was found to increase after semantic mapping is in contrast with the findings of Ghonsooly and Houseinpour (2009) however. The main reason for this inconsistency could be due to the design of the instructional activities aimed at achieving improvements in the fluency of speech. In this study, the participants in the experimental group practiced a single conversation in each session and the new lexical items in the assigned conversation were presented with the application of concept maps. However, the process of mapping simply comprised feeding the map with lexical items related to the main topic, the meaning was provided by presenting synonyms or equivalent words, and there was no discussion or semantic encoding of meaning. Moreover, additional grammatical structures were presented explicitly at the same time as the semantic mapping and presentation of new lexical items. This in turn could have triggered a trade-off between accuracy and fluency due to the capacity limitations of the L2 language learners (Skehan, 2009). Hence, it might be the case that the participants prioritized focusing on delivering sentences with correct grammatical forms before moving to the next sentence, which in turn could have had a negative impact on the density of the lexical items they could have produced in one go.

Turning to the control group, the participants exhibited a significant increase in syllable run in the immediate post-test, but it was not sustained in the delayed post-test. Although the changes in this measure could be attributed to meeting the lexical items more frequently during the six sessions of intervention, it could also be related solely to faster production of speech. As syllable run was measured as the mean number of syllables between silent pauses, it can be considered a measure of rate of articulation as well as pauses (De Jong, 2018). Hence, a higher number of syllables in a stretch of speech could be attributed to faster production, particularly as the correlation analysis showed no significant relationship between word association scores and syllable run in the control group, constituting a temporary result of meeting and retrieving lexical items during the intervention sessions. Accordingly, as the practice was shallow and not semantically organized, only a short-term memory retrieval path was created, which disappeared not long after the training sessions in the delayed post-test.

In terms of explaining the increase in the syllable run measure, I suggest that organizing lexical items in semantic groups and presenting practice opportunities for learners to retrieve and discuss the items helped them produce utterances with higher lexical density, which could in turn be related to underlying changes in the structural properties of the lexicon in terms of connectivity and the strength of links among items. Consequently, the results of this study confirm the effectiveness of the semantic mapping technique in enhancing the elaboration of lexical networks, as represented in changes in syllable run and its strong correlation with word association scores.

6.4.2. Semantic mapping and phonation run

As explained in Chapter 3, phonation run is the mean duration of uninterrupted speech between silent pauses. In this study, the threshold was set at 250 ms, anticipating a strong relationship between the frequency of silent pauses and vocabulary knowledge (e.g., De Jong & Bosker, 2013; De Jong et al., 2013; Kahng, 2014, 2020).

This measure is equivalent to the frequency of silent pauses per minute. It is operationalized as the time spent speaking divided by the total number of silent pauses longer than 250 ms; therefore, prior to discussing the results, it should be noted that phonation run is the length of fluent phonation between pauses and not a measure of the density of lexical items. Therefore, there is a high chance that any increase in this measure might not be associated with the production of a higher number of lexical items, but rather longer voiced pauses (included in phonation time) or elongated syllables, resulting in a longer duration of phonation. To ascertain that this is not the case and interpret the results with greater certainty, the results of correlation analysis between syllable run and phonation run are consulted (see Appendix U). These clearly indicate that a higher number of syllables in a syllable run is associated with a longer run between silent pauses rather than voiced pauses or elongated words.
Overall, examining the results presented in 5.3.4, the greatest significant increase in phonation run was observed only in the experimental group between the pre-test and immediate post-test,

which in turn resulted in significant gains in favor of this group from the pre-test to the immediate post-test for between group comparison of the results; however, the results of the delayed post-test revealed that the gains were not sustained. Finally, the correlation analysis revealed a moderate positive relationship in the immediate post-test phase between the word association result and phonation run in the experimental group and also a strong positive relationship in the delayed post-test in the same group, but not in the control groups.

As mentioned in Chapter 2, no study has yet looked into the impact of the development of lexical knowledge and changes in the organizational structure of the mental lexicon on L2 fluency of speech; however, the previous literature has explored the relationship between various dimensions of lexical knowledge and fluency of speech, more particularly the measure of phonation run (inversion of frequency of silent pauses). Clenton et al. (2020) investigated the relationship between productive lexical knowledge, measured through a productive word association test (Lex30), and the frequency of silent pauses. They found that the higher the participants' scores in Lex30, the fewer the silent pauses detected in speech production. In addition, De Jong et al. (2013) explored L2 fluency in terms of the frequency of silent pauses and productive vocabulary knowledge, and reported a negative correlation between this measure and the Dutch version of the test of vocabulary knowledge (r = -.39), which comprised 77% lexical questions accounting for the relation between meaning and form, and 22% for depth of vocabulary knowledge. They argue that this relation could be due to smoother speech processing for learners with greater knowledge of lexical items, which in turn helps them in recall adequate items more rapidly and use them in speech production. The findings of this study lend further support to research that has consistently reported a moderate significant relationship between these two measures, while adding to understanding of the association between this fluency measure and organization of the lexicon. Not only is this association related to the breadth dimension of the lexicon, but there is also a relation between the organization of the mental lexicon and the number of silent pauses in speech. Thus, it is not only learners with a greater number of lexical items available in their repository who pause less, but also learners with more structured mental lexicons, reflected in the number and strength of connections in the network knowledge, manage to produce longer lengths of fluent runs and fewer pauses.

Besides finding a moderate correlation between the two measures in the semantic mapping group, this study explored the extent to which semantic mapping, as a technique for enhancing lexical development and generating changes to the structure of the lexicon, can actually lead to changes in phonation run. As mentioned earlier, the results indicate significant changes in phonation run in the semantic mapping group right after the process of intervention, namely a decrease in the number of silent pauses, resulting in longer fluent runs. In addition, increases in the phonation run in the semantic mapping group created significant difference between the two groups regarding gains from the pre-test to the immediate post-test.

The increase in phonation run attested to here represents a reduction in the need to pause, but why is this so? It seems likely that the increase in phonation run can be explained by a reduced need to search for lexical items as they were more readily available (De Bot, 1992; De Jong et al., 2013; Hilton, 2008; Kahng, 2014). Therefore, the high number of silent pauses in the pre-test in comparison to the two other phases of testing can be considered a sign of lexical retrieval problems, which could originate from rather loosely structured network knowledge. According to Hilton (2008), 78% of pauses in speech are followed by either an overt or a possible lexical search, that is, "an avowed incapacity to retrieve the necessary lexical items" (p. 159). Thus, an increase in phonation run in the semantic mapping group can be explained by the development of lexical knowledge, and creation of denser and stronger clusters of closely related networks, i.e., a more organized lexicon upon which the learner could draw more quickly in spontaneous speech. This can further be confirmed with reference to the results of the correlational analysis discussed above, which present a strong positive relation between the phonation run measure and the word association test in the immediate post-test for the experimental group. It can be argued that a more organized lexicon, as reflected in the connectivity and strength of associations in the word association task, is positively related to less frequent pausing in the speech sample.

An earlier study by Ghonsooly and Hoseinpour (2009) reported similar results in relation to the application of semantic mapping in teaching new lexical items and its impact on the frequency of silent pauses in speech. They found significant changes in the number of silent pauses between the group receiving practice in semantic mapping and the control group in the post-test; however, taking their results for the number of silent pauses and syllable run together, it is likely that the decrease in the number of silent pauses could be attributed to the elongated syllables or longer duration of filled pauses, rather than longer utterances with more lexical items, since the results for syllable run showed no changes in the number of syllables uttered between silent pauses.

As noted above, there were clear changes in the immediate post-test in terms of a longer phonation run, although such changes were not sustained in the delayed post-test. Looking at the descriptive statistics (see table 5.4), it is apparent that the mean length of utterances in the delayed post-test ($M = 2.16$) was still higher than in the pre-test ($M = 2.04$), and significantly correlated with the word association test results, although not enough to reach significance. To gain a better grasp of the non-significant changes in the delayed post-test, it is perhaps worth interpreting the results in light of the syllable run measure, as they are both measures of fluent, uninterrupted runs.

Re-examining the syllable run results, there was a significant difference between the pre-test and delayed post-test, although the same result was not found for phonation run. As both these measures have the number of pauses as a constant measure in their calculation, the difference can be attributed to the number of syllables, which differed significantly between the pre-test and the delayed post-test, and the duration of uninterrupted speech in the phonation run measure, which was still longer, although not significantly different from the pre-test. Therefore, it seems that the participants were able to produce a higher number of syllables, but the duration of utterance was not as long as in the immediate post-test. Accordingly, the shifts can be attributed to the changes in the speed of articulation (confirmed by examining syllable duration in 6.4.3). It might be the case that as the participants had the lexical items required ready in their depository and had created a network of dense links, they managed to produce faster speech with more syllables in a shorter period of time between silent pauses. This can also be confirmed by the strong correlation between word association and phonation run in the delayed post-test.

Consequently, taking all the results together based on the current discussion, as the number of silent pauses is a fairly good indicator of the underlying processing skill of lexical retrieval and changes in this measure can be interpreted as changes in the speed of retrieval of lexical items (De Jong & Bosker, 2013; De Jong et al., 2015, 2013; Hilton, 2008; Kahng, 2020; Segalowitz & Freed, 2004; Tavakoli, 2011), it can be argued that changes in the structure of the lexicon, resulting from developmental shifts in knowledge of the lexical items through semantic mapping, are the key factor underlying these changes.

6.4.3. Semantic mapping and syllable duration

Syllable duration is a measure of the duration of the syllables used in speech, reflecting the size of the syllables making up the speech flow. Prior to discussing the results of the application of the semantic mapping technique, it is worth mentioning that syllable duration has been operationalized as a pure measure of speed in this study, since the impact of pausing was minimized by calculating the phonation time minus both filled and silent pauses. Hence, the results are a mere representation of the true phonation time as the use of both filled and silent pauses differs according to the individual, and sometimes reflecting speech processing difficulties (Cenoz, 2000). Therefore, as syllable duration is not confounded with pauses in the calculation, any increase in this measure can be considered either a reflection of the lengthiness of sounds in speech, that is, the use of elongated syllables, or the speakers' inclination to use micropauses instead of macro pauses, which in turn can be a matter of L1 behavioral differences or individual personality traits (Goldman-Eisler, 1961, p. 173). Both prolonged sounds and micropauses help to restore continuity (Clark & Wasow, 1998), while buying time to activate and retrieve the lexical items required. Therefore, higher measures of syllable duration can be a representation of a greater need to hesitate, while lower measures suggest the speaker moving at speed from one word to the next.

The results revealed a significant decrease in the measure of syllable duration in the semantic mapping group from the pre-test to the immediate post-test, and also the delayed post-test; however, this was not the case for the control group. Furthermore, exploring the between group results indicated significant differences in gains from the pre-test to the immediate post-test, and from the pre-test to the delayed post-test between the experimental and control groups. Finally, the results of correlation analysis revealed a weak, non-significant negative correlation between the word association scores and syllable duration in both the immediate and delayed post-tests.

Drawing on the literature, several studies in the field of second language acquisition have reported partially close results concerning the relationship between lexical knowledge and syllable run (e.g., De Jong et. al., 2012, 2013; Kahng, 2020; Uchihara et al., 2020). For example, Kahng (2020) reported a similar weak correlation between the breadth of lexical knowledge and syllable duration ($r = -.13$), and a significant moderate relation between the same fluency measure and speed of lexical retrieval ($r = .50$). Uchihara et al. (2020) reported a strong moderate correlation between articulation rate (inverse of syllable duration) and breadth of lexical knowledge measured by Lex30 ($r = .50^{*}$). Clenton et al. (2020), using the same word association test, reported a positive weak correlation with articulation rate ($r = .12$). Broadly speaking, the results of this study align with the findings of some previous research indicating a relationship between the two measures, revealing a weak correlation between them.

The inconsistencies in the results could be due to two major factors. First, one possible reason might be the dimension of lexical knowledge tapped with the use of word elicitation tasks in such studies. Various studies have drawn on different conceptualizations of lexical knowledge. For example, Uchihara et al. (2020) investigated productive knowledge of vocabulary through the application of Lex30, which is a word association test of vocabulary assessing the breadth of the lexicon (r = .48 to .31). De Jong et al. (2013) also made use of a Dutch version of a sentence completion task assessing breadth of lexical knowledge along with a single aspect of depth of knowledge (collocations) (r = .58). This study benefited from word association tasks specifically designed to tap the structural properties of the mental lexicon, that is, the depth of vocabulary knowledge from a lexicon-based perspective (r = -.38).

Second, another explanation might be the different scoring methods used for vocabulary testing. A clear example can be found in the different studies conducted by Clenton et al. (2020) and Uchihara et al. (2020), both of which made use of Lex30 as a productive word association test. Although the lexical test was identical, the scoring procedures were different; hence, different results were gained for the measure of syllable duration: Clenton et al. (2020) reported r = .12 for syllable duration, while Uchihara et al. (2020) reported r = .58 for articulation rate.

The results of this study in general are in tandem with earlier research (e.g., De Jong et al., 2012, 2013; Kahng, 2020; Uchihara et al., 2020), providing further support for the proposition that speech production is lexically driven (Kormos, 2006a; Levelt, 1989). It is argued that not only will learners with smaller mental lexicons have difficulty producing smooth and fluent speech, but also that the depth of knowledge of lexical items contributing to the organization of the lexicon will add to the relative ease of lexical retrieval.

As discussed above, previous studies have only explored the relations between the measures of syllable run and vocabulary knowledge, so it is not possible to state the extent to which a pedagogical technique, such as semantic mapping, impacts syllable duration as a result of developing lexical knowledge and structuring the organization of the lexicon (i.e., if syllable run would change as a result or remain constant). The only other study to do so, to my knowledge, was conducted by Ghonsooly and Hosienpour (2009), which likewise reported fluency gains with respect to articulation rate, which is comparable to syllable duration in this study.

Exploring the results (see Section 5.3.5), it can be seen that the mean syllable duration decreased significantly after training with semantic mapping in the immediate and delayed post-tests. The decrease in syllable duration in the experimental group concerns "the speed at which a speaker moves from one sound to the next" (Hunter, 2017, p. 241), which also suggests a reduction in the need to micropause or elongate syllables. This measure can be considered a gauge of the extent to which a participant "buys time" when speaking using elongated syllables (Hilton, 2008).

Table 6.5 presents an example showing the difference between the performance of a randomly selected participant from the semantic mapping group, at two test points, before and after the intervention. This sample, extracted from the last 20 s of the speech, shows that the production of the participant's speech in the immediate post-test was faster than in the pre-test. Aside from the considerable difference related to pausing, the first sample includes four elongated syllables, while the second one contains only one. Lengthening the sound can be interpreted as an indication of lexical search, the result of which can be retrieval of a lexical item and using it in speech; however, this is not always the case and there is sometimes a need to fall back on rephrasing the utterance to avoid lexical retrieval problems, leading to an increase in syllable duration.

In addition, examining the second sample in the immediate post-test, the presence of lexical items such as *coach*, *clap*, *ahead*, and *cup* shows that the reduced need for lexical searching originates from an increased ability to activate and retrieve lexical items in a network of partially or fully integrated items with stronger connections in comparison to the pre-training phase.

Table 6.5 Example of elongated syllables

Pre-test:	**Immediate post-test:**
(4.45ec) after that˜ (.73mu) he was looking on the (.27mu) poor frog that is alo˜ne a˜nd (.83mc) so˜ upset (1.47ec) they invited from to come and play with them	His coach is clapping for him (.27eu). when he woke up, he saw that the others are ahead (.45eu), and he ran but he˜ (.26mu) couldn't (.98mc) he couldn't get to others (.40eu) and (.52mv) his friends (1.90mc) got the cup

Note: ~ = elongated syllable

The final alternative explanation, which seems somewhat far-fetched, although still possible, is the possibility of a reduction in syllable duration as a result of a reduced need to plan the content of the message. Although the picture stories were not at all about the same content, the lexical items that could be used to narrate them were from a particular stratum of the lexicon. It is plausible that familiarity with the general content helped participants increase their speed of delivery as there was less need to plan and conceptualize the talk (i.e., Levelt's *conceptualization*). To ascertain if this was the case, further investigation of elongated sounds is needed as no pause type was included in calculating syllable duration, and elongated sounds function as filled pauses to fill the void and create an illusion of continuity of speech. Accordingly, differentiating between elongated words occurring at clause boundaries, reflecting problems in conceptualization, or within clauses, representing dysfluencies resulting from the process of formulation, could help to decide if the differences were associated with changes in formulator or not (Skehan et al., 2016).

Consequently, the results of this study are predominantly in alignment with the literature concerning the close relation between the measure of syllable duration and productive lexical knowledge (e.g., Clenton et al., 2020; De Jong et al., 2012, 2013; Hilton, 2008; Kahng, 2020; Uchihara et al., 2020), and more specifically the depth of vocabulary knowledge from a lexicon based perspective attributing to the organization of the mental lexicon through assessing the extent of connectivity and strength of associations among lexical items.

6.4.4. Semantic mapping and silent pause duration

Silent pause duration is a measure of dysfluency in speech, operationalized as the duration of silent pauses produced in an utterance over the number in total. In this study, the results revealed that silent pause duration was the only measure of L2 utterance fluency, among the four analyzed, that was not significantly affected after the intervention. Analysis of the mean duration of silent pauses indicates a trend toward a significance difference within the semantic mapping group from the pre-test to the immediate post-test, and also gains between the groups from the pre-test to the immediate post-test phase; however, the results were not significant in any phase of testing or analysis. Furthermore, the results of correlation analysis between this measure and word association scores also revealed a weak negative relation in the immediate post-test for the semantic mapping group, which was not sustained in the delayed post-test.

The results of this study in general mirror those of most research in this domain, stating that the duration of silent pauses scarcely reflects L2 cognitive fluency with respect to lexical knowledge or the processing skills of lexical retrieval (e.g., Clenton, et al., 2020; De Jong & Bosker, 2013; De Jong et al., 2015, 2013; Kahng, 2020). Clenton et al. (2020) found that the mean duration of pauses, both between and within AS units, are minimally related to the scores for productive vocabulary knowledge assessed through the Lex30 word association test (r = -.06 and r = -.15, respectively). This is in accordance with the studies of De Jong et al. (2013) and Kahng (2020), which both investigated the relation between this fluency measure and productive knowledge of lexical items. Although the aforementioned research taps into various dimensions of the mental lexicon, overall, the results are consistent with the findings of this study.

The findings in this section are interesting in that not only do they endorse the literature in this domain, but they also reveal that modifications to the network knowledge resulting in changes in the connectivity and strength of connections in the lexicon, as reflected in gains in almost all other three L2 fluency measures examined, did not yield significant changes in the outcomes for pause length. Although there is a strong consensus in the literature that the more words learners know and the larger their lexicon, the more fluent they can be in their speech production (Hilton, 2008), this was not strongly confirmed by the mean duration of pauses as a measure of fluency. The assumption in this study was that were the lexicon organized for optimal access during online encoding, aiding the process of retrieval as a result of the activation of many more items with numerous links in a specific stratum of the network in question, it would help by yielding shorter pauses in the process of oral production; however, the results did not meet the expectations.

A possible explanation for this concerns the nature of silent pauses, reflecting language-general fluency, that is, the personal speaking style of the speakers as opposed to L2-specific issues (de Jong & Mora, 2019; de Jong et al., 2013, 2015; Kahng, 2020; Riazantseva, 2001). Drawing on the literature, it has been argued that some people pause longer when talking than others as an idiosyncratic attribute; however, as the purpose of this study was not to investigate the impact of L1 speaking style on L2 speech performance, no precise conclusion can be drawn.

In addition, looking at the overall speech production model, it might be possible to explain the minimal changes in this measure by considering the possibility that the participants used long pauses to plan and conceptualize prior to speaking. Scrutinizing the speech samples elicited from participants, it can be observed that as they were provided with planning time prior to delivering their speech, they were prepared with ideas and content for delivery as soon as they started talking, which in turn could have led to a shorter duration of pauses and a higher rate of delivery. However, over time, the participants ran out of ideas and had to look for additional content and ideas to fill the remaining speech, leading to a longer pause duration for conceptualization. Nonetheless, the possibility that this time was used to formulate the utterances cannot be disregarded.

On the whole, then, the findings of this study provide support for much of the literature, particularly the work of Clenton et al. (2020), who applied the word association test to assess knowledge of vocabulary, and reported a very small non-significant association between productive knowledge of lexis and silent pause length, confirming that this L2 measure is not significantly related to knowledge of lexical items, either in terms of size or network knowledge, or the structure of the lexicon.

6.5. Summary

In this chapter, the findings of the present study were discussed on how each part can answer the research questions posed. In the first place, the impact of semantic mapping technique on word association test, as a representation of the network knowledge and organization of the mental lexicon, concentrating on the connectivity and strength of the links among items stored on lexicon, was explored. The results of the study along with the related literature suggest that semantic mapping can create changes to the level of lexical knowledge as a result of semantic encoding of the lexical items, contributing to shifts in the number and strength of connections between items in various levels of representation in the lexicon, which in turn can have implications for fluency of spontaneous production of speech discussed in the following section.

In the second section, the findings related to various measures of utterance fluency in relation to semantic mapping were discussed in detail. As stated previously, participants were more fluent after semantic mapping compared to the pre-training phase for most L2 utterance fluency measures, including syllable run, phonation run, and syllable duration. However, it is noteworthy to mention that mean duration of silent pauses didn't significantly change. In addition, correlation analysis showed that for word association and L2 utterance fluency measures, syllable run and phonation run exhibited the strongest correlation, whereas syllable duration and duration of pauses showed moderate and weak, respectively. Overall, the findings suggest that semantic mapping helped the participants in the experimental group to activate and retrieve higher number of lexical items as a result of a more organized and densely structured network knowledge. The results clearly suggest that participants managed to produce higher number of syllables in longer duration of phonation, that is less frequent silent pauses, with higher rate of articulation. This all was suggested to be attributed to creation of higher number of links among items in various levels of presentation in network of knowledge, which is in turn originated from increase in degree of word knowledge. The results implied that in case of lexical items acquired mainly through translation equivalents in Iranian context of teaching and learning, the L1-L2 link in the lexical level is most often unidirectional (receptive) while by semantic mapping and the process of productive lexical retrieval, it was possible to strengthen the bidirectional links in the lexical level. Finally, retrieval of lexical items in such a dense local network can leave memory traces, which can in turn be strengthened when processing is mainly semantic as the case in semantic mapping. This finally leads to faster production of lexical items in speech as indicated by various measure of utterance fluency.

7. Chapter 7: Conclusion

7.1. Overview

In this chapter, an outline of the key elements and main implications of the current research study is presented in eight sections. Initially, the substantial elements of the current study including the aims along with the research questions (see Section 7.2), summary of the study (see Section 7.3), and summary of the key findings (see Section 7.4) are addressed. Next section deals with the implications of the study from theoretical (see Section 7.5.1), methodological (see Section 7.5.2), and finally pedagogical perspectives (see Section 7.5.3). Next is a section on discussion of the limitations of the study (see Section 7.6). The chapter will be finalized with a discussion of possible directions for future research (see Section 7.7) and final considerations (see Section 7.8).

7.2. Aims of the study

There have been many studies in the realm of fluency exploring various aspects of lexical knowledge and the related subprocesses involved in the production of fluent L2 speech, as lexical knowledge has been introduced as "the greatest impediment in spoken L2 fluency, at least from a temporal point of view, since lack of lexical knowledge, or of access to this knowledge would appear to be the primary cause of most dysfluencies" (Hilton, 2008, p. 163).

To address this main challenge, this thesis set out to explore the impact of semantic mapping technique on both organizational features of mental lexicon and L2 fluency of speech among intermediate Iranian learners of English. The study had the following main aims; first to explore the extent to which semantic mapping technique is beneficial in structuring the mental lexicon. Second, to probe the extent to which this same technique can lead to improvements in measures of L2 utterance fluency, and finally, to explore if there exists any relationship between the underlying changes in the organization of the mental lexicon and L2 fluency of speech. To achieve the main overarching aims, the following contributing research questions were posed:

RQ1: To what extent does semantic mapping affect organization of bilingual mental lexicon as measured through free productive word association?

RQ2: To what extent does semantic mapping affect L2 speech fluency measured through empirically established means of gauging L2 fluency?

RQ3: Is there a relationship between organization of bilingual mental lexicon and L2 fluency of speech?

7.3. Summary of the study

As stated above, the main aim of the current study was to investigate the impact of semantic mapping technique on the structure of the mental lexicon along with the extent of its impact on L2 fluency of speech. Accordingly, it was also decided to further investigate whether there was any relationship between changes in the degree of structuredness of the lexicon and gains in speech fluency. A purely quantitative research strategy was adopted to find answers to the research questions proposed.

The participants were thirty-three adult EFL learners from Iran, studying General English Language courses at a private language school. On the first place, the volunteered participants took part in a baseline English language proficiency test to ascertain that they were in the same level of language proficiency. Next, they were randomly assigned into groups of experimental and control in four different classes, two of which were participants in the control group and two other experimental group. The main reason for utilising a non-active control group was to control for any potential test effect influencing the learners' performance.

Next, in order to find out if training participants with semantic mapping on the already known vocabulary in specific semantic fields have any impact on the lexicon and speech fluency, multiple pre-tests, immediate post-tests, and delayed post-tests were administered in weeks 1, week 5, and week 8, respectively. The study lasted for eight weeks in total. The intervention sessions were administered in six sessions over a period of three weeks, giving a total duration of four hours and thirty minutes. The participants in the experimental group received training on the application of semantic mapping technique as a way to practice the already known lexical items, while the study made use of a control group at the same time to isolate the impact of independent variable.

The key dependent variables in the current study were organization of mental lexicon and L2 utterance fluency. Any changes in the organization of the mental lexicon was measured through overall productive word association test, which was adapted from Henriksen (2008). Along with the calculation of this measure, type of responses provided were also scrutinized in order to keep track of changes after application of semantic mapping technique. In addition, L2 utterance fluency was assessed by measuring four temporal measures of syllable run, phonation run, syllable duration, and duration of silent pauses, which believed to be among the best indicators of utterance fluency reflecting the underlying cognitive processes.

7.4. Summary of the findings

The current study has provided substantial evidence that semantic mapping technique can have a positive impact on the degree of structuredness of the mental lexicon, as a result of which improvements in certain aspects of L2 utterance fluency can be gained. The result extends the findings of previous studies in that not only this technique can be applied as a beneficial vocabulary teaching technique supporting the breadth of vocabulary development but also based on the results of the current study, it supports the development of depth of lexical knowledge of the already acquired lexical items from a lexicon-based perspective. That is, further evidence in support of changes to the structure and organization of mental lexicon in terms of the number and strength of connections have been offered, which have been complemented by fewer number of sparse connections and higher number of local clusters of networks, exhibiting stronger connection at the center of the network. This in turn demonstrates that lexical items will be more readily available to be produced in spontaneous speech utterances.

The following is the summary of the findings in relation to the proposed research question;

RQ1: To what extent does semantic mapping affect organization of bilingual mental lexicon as measured through word association?

Findings of productive word association test revealed a trend in favor of semantic mapping. More specifically, the results indicate that semantic mapping technique has a significant positive effect on changing the structure and organization of the mental lexicon, more specifically in terms of the network knowledge of that particular domain, both at the immediate post-test and even after the interval of two weeks in the delayed post-test phase. In contrary, no significant impact was reported on the performance of the control group.

RQ2: To what extent does semantic mapping affect L2 speech fluency measured through empirically established means of gauging L2 fluency?

To answer this question, four different measures of L2 utterance fluency, syllable run, phonation run, syllable duration, and duration of silent pauses, were assessed. The results of within group analysis revealed a positive significant effect of semantic mapping technique on measures of syllable run, syllable duration, and phonation run from pre-test to immediate post-test, while no significant impact was reported for the measure of mean duration of silent pauses. Subsequently, the results of pre-test to the delayed post-test indicate that two measures of syllable run and syllable duration were still significantly different; however, changes in phonation run was not sustained. Finally, no significant difference was again reported on duration of silent pauses.

In addition, with respect to the comparison between semantic mapping and control group, the results confirmed that semantic mapping had a significant positive effect on gains of pre-test to immediate post-test in three measures of measures of syllable run, phonation run, and syllable duration, while the gains were sustained only for two measures of syllable run and syllable duration to delayed post-test.

RQ3: Is there a relationship between organization of bilingual mental lexicon and L2 speech fluency?

The findings of the correlational analysis right after intervention between the scores on the productive word association score and four utterance fluency measures suggested a strong relation between two fluency measures of syllable run and phonation run and overall score on word association test. However, the correlation was weaker with syllable duration and mean duration of silent pauses. The correlation on the delayed post-test phase indicated very similar results with the exception of the silent pause duration which is positively correlated instead. Considering the control group, no strong correlation between the measures were reported.

The results generated various number of implications and suggestions, which will be discussed in the following sections of this chapter.

7.5. Theoretical, methodological, and pedagogical implications

In this section, an outline of the implications of the study is presented in three main sections of theoretical, methodological, and finally pedagogical. With regard to the theoretical implications, I focus on how the present study provides insight into the role of organization of mental lexicon and network knowledge on L2 fluency of speech and suggesting a more in depth look into the influence of linguistic knowledge and processing on oral fluency. Next, the methodological implications outline contributions in terms of the application of quantitative method in researching organization of the mental lexicon, speech analysis and use of PRAAT, and design of a productive word association task for the purpose of the study. Finally, the pedagogical implications include a discussion of the applicability of semantic mapping technique in relation to lexical knowledge, organization of mental lexicon, and fluency of speech in L2 context, as well as how to sell the idea to students as a way to practice fluency. Finally, a discussion of L2 fluency assessment in L2 classroom context is presented.

7.5.1. Theoretical implications

Semantic mapping has always been regarded as a technique to teach new lexical items, targeting the breadth of lexical knowledge. The literature on semantic mapping has primarily drawn the efficacy and usefulness of this technique in establishing high number of form-meaning connections, grounding the discussion on depth of processing hypothesis explaining that new lexical items become related to the ones already acquired and previous experience, leading to creation of durable memory traces due to semantic encoding of the lexical items (Craik & Lockhart, 1972). The results of this study, however, promote the effectiveness of this technique in developing deeper knowledge of lexical items, which in return can increase the likelihood of L2 oral fluency development. Therefore, the results lend support to models of lexical development in traditional context of L2 classroom as well as theories of speech production.

Following Meara (1996), proposing that lexical items are interlocking networks which should be explored in relation to one another rather than in solitary, this study offers substantial evidence that semantic mapping is effective in the way of developing depth of lexical knowledge from a lexicon-based perspective. That is, it offers chances of creating changes in the organization and degree of structuredness of the bilingual mental lexicon. The support for this line of argumentation comes from the changes in the number of connectivity and strength of associations created in the lexicon as the outcome of the application of sematic mapping technique, which is envisaged in creation of denser local connections in the center of network and fewer sparse connections at the periphery, as a result of less moderately known items.

Subsequently, I would argue that this goes some way to explaining the findings in literature that semantic mapping seems to result in development of fluency of speech. I would argue that as speech processing is lexically driven, a dense network of connections to draw on during spontaneous production of speech, meaning higher degree of structuredness of mental lexicon, can be a contributing factor in development of fluency.

There is large literature explaining that one of the main impediments in fluent production of speech is problems with lexical retrieval (e.g., De Bot, 1992; Hilton, 2008, Segalowitz, 2010). This study has suggested that the gains in L2 fluency of speech are linked to the structure of one of the most significant knowledge components in language processing; that is, the bilingual mental lexicon. Segalowitz (2010) proposed various fluency vulnerability points in the process of speech production, all of which are the critical aspects underlying the cognitive processes. However, I would suggest that the structure of mental lexicon, although not sharing any processing function (Kormos, 2006), can have a major impact on fluency of speech.

Finally, the current study has also attempted to shed some light on the relationship between oral fluency and depth of lexical knowledge, from a network lexicon-based perspective. The literature has been mainly drawn on the association between L2 fluency and breadth dimension of lexical knowledge, and even cases reporting the relation with depth of knowledge, the concept has been scrutinized from a word-based perspective. This study has suggested that organization of mental lexicon is closely associated to L2 oral fluency of speech. That is, although it has often been suggested that learners with bigger lexicons (higher number of known lexical items) are more fluent speaker of L2, I have suggested that a more structured lexicon with closely tied links in the center of the network, have a significant association to L2 fluency of speech.

7.5.2. Methodological implications

The present study has a number of findings which may be of methodological value to fluency and vocabulary research. These will be addressed in four sections: (1) The use of quantitative method in researching organization of mental lexicon; (2) Speech analysis and use of PRAAT; and (3) Designing a productive word association task for the purpose of the study; and (4) The use of counterbalancing method in research.

7.5.2.1. The use of quantitative method in researching organization of mental lexicon

The current study has shown that a quantitative approach provides an ideal platform from which to investigate the organization of the mental lexicon, since it was possible to present a general but detailed reflection of the network knowledge.

As it has already been discussed (see Section 4.4), most studies in this realm either take the weighted/non-weighted scoring procedures comparing participants responses to the list of the norms (e.g., Wolter, 2002) or lexico-syntactic pattern of associative responses (e.g. Fitzpatrick, 2006; 2007, Fitzpatrick & Izura, 2011) into account in the way of drawing conclusions on the organizational properties of the lexicon; however, the current study, adapting a coding and scoring scheme from Henriksen (2008), managed to incorporate various aspects of connectivity, strength and type of the connections along with the level of frequency and canonicality of the responses in a single score.

The new categorization system was proposed based on the findings of previous studies, primarily drawn on Fitzpatrick's work (Fitzpatrick 2006; 2007; Fitzpatrick & Izura, 2011). The previous coding system (Henriksen, 2008) has the main challenge of having responses which either could not be assigned to any of the groups or various responses of different natures being assigned into a single group (semantically related to the cue word). For example, two-step associated responses were required to be assigned either to 'formal link', 'semantic link to the first response' or 'no association' groups as there was no category defined for responses associated with the cue word through a formal link (phonological/orthographic) to the previous response provided. Therefore, I would argue for the application of the newly designed coding scheme in future research, which is developed in a more systematic way in line with the previous literature, taking various types of associations into account, which can in turn reveal important information about the structural properties of the lexicon.

Secondly, as the main aim of the current study was exploration of the organization of the mental lexicon in particular strata, the number of connections along with the strength of the connections, both of which leading to determining the centrality of the items in the network, were significant and taken into account in designing the scoring rubric. The novel dimension of the new scoring rubric, which is grounded on addressing the main aim of the current study, was assessing the semantic density of the network. This was made possible by assigning a score to the *sematic strength* of the connections, a score which was not included in the original scoring scheme by Henriksen (2008). Hence, I would therefore argue for the benefits of application of this scoring scheme in future research on the grounds that it allows researchers to have a more detailed and subjective perspective of the organization of the mental lexicon, which is a point missing in previous researcher. In addition, the overall word association score provides an approximate estimate of network knowledge for a certain stratum of the lexicon, taking the semantic strength of the connections into account, which represents the stronger connections among the items in the lexicon.

7.5.2.2. Speech analysis and use of PRAAT

As it has been explained in Chapter 3 (see Section 3.7.1.2), the temporal measures of utterance fluency were analyzed using the computer program PRAAT (Boersma & Weenink, 2018). Along with application of PRAAT computer software, two different scripts were made use of; firstly, De Jong and Wempe (2009) to automatically detect sounding and silent segments in speech data, and secondly, Handley's (2017) script to extract particular fluency measures for the purpose of the current study.

In order to use PRAAT, training on how to make use of this computer software, to annotate and code the data effectively was necessary as manual annotation of the data was essential after the initial automatic coding. Although automatic coding of the data is accurate and objective, its application is more advantageous in the case of speech data recorded in silent soundproof rooms rather than natural context, which is not possible in the case of the present study. In addition, it was necessary to learn the basics of how to develop a PRAAT coding script for the purpose of the analysis of the data. Although the programming script was adapted from Handley (2017), it required further changes to fit the measures of the current study. The script was a comprehensive one, calculating almost all measures of fluency; however, it was not designed to elicit two of the measures, namely of phonation run and syllable duration; hence, further changes to the scrip was required.

The time and energy were all well spent as annotating and coding the data for the second time after the automatic coding proved to worth it. Accordingly, I would finally argue in support of manual annotation of the data due to three main reasons. First and foremost, it provides chances for the researcher or the practitioner to gain a comprehensive overview of the data and get to know if there are any points needing further attention and consideration. In the second place, I would argue that manual coding of the speech can lead to more reliable results in case the voices are not recorded in a soundproof quite room. In case of this study, although the voices were recorded in an empty and quite class with no other students around, the noise from other classes or the students in corridors would lead to the detection of syllables or sounding segments which were the result of noises rather than words produced. Finally, manual encoding of the data would make it possible to code the speech on the basis of the particular fluency measures believed to reflect the underlying cognitive processes.

7.5.2.3. Designing a productive word association task

Word association task appears to be "deceptively simple" (Fitzpatrick & Thwaites, 2020); however, as the thread running through the current thesis and more specifically chapter 4 proves, its application as a way of investigating mental lexicon is not straightforward and should be carried out with extreme caution. For the purpose of this study, it was decided not to use any of the pre-prepared word lists and associated responses available in literature but to design a productive free word association test (see Chapter 4) considering different factors, including careful selection of the cue words, developing a normative list from the target population, defining the connection between the cue word and the responses produced, deciding on the way of categorizing the responses, and finally adapting a coding scheme and scoring system, which made it possible to look into the data from a number of various angels.

Two main methodological challenges faced in the process of designing and implication of the test which needed to be overcome were as follows. First and foremost, the primary challenge in the way of designing the test was selection of the cue words to be used in the study. Although in the process of choosing the correct list of cues, Fitzpatrick's (2006) framework had been closely followed, it was still proved to be a real challenge requiring numbers of piloting the lists to know what group of words to choose. The point of novelty of the designed test, overcoming the challenge of what words to include, was keeping the cues specific and confined in a way of tapping into specific local stratum of lexicon, meaning that particular layer of network was only tapped on in order to make more specific judgments. Other word association tests, being widely used in literature, make use of range of cues which are located in different strata of lexicon and the conclusions are drawn about the global network of lexicon which might not lead into very accurate results. For example, Lex30 is a productive word association test, in which the respondents are presented with a list of 30 cue words from the list of highly frequent words. Uchihara et al. (2019) made claims on the basis of the results collected via this test on the structure of the lexicon as a whole; however, making such a claim needs further empirical investigations.

Secondly, although the process of administration of the word association test is straightforward and easy, a challenge faced was avoiding students from referring back to the previous cue words and adding more responses. As the aim of the test was collecting data on the most accessible links, the time for providing responses for each cue word was limited and referring back to the responses of the previous cues in the process of testing would affect the accuracy of the results. In order to soften the blow, a booklet was designed with ten pages, each page allocated to responses of one single cue. Hence, participants had to flip the page to answer the following cue, which hinder them from adding more responses to the previous words.

I would therefore argue that the methodology applied in the present study, which is partially in compliance with Fitzpatrick (2006, 2007), along with the analysis, being pursuant to Henriksen (2008), proves a beneficial tool in a move towards a more consistent solution which has the potential of resulting into less discrepancies in word association behavior probing into mental lexicon; however, there is still plenty of work left to be done. Consequently, I would argue for the benefit of designing the word association task and selection of analytical approach to take into account the particular requirements of specific research questions. I feel like the results make the additional hard work involved rewarding and worthwhile.

7.5.2.4. The use of counterbalancing method in research

Next, the current study benefited from counterbalancing technique for the evaluation instruments of both fluency and word association tests. The instruments, both picture narrative and word association tests, were counterbalanced by using three parallel versions of each test. As the data was collected in three different phases of pre-test, immediate post-test, and delayed post-test, three versions of the same tasks were developed, and each version was administered in one phase of testing. Counterbalancing the assessment tasks was advantageous due to two different reasons; firstly, to maintain the internal validity by controlling any effect created as a result of order and sequence of presentation of the tasks and secondly to control the test effect and the chance of learning as a result of the test itself, such as participants remembering the task.

7.5.3. Pedagogical implications

Based on the findings of the current study, there are various significant pedagogical benefits attributed to the effective integration of semantic mapping in improvement of lexical knowledge and fluency of speech in EFL classroom context. This section discusses measures that could improve teachers' practice in using semantic mapping technique to help EFL learners develop their depth of knowledge of lexis in tandem with oral fluency. Therefore, specific pedagogical implications from two different perspectives of language processing and more instructional viewpoint will be presented.

7.5.3.1. Applicability of semantic mapping in relation to knowledge of lexis in L2 context

One of the significant pedagogical implications of the current study is the feasibility of implementation of semantic mapping technique in the way of developing lexical knowledge in Iran's EFL classroom context, a context which sets the primarily focus on developing the size of lexical knowledge without providing many opportunities for further improving the depth of knowledge. As Nation (2001) argues it is of utmost importance for EFL language teachers to provide chances for creative use of the lexical items for the language learners, meaning offering possibilities of meeting previously learned words subsequently in ways which differ from previous meeting with the words, as a result of which students can reconceptualize their knowledge.

Based on the findings of this study, semantic mapping can be considered as a technique targeting development of already known lexical knowledge; however, not in terms of individual items in isolation, but rather in conjunction with knowledge of other items, resulting in changes in number and strength of connections. For researchers and teachers seeking to exploit a lexical instructional practice, Meara (1996) suggested that exploring the lexical items as single entities expose the danger of "losing sight of the wood through concentrating too hard on the individual trees" (p. 9). That is, lexical items are best to be practiced in interlocking networks with items semantically related to one another. In this respect, the present study's findings can be useful for reference on how semantic mapping technique can be maintained in classroom context to present lexical items in a network of closely and strongly related items.

Next, to facilitate the effective application of this technique in developing knowledge of the lexis, it would be beneficial for teachers to organize explicit well-structured training sessions on semantic mapping and integrate them as part of the classroom culture. It is important to ensure that students understand the process and are aware of the differences between this type of mapping as a way to develop deeper knowledge of the previously known words and the technique of semantic mapping they frequently apply in the classroom as a means of learning new lexical items. It is crucial for teachers to highlight the importance of contribution of students in building up the map and also the significance of discussion (Stahl & Vancil, 1986), which is the most contributory factor in the process of developing deeper knowledge of lexical items.

However, it is not only the students who needs to be trained on how to practice semantic mapping but also teachers are required to know what to focus on and how to deliver the session to maintain the best and most desirable outcome. Hence, the first question worth attending to, prior to practicing semantic mapping and in the process of planning the session, is 'what to include and what to exclude in the process of semantic mapping?'.

The active nature of semantic mapping gives encouragement and motivation to the language learners to produce many lexical items and further broaden their knowledge by using them in the semantic context created in the process of discussion and dialogue with the teacher and peers. However, the process can be overwhelming given the amount of idiosyncrasy of responses, most particularly with the high frequency words. Therefore, the challenge for the language teachers would be not impeding the creation of a natural map but controlling the process of learning. In fact, the idea of mapping and creation of a network of lexical items should not be the case of putting the cart before the horse. Rather than teaching the associations, letting students to develop a word network in the form of a semantic map would be a better outcome of language teaching to target students' growth in this domain. This calls for the teacher's expertise and level of preparedness to control the process of mapping and the discussion so that it won't deviate from the main purpose and keep students on the track.

In addition to this all, the sense of variety and creativity semantic mapping brings about to the teaching context aids to remove the formality and strictness of the classroom and to move away from the tedious and boring processes of vocabulary practice. In fact, participants in the current study showed enthusiasm for engaging in the class and there were no dropouts, which in turn suggests that students value the practice of semantic mapping in the classroom context.

Consequently, although semantic mapping has been applied in classroom context for long, it has primarily been applied as a vocabulary teaching instrument (Nation, 2008). Nevertheless, it is advisable for language teachers to apply semantic mapping into classroom context not only as a way to teach language learners higher number of lexical items but also to practice the already known ones to provide chances for developing deeper knowledge of the words. Although it is of utmost importance that students know many lexical items, the quality of learning and how well each item is integrated with other words, is also of paramount significance. Finally, our findings showed that when semantic mapping is used on its own entity, it would be an ideal tool to support development of lexical knowledge and bringing the semantic relationships between items to consciousness, which in turn has processing benefits associated with it.

7.5.3.2. Applicability of semantic mapping in relation to the fluency of speech in L2 context

Another important implication of the current study is the possibility of undertaking semantic mapping technique in a foreign language classroom context that has often been known for practicing oral fluency solely incidentally in the process of free communicative speaking sessions. Tavakoli and Wright (2020) advise teachers to implement a variety of communicative and interactive activities in the classroom context to help EFL learners effectively develop fluency in instructional classroom setting, which is of more significant importance in Iran language teaching and learning context due to scarcity of opportunities for developing oral fluency outside classroom context.

The findings of this study have shown that semantic mapping technique increases cognitive processing speed and that this manifests itself in higher levels of utterance fluency. It can therefore be seen as an instructional practice which targets cognitive fluency in application of lexical items, in a very particular form which is specific to the items practiced by the task. Therefore, the findings of the current study suggest that the development of productive lexical network knowledge can play a key role in improving L2 fluency of speech in the traditional classroom context, and semantic mapping technique appears to provide learners with opportunities to integrate their knowledge of lexis in active spontaneous speech.

However, as it has already been mentioned (see Section 1.4), in traditional classroom context of Iran, vocabulary is often taught as a by-product of meaning-focused activities while consulting dictionaries to look up the meaning of words and finally memorizing the form-meaning relations. In other words, teachers in such context mainly concentrate on developing the breadth of lexical knowledge and rarely provide opportunities for students to be further engaged with the previously learned words to develop deeper knowledge of the words, which can ultimately support development of oral fluency.

The lack of practice is not only limited to the teachers but also González-Fernández and Schmitt (2017) mentioned that “textbooks usually do not recycle words to any great extent. The creation of supplementary materials (e.g., word games, speaking activities with a target word list) focusing on already-taught words will aid in their retention and elaboration” (p. 291). Accordingly, semantic mapping, as a fluency-focused classroom practice, can be justified by the argument that although many lexical items have already been met by students, chances for productive use of them (in the form of network of connections) as a way to develop oral fluency can be provided by its application. Hereby, by the application of this technique, teachers can provide chances for learners to semantically encode the lexical items which in turn paves the way to deeper processing of the elements, leading to easier and more efficient later retrieval (DeKeyser, 2017).

Finally, teachers might like to think about the particular aim of the students taking part in English classes prior to applying this technique. In case the aim is to integrate the language acquired into active speech, then application of semantic mapping might be a good choice to help learners in the way of developing oral fluency. However, some English classes, particularly in Iran context, are primarily grammar-based or focus on helping students to get prepared for the National University Admission Examination; hence, application of semantic mapping might be of less use. Nevertheless, semantic mapping has been traditionally practiced in EFL classroom context as a means of teaching new lexical items, and also sometimes as a pre-task phase of listening, reading, or writing; therefore, selling the technique to students as a means of improving fluency might not be an easy task. The section below discusses how semantic mapping can be sold as a fluency practice technique in classroom context.

7.5.3.2.1. Selling semantic mapping as a way to practice fluency

In case semantic mapping is going to be introduced in the classroom as a technique paving way towards more fluent production of L2, it is advisable to provide students with a clear rational so that they have a clear understanding of the underlying reasons for practicing it and don't assuming it as either a practice of lexical items or an exercise with no use and worthlessness. In terms of justifying and explaining the value of semantic mapping, teacher could take various number of approaches.

Initially, semantic mapping could be presented as a type of 'fluency training'. In this case, students were needed to be provided with detailed and clear explanations of fluency of speech and asked to reflect on their performance in terms of fluency. This reflection could take the form of a questionnaire of self-assessment in order to monitor speech performance with regard to fluency. In addition, as a part of providing chances for students to understand the reason for semantic mapping, by raising their awareness and getting them notice this particular aspect of language, chances were raised to contribute to the development of independent language learners.

However, prior to explaining the value of the semantic mapping to students, teachers' understating of fluency and the corresponding practice should have been promoted. Despite the fact that there has been notable developments in the realm of fluency, the practical implications of research are not effectively applied in the L2 pedagogy (Tavakoli, 2020). Tavakoli and Hunter (2018) argue that L2 teachers most often are either "using fluency and speaking ability interchangeably" (p. 330) or define it as general language proficiency. Accordingly, in order to make the best of the findings, teachers were required to be trained on how to explain the concept of L2 oral fluency on the first place, which would be in turn followed by raising self-awareness of the students to reflect on their practice.

The second approach in the way of explaining the value of the semantic mapping technique can be the use of PRAAT in order to show the students how the technique had helped them develop fluency of speech. Although this wouldn't be an easy task to do as either students or teachers need to be trained in analyzing speech data with the application of PRAAT, it would be really useful. Hilton (2014) recommends that L2 teachers could

> have at their disposal a simple application for easily marking up sound files: hesitations could be selected and tagged when occurring within a clause. An automatic count of numbers and lengths would give a clear, quantitative picture of the processing effort required by the task…[with] a much simpler user interface, for purposes of classroom assessment and action research (p. 45).

Accordingly, either PRAAT or a more user-friendly computer program to analyze speech samples would provide chances for the students to track changes in their speech.

7.5.3.3. Applicability of semantic mapping in relation to the organization of the mental lexicon in L2 context

Little attention in literature has been devoted to the implications of semantic mapping findings from a language processing perspective and its application into language teaching and learning context. That is, semantic mapping's findings have been rarely translated into language learning context taking into account models of bilingual lexicon. It can be argued that the realm of language processing has been mostly devoted to research rather than applying the findings into educational context due to various reasons, one of which can be inconsistencies in results and challenges faced in interpreting them. If we accept certain commonalities among models of bilingual mental lexicon, that is, the hierarchical organized conceptual and lexical levels of representation with links of various strength within and between the levels (see Section 2.4), it is feasible to consider a particular aspect of applicability of the present research to L2 language instruction.

Initially, one of the aims of this study was strengthening the links between L2 lexical representations and concepts with the aim of reducing L1 mediation to promote automaticity and fluency (Jiang, 2004); however, in EFL classroom context, teaching and practicing vocabulary through the use of L1 translation equivalents is very popular due to the relative efficiency and speed attributed to it. Jiang (2004) demonstrated that interlingual strategies, that is utilizing L1 for semantization, is desirable since it reduces the uncertainties associated with guessing from the context, by posing that "there is no reason not to use L1 as a means of semantization or as a tool for checking and validating learners' understanding of word meaning" (p. 426); however, this leads to strengthening the connections in the lexical level and mediation through L1. As it has been demonstrated in L2 lexical developmental stages, the connection between L2 item and concept is created in later stages of learning. This implies that learners at the earlier stages of L2 learning rely on the L1 translation equivalent of the lexical items to access the concept. This process is not only less direct but also slower and might have implications for the dysfluencies in speech.

The current study by the application of semantic mapping technique provided a platform for the students to develop and strengthen the links between the L2 lexical item and the concept more directly. This has been reached in three ways in the classroom context. Firstly, students were provided with chances of broadening their L2 lexical knowledge with more frequent meetings with the words and semantically encoding them, as a result of which activation of L1 translation equivalent is no longer necessary to access the concept. Secondly, chances for falling back into L1 translation equivalent as a way of practicing vocabulary was cut short by using L2 as the sole medium of communication. Finally, semantic mapping provided opportunities of instructional exposure to the lexical items already learnt with the purpose of semantic restructuring, as a result of which a direct link from L2 lexical item to concept will be formed. That is, semantic mapping, as a meaning-focused pedagogical activity, helped students to classify the lexical items into semantic groups while reflecting on both points of similarities and differences between lexical items in the same semantic field. Jiang (2004) argues that "once learners become aware of the differences, the word will assume a new identity. It will no longer be seen through the lens of the L1 but will become a lexical and semantic entity in its own right" (p. 427). Similar example of meaning-focused activities in the form of a word mapping and goal of semantic restructuring was presented in a study by Jullian (2000). This process has been tremendously more beneficial and feasible in the current study as the participants were not struggling with the core meaning of the items due to previous familiarity.

In conclusion, although use of L1 as a means of helping students understand the meaning of L2 words should not be considered as having a detrimental effect on the very first stages of learning a new lexical item, on the later stages of broadening lexical knowledge after creation of the initial form-meaning link, chances of strengthening of the direct link between the lexical item and concept should be provided by reducing chances of meeting L1 translation as a means of practice. Semantic mapping provided opportunities for L2 language learners to semantically encode the lexical items in L2 through the process of grouping the words into semantic fields, and discussion.

Finally, the results of the study can provide language teachers an effective technique in the way of developing students' vocabulary knowledge in a short duration of time, while subsequently proving that the learning gained with this technique allows students to apply the lexical items in spontaneous speech more fluently. This study provides insights into the ways of extending the application of semantic mapping as a vocabulary teaching and learning technique to supporting the development of known words and their use in real time communication with more automaticity.

7.5.3.4. L2 fluency assessment in classroom context

Based on the findings of the current study, it is argued that productive word association test demonstrates a potential usefulness to serve as an assessment tool in obtaining broad estimate of EFL learners' fluency of speech. It can be of tremendous usefulness for EFL teachers to administer such tests at intervals for diagnostic or monitoring purposes to assess students' fluency and oral ability in general terms.

Collecting speech samples from language learners, transcribing, coding, and analysing the data using speech analysis computer software, such as PRAAT, requires substantial amount of time and energy and might not be possible to practice frequently in classroom context. On the other hand, tests of productive word association, such as the computerized version of Lex30 designed by Meara and Fitzpatrick (2000), are convenient, easy to administer, and user-friendly; hence, the chances for teachers applying them in the classroom context to assess fluency of speech can be higher. Therefore, teachers are advised to look beyond knowledge of lexical items in productive tests of vocabulary to assess oral proficiency. Consequently, although the validity of such a practice is yet to be determined, the possibility seems pedagogically appealing.

7.6. Limitations of the study

While care has been taken to maximize the reliability of this study, a number of limitations must be acknowledged.

The first limitation regards the number of topics practiced during the intervention sessions, based on which word association tests were designed. Word association tests, based on which conclusions regarding the structure of the mental lexicon, have been drawn were only taken from two single topics; hence, exploring the local networks of associations in the lexicon. It needs to be pointed out that the findings of the word association tests can only provide a static approximation of what takes place in particular local strata of the lexicon, and it is necessary to acknowledge that only a small-scale snapshot of the large-scale phenomenon of the mental lexicon was taken into account. However, assessing the features and properties of mental lexicon as a global network of associations seemed impossible considering the constraints of the current research.

Secondly, the number of items used in word association test was limited to ten items per test. As Table ***4.1*** shows a few examples of word association studies, different researchers have applied various number of cue words; however, it seems that the number chosen heavily depends on the number of responses elicited. As the current study benefited from multiple responses with the aim of exploring the structural properties of the mental lexicon, higher number of cue words would make the process of coding and scoring overwhelming.

Next, the scope of the study was confined in terms of the number of participants (n = 33), which can be considered as a limitation of many empirical works. As the number of participants is not big and significance level is dependent on sample size, it was decided not to lower the alpha level and set it at 0.05; however, not setting the alpha level lower in case of multiple tests on same data set is considered another limitation of the current study.

In addition, the number of cue words in the word association test was limited to ten for each phase of the testing. This in fact leaves most studies, including the present one, vulnerable to the impact of individual differences in responses preferences. Although utmost measures were taken into account to ascertain that the drawn conclusions were theoretically and empirically justified, the chances would still remain that different participants would produce different responses on a different testing context leading to results not supporting the present conclusions. Hence, in order to generalize the findings of the current study, further research with higher number of participants and different cue words is encouraged.

Furthermore, this study investigated fluency in a monological mode of oral language, which imposed a semi-artificial context of speech. Although this was done deliberately to eliminate the less controlled and less predictable nature of speech performance, the speech does not necessarily reflect the natural flow.

Finally, an additional limitation from the methodological point of view, is the fact that participants received six sessions of training, practicing the lexical items in each topic for three sessions only, which seemed to be short for full lexical development and integration the lexicon. Nation (2020) has argued in favor of multiple meetings of lexical items in variety of context in addition to the quality and depth of processing each time each word is met. Although opportunities for elaboration and enrichment was provided in each meeting with the items, only few chances of practicing and meeting of the words was provided. However, the main reason for keeping the training sessions fairly short was reducing the level of classroom learning of the participants as a cofounding variable. Nevertheless, it needs to be acknowledged that the level of practice and familiarity with the lexical items was reduced at the same time.

7.7. Suggestions for future research

In this section, I will highlight a number of potential directions for future research in the realm of vocabulary and L2 fluency of speech.

In the very first place based on the results of the current study, it has been suggested that the changes in various measures of L2 fluency is highly correlated with the word association scores as a measure of structure of the mental lexicon. That is, improvements in measures of syllable run, phonation run, and syllable duration is associated with higher level of lexical availability as a result of higher number of closely related items in the center of the lexicon. However, it will be favorable if future research considers assessing the number of different lexical items in the elicited speech sample; that is, inclusion of measures such as lexical diversity

Subsequently, as the results of the current study revealed, it can be admitted that having access to a strong local cluster of associative networks with short average path among lexical items, that is an organized and advanced mental lexicon, is necessary to the fluent production of speech, however, it is not the only contributing factor. Further research is required exploring the impact of the syntactic patterns practiced via the application of the semantic mapping technique and through the process of discussion of the lexical items which could have resulted into fast and accurate application of the syntactic patterns in spontaneous production of speech, and consequently attributed to more fluent production of language.

Finally, drawing on the literature of the nature of pausing in speech, it has been argued that higher frequency of pauses is associated with lexical retrieval problems (De Bot, 1992; De Jong et al., 2013; Hilton, 2008; Kahng, 2014), and this fact warrants further investigation. For example, it would be very interesting indeed to carry out research on what each pause precedes, and if the reduction in the number of pauses is associated with the lexical items practiced ad retrieved during the intervention period. An alternative can also be exploring if a negative correlation can be found between the lexical items preceding the pauses and the items activated and retrieved during the word association task.

7.8. Final consideration

The study was carried out to respond to one of the main challenging issues frequently reported in literature considering the ultimate goal of many EFL learners; that is, to be able to talk easily, to express ideas and thoughts effortlessly and smoothly, without paying attention to its production, namely, to speak fluently. This study has shown that implementing semantic mapping technique as an instructional activity can be regarded as an effective hands-on experience that addresses this challenge and facilitate the process of production of fluent speech. The study also highlighted that an advantageous factor contributing to the desired outcome was the changes occurred in the organization of the mental lexicon in the process of semantic mapping. That is, the shifts in the structural properties of the mental lexicon, namely the connectivity and strength of the connections, contributed to the fast and easy availability of lexical items in the process of speech production. The findings of the current study extend understanding of the value of this technique as a valuable instruction tool and teaching technique of new lexical items to supporting the development of knowledge of the known words.

Appendices

Appendix A – Background questionnaire

Please complete the questionnaire with accurate information. Your participation in this study is anonymous; therefore, all identifiable data will be kept confidential by the researcher and will be destroyed two weeks after the data collection period of time. Thank you for your participation.

1. Name: ……………………………….
2. Email address: ………………………
3. Age: …………………………………
4. Job: …………………………………...

5. Gender

    - Male
    - Female

6. Native language:

    - Farsi
    - Other ……..…….

7. Highest level of education:

- Guidance school
- High school
- Associate degree
- Bachelor degree
- Master degree
- Doctorate

8. Do you know any other languages except your native language and English?

- Yes
- No

If Yes, which

9. Do you have any hearing, speaking, or vision problems?

- Yes
- No

If Yes, please explain

10. How old were you when you started studying English?

11. Have you continuously studied English since you started?
- [ ] Yes
- [ ] No

If No, please indicate the length of time you were not studying English

12. Are you currently taking any other English courses?
- [ ] Yes
- [ ] No

If Yes, please indicate the number of hours per week and type of classes that you are taking

13. Have you ever been to an English speaking country?
- [ ] Yes
- [ ] No

If Yes, where were you and how long did you stay

Thank you in advance for your time.
Best
Sara

Appendix B – Baseline proficiency test (C-test)

Participant no.: ____________

Task 5: Text reconstruction task

Please **complete the following texts**. You have **25 minutes** to for this task. Do not worry if you cannot complete all of the blanks in that time.

1. Geography

The UK is located on a group of islands known as the British Isles, which lie between the Atlantic Ocean and the North Sea, northwest of France. At i___ widest t___ UK i___ 300 mi___ across a___ 600 mi___ from No___ to So___. It sha___ a sin___ land bor___ with the Irish Repu___. Despite i___ relatively sm___ size t___ UK boa___ incredibly var___ and of___ very beau___ scenery, fr___ the mountains and valleys of the North and West to the rolling landscape of the South, and from downland and heath to fens and marshland.

2. UK Passport Service

A new passport office that has opened in London will help the UK Passport Service provide a much better service to customers who need a passport urgently. The n___ office ru___ on a___ appointment-only ba___, removing t___ need f___ a len___ wait bef___ being se___. The n___ building, Globe House repl___ the Petty France off___, which af___ 50 ye___ of conti___ service, h___ now clo___ its do___. The London Passport Office h___ the capa___ to issue 5000 passports weekly.

3. Record employment

Latest employment figures show that there are 28.2 million people in work. Work & Pensions Secretary Alistair Darling said this showed the UK labour market has coped well so far with the current international economic uncertainty. Mr Darling said: "Employment cont___ to ri___, with th___ month's fig___ showing a rec___ 28.2 mil___ people i___ work. Th___ are 65,000 mo___ people i___ work th___ last qua___ and 252,000 mo___ than la___ year. Alth___ both meas___ of unempl___ have ri___ slightly, th___ are st___ significantly lower than they were a year ago." The latest claimant count figures, for the month on Dec 13 2001, show a rise of 3,200 on the previous month. At 963,500 claimants, it remains 70,000 lower than this time last year.

4. Government consults on plans to modernize animal welfare

Plans to review, modernize and simplify outdated laws on animal welfare have been announced by the Government. Animal wel___ groups, lo___ authority represe___, courts, pol___ and indu___ are t___ be cons___ in wh___ will b___ a f___ reaching rev___ drawing toge___ the enviro___ and indus___ concerns o___ animal wel___. The Depar___ for t___ Environment, Fo___ and Ru___ Affairs (DEFRA) wants to hear views on the existing 11 Acts of Parliament governing the welfare of pets and farm animals.

5. Government is going in right direction on crime says Blunkett

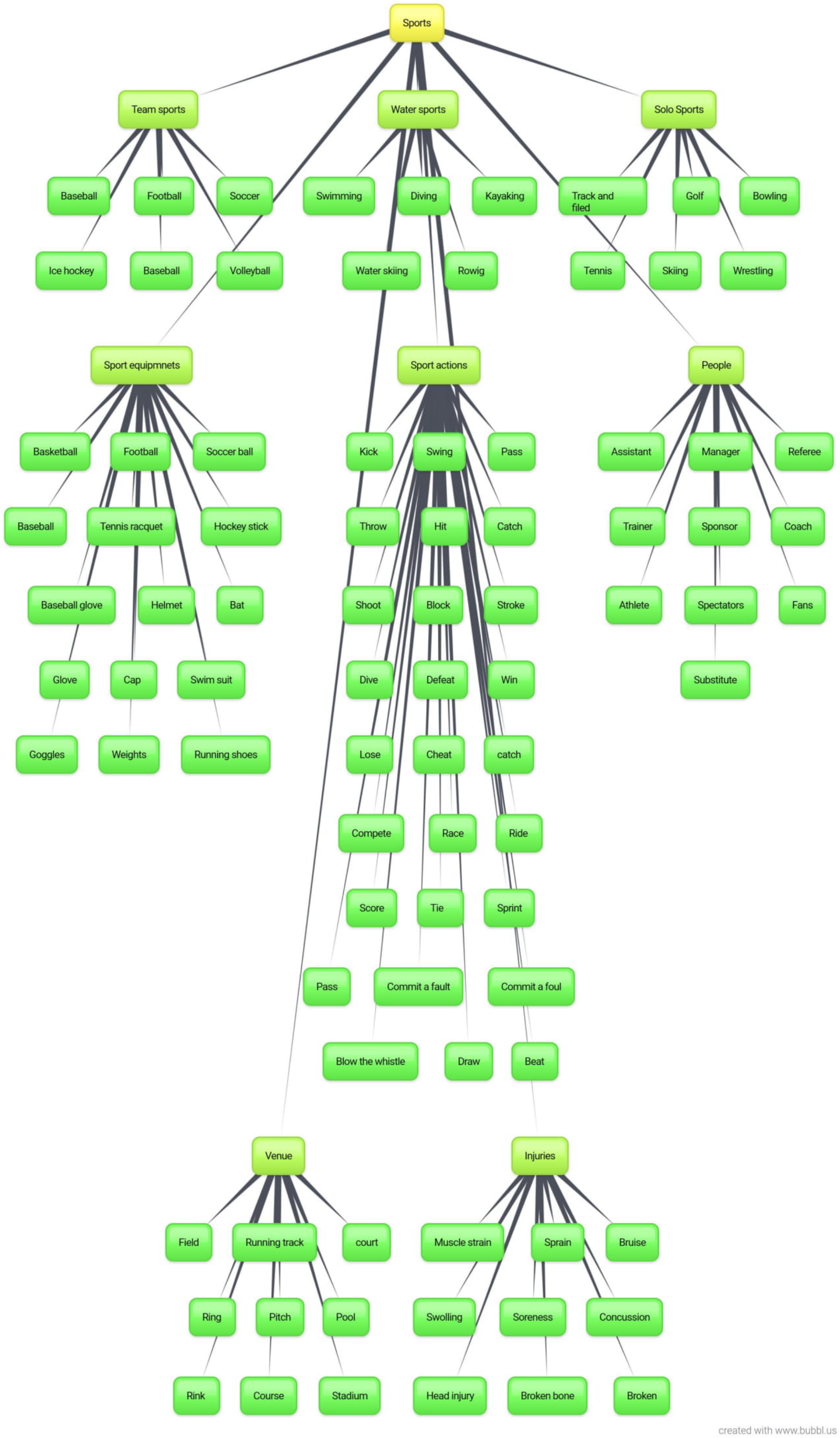
The chances of being a victim of crime are at their lowest level for 20 years – despite the worrying increases in mobile phone robbery – Home Secretary David Blunkett said today in a speech to a resident's group in Sheffield. This i___ being achi___ through mas___ Government inves___ not ju___ in t___ police ser___ but i___ education, emplo___, neighbourhood a___ community ren___, the devel___ of comm___ partnerships a___ record lev___ of spen___ on cr___ reduction. I___ his fi___ major spe___ of 2002 Mr Blunkett said the Government is delivering on its promises on crime, speeding up youth justice and targeting persistent offenders.

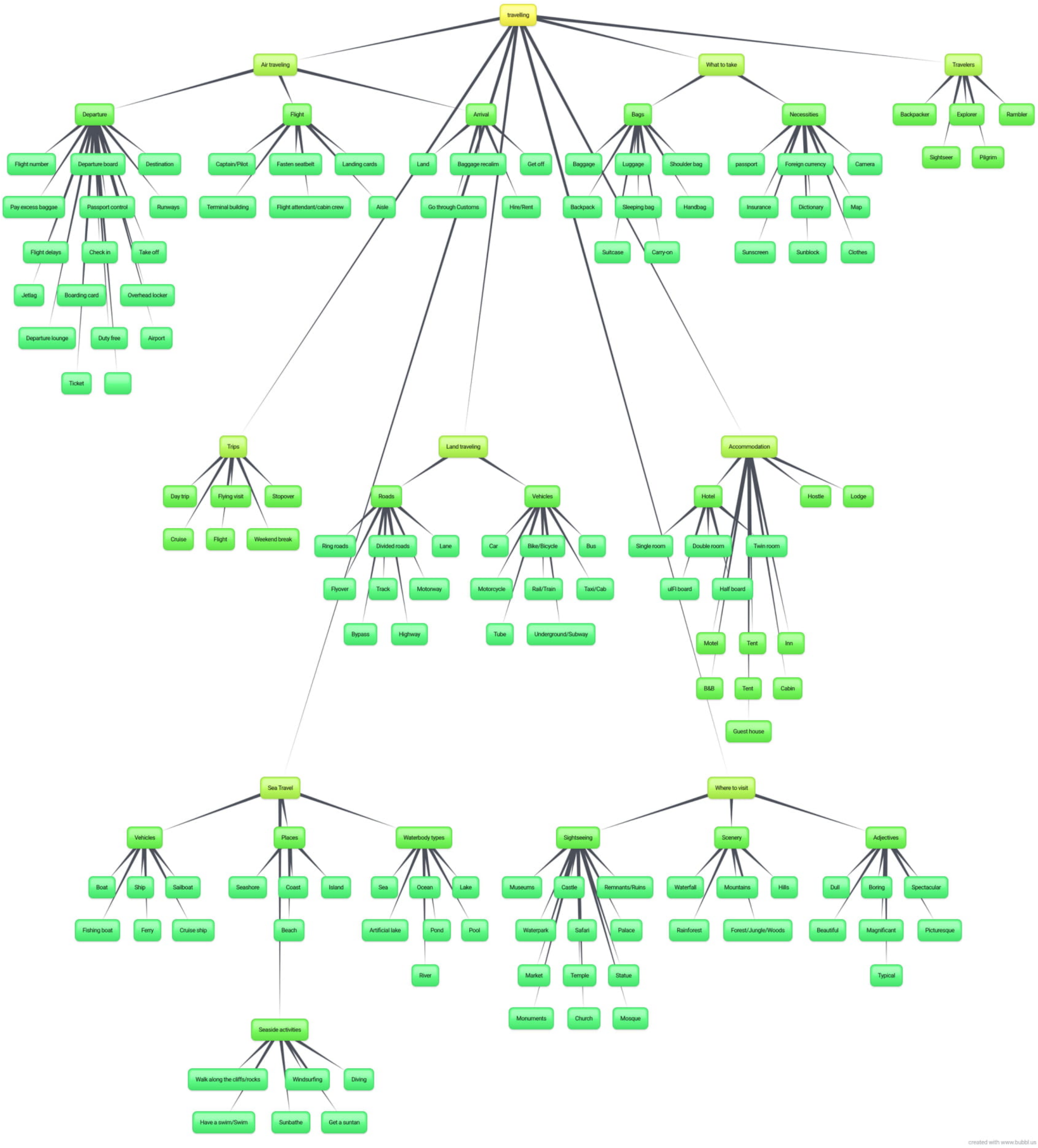
6. School for heroes

The internationally acclaimed Fire Services College in Gloucestershire is widely regarded as the best fire college in the world. The si___, a for___ airfield wh___ once laun___ Wellington bom___ for ra___ on Ger___, was rece___ transferred fr___ the Home Office t___ the Fire Services. A___ such i___ its reput___ that peo___ come fr___ as f___ away a___ China a___ Trinidad t___ sharpen u___ their leadership skills. Even fire chiefs from the other side of the world have attended as College students.

Daller, H. & Phelan, D. (2006). The C-test and TOEIC as measures of students' progress in intensive short courses in EFL. In Grotjahn, R. (Ed.). *The c-test: Theory, empirical research and applications*. Frankfurt: Lang.

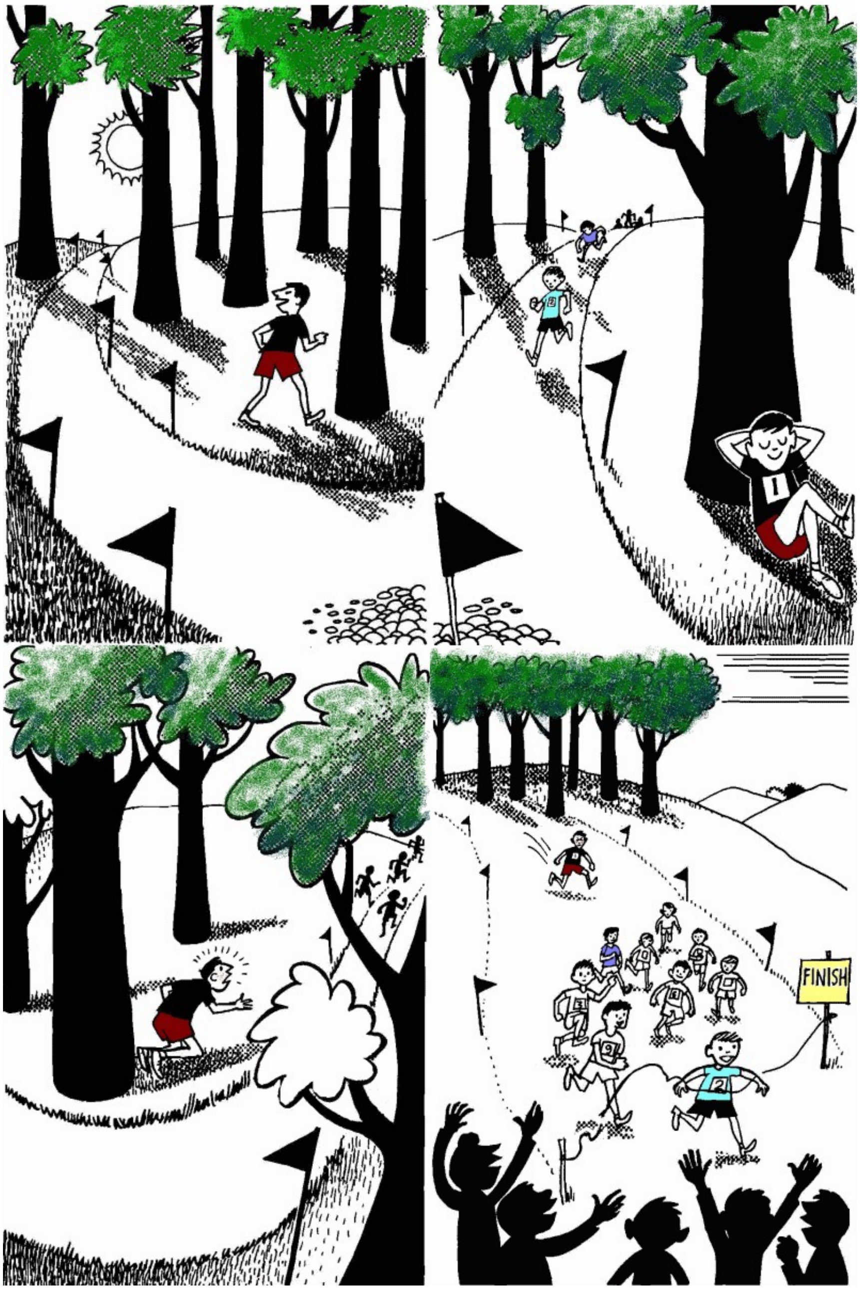
Appendix C – Semantic maps




travelling
Air traveling
What to take
Travelers
Departure
Flight
Arrival
Bags
Necessities
Backpacker
Explorer
Rambler
Sightseer
Pilgrim
Flight number
Departure board
Destination
Pay excess baggae
Passport control
Runways
Flight delays
Check in
Take off
Jetlag
Boarding card
Overhead locker
Departure lounge
Duty free
Airport
Ticket
Captain/Pilot
Fasten seatbelt
Landing cards
Terminal building
Flight attendant/cabin crew
Aisle
Land
Baggage recalim
Get off
Go through Customs
Hire/Rent
Baggage
Luggage
Shoulder bag
Backpack
Sleeping bag
Handbag
Suitcase
Carry-on
passport
Foreign currency
Camera
Insurance
Dictionary
Map
Sunscreen
Sunblock
Clothes
Trips
Day trip
Flying visit
Stopover
Cruise
Flight
Weekend break
Land traveling
Roads
Vehicles
Ring roads
Divided roads
Lane
Flyover
Track
Motorway
Bypass
Highway
Car
Bike/Bicycle
Bus
Motorcycle
Rail/Train
Taxi/Cab
Tube
Underground/Subway
Accommodation
Hotel
Hostle
Lodge
Single room
Double room
Twin room
ulFl board
Half board
Motel
Tent
Inn
B&B
Tent
Cabin
Guest house
Sea Travel
Vehicles
Places
Waterbody types
Boat
Ship
Sailboat
Fishing boat
Ferry
Cruise ship
Seashore
Coast
Island
Beach
Sea
Ocean
Lake
Artificial lake
Pond
Pool
River
Seaside activities
Walk along the cliffs/rocks
Windsurfing
Diving
Have a swim/Swim
Sunbathe
Get a suntan
Where to visit
Sightseeing
Scenery
Adjectives
Museums
Castle
Remnants/Ruins
Waterpark
Safari
Palace
Market
Temple
Statue
Monuments
Church
Mosque
Waterfall
Mountains
Hills
Rainforest
Forest/Jungle/Woods
Dull
Boring
Spectacular
Beautiful
Magnificant
Picturesque
Typical
created with www.bubbl.us

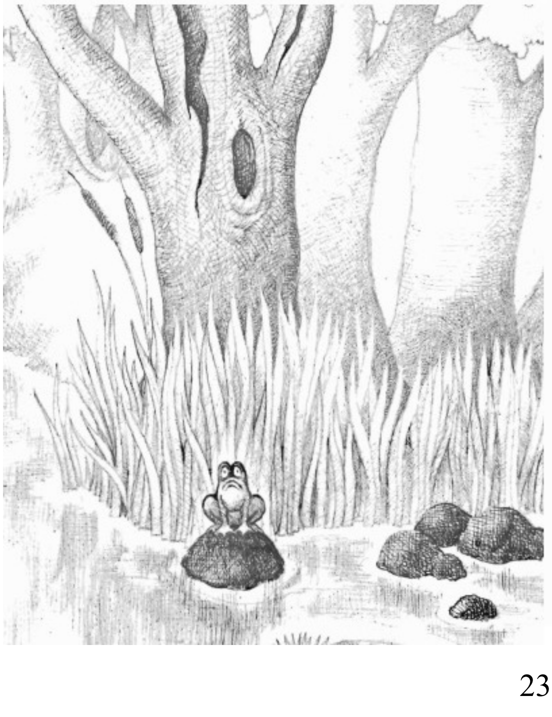
Appendix D – Picture narratives


START


FINISH








Appendix E – Picture narrative instruction

You are going to see a set of pictures telling a story. You have three minutes' time to plan what you want to say before telling the story. Meanwhile, you are provided with a piece of paper and a pen to make notes of what you want to say. Also, there are some questions on cards, presented along with the pictures, which can help you in your narration. You are asked to tell the story after you finish the three-minute planning time. Before you start to tell your story, your notes and the question cards will be taken away. Include as much details as you wish to talk for three to four minutes uninterruptedly.

Appendix F – Picture narrative question cue cards

Bicycle story

- In picture 1, what is the driver of the car doing?
- Why can't the car pass the boy on the bicycle?
- In picture 6, what is the boy doing?
- How do the people feel?

Frog story

- In picture 1, what is the boy trying to do?
- In picture 2, how was the dog trying to help the boy?
- In picture 4, where are the boy and dog going? Why?
- In picture 5, how does the frog feel?
- In picture 6, how do the boy and dog feel?

Race story

- In picture 1, what might the tall runner be saying?
- Why does the tall runner decide to take a nap?
- In picture 5, what is the tall runner thinking?
- Who wins the race?
- How do the boys feel in the end?

Appendix G – Word association test booklet and instructions

In this booklet, there are ten pages with spaces provided for your responses. Ten words will be presented via a power point. There will be a single word on each slide. Write down any words it brings to your mind. You can write as many responses as you can think of in the time provided. Flip the page as soon as the next word appears. For instance, if the word is **dog**, the words you think of can be **cat**, **bite,** or **animal**. In the case no word comes to your mind, just leave the page blank. Do not go back to the previous page and add more words or change your mind after you have written your responses.

In this booklet, there are twenty pages with spaces provided for your responses. Twenty words will be presented via a power point. There will be a single word on each slide. Write down any words it brings to your mind. You can write as many responses as you can think of in the time provided. Flip the page as soon as the next word appears.

Cue word: **Dog**

Animal	Loyal
Cat	Puppy
Bite	

Name:

Date:

-----------------------------------	-----------------------------------
-----------------------------------	-----------------------------------
-----------------------------------	-----------------------------------
-----------------------------------	-----------------------------------

-----------------------------------	-----------------------------------
-----------------------------------	-----------------------------------
-----------------------------------	-----------------------------------
-----------------------------------	-----------------------------------

-----------------------------------	-----------------------------------
-----------------------------------	-----------------------------------
-----------------------------------	-----------------------------------
-----------------------------------	-----------------------------------

Appendix H – Reading materials

Short stories were selected from American File book series. They were mainly chosen from the previous books and units that participants had already covered; hence, it was possible to claim that they had studied and met the vocabularies in the texts at one time. The mean length of the stories was tried to keep short at around 400 words. Although no new vocabulary was going to be taught, the comprehensibility of the text was important to be taken into account. The reason was that lack of knowledge of any lexical item in the text might shift the focus of attention of the students and learning new items incidentally would replace reviewing and practicing the already known ones, which is the aim of the current study.

Approximately 80% to 90% of the words in the stories belonged to the 2000 most frequent words of the British National Corpus (BNC; determined by Compleat Lexical Tutor; Cobb, n.d.). Many of the off-list words were proper names (e.g., Fre Lorz, Boris Onischenko, etc.) and also the academic words, which were less than 4% of the total, were the ones that they had already learnt previously and were intentionally inserted into the reading texts.

	Travel 1	Travel 2	Travel 3	Sports 1	Sports 2	Sports 3
	Welcome to Folkstone	Air Babylon	A night at White House Hotel	Famous cheating moments in sports	When you hear the final whistle	The London Marathon
Number of tokens	425	567	317	311	369	394
K1 + K 2	94.70	90.02%	97.20	88.06	81.62	81.21
AWL	.92	2.41	0	2.23	3.78	2.01

A night at White House Hotel

This happened about ten years ago when I was on holiday in England. We arrived in Cambridge very late. However, I knew the name of a hotel as a friend had recommended one before I left. It was the White House Hotel. I told the name to the taxi driver and he drove off, giving me a slightly strange look. After a while we arrived in a dark, narrow street and the taxi stopped outside an old building.

There was no lift in the hotel and the man at reception looked about a hundred years old. He showed us up to a room on the top floor – a long climb. There was no bathroom and the room smelled of damp but it was now midnight. There was nothing we could do, so we got undressed and went to bed.

We'd only been in bed for a few minutes when I heard my friend scratching. And then I began to feel itchy too. I switched on the light and saw that I was covered with red marks. Bed bugs! We went downstairs and asked the ancient receptionist for another room.

But he smiled and shook his head. 'All our rooms are the same', he said. Not wanting to get bitten all night, we decided to check out there and then. The old man gave us our money back without complaining – perhaps he was used to it. It was too late to find another hotel so we ended up sleeping in the bus station. Fortunately, it was August, so we didn't freeze to death. When I saw my friend again a few weeks later, I asked her why she'd recommended such a terrible place. 'White House?' she said, pulling a face. I never told you to go to the White *House* Hotel – everyone knows it's a dump. I said the White *Horse* Hotel!'

Air Babylon

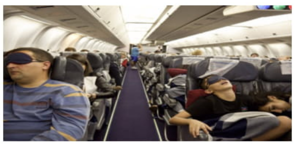
Air Babylon is the best-selling book, co-written by Imogen Edwards-Jones and anonymous airline employees whose identities must remain secret. It tells the "Inside story" flying and answers all questions.

Depending on what computer system the airline uses, check-in agents can talk to each other via simultaneous e-mail. So, when they seem to be taking a very long time to type you rather short name into the computer, they are probably sending one of their co-workers a message-usually about you or about someone in the line behind you. These messages range from "Have you seen this incredibly good-looking woman/man?" to "I've got a really difficult passenger here-does anyone have a seat next to a screaming child?" So, you can see it really does pay to be nice to the person in the counter.

There is a sensible drinking policy on all airlines, which means that we are not supposed to serve passengers if they start getting noisy, but some cabin crew members think that if you give them enough to eat and drink, they will eventually fall asleep and give you no trouble at all. And as every flight attendant knows, a snoring plane is a happy plane. That's the reason, of course, why we like to turn the heat up halfway through a flight…

Some airports are notorious for losing passengers' luggage. Heathrow Airport has a poor reputation – most airports lose about two in every thousand bags, but Heathrow loses 80 per thousand, which means for every 500 people who check in, 40 won't get their bags or suitcases at the other end! This is mainly because the transport times between terminals are so tight! When the airport is busy, which it always is, there is so much luggage being transported between the terminals and so little time to do it that a lot of the transferred luggage gets left behind.

Wheelchairs are a big problem for us. Not only is there always a shortage of them for the people who really need them, but worse still, some of the people who request them often don't need them at all. I've lost count of the number of times I have pushed someone through the airport taken them through the customs and passport control, and gotten a porter to pick their luggage, and then seen the person jump up in Arrivals and sprint toward their waiting relatives. One flight attendant I know gets so annoyed when this happens that as soon as the passenger gets out of the chair, she shouts, "Ladies and gentlemen! I give you another miracle, courtesy of the airline industry! After decades in a chair, he walks again!" the passenger is usually so embarrassed that he (and it is usually a he) disappears as quickly as he can.



Birds are one of the major problems for ant airport when planes are taking off or landing. A swan or any large bird can easily cause an accident. It flies into the engine, totally destroying itself and the machinery. Smaller birds are less of a problem. In some cases, they can do some damages, but more often than not they are just roasted. When this happens, there is often such a strong smell of roast bird that passengers on the plane think that chicken is being cooked, and they are often surprised when they are given a choice of fish or beef for dinner!

Famous cheating moments in sports

Divine intervention?

Soccer

Argentine was playing England in the quarterfinals of the 1986 World Cup in Mexico. In the 52^{nd} minute the Argentinian captain, Diego Maradona, scored a goal. The English players protested, but the referee allowed the goal. However, TV cameras showed that Maradona had scored the goal with his hand! Maradona said the next day, "it was partly the hand of the Maradona, and partly the hand of the God." Later in the game, Maradona scored another goal and Argentina won 2-1. They went on to win the World Cup.

With a little help from my friends

Track and Filed

Fre Lorz, from New York, won the marathon at the St. Louis Olympic Games in 904. He finished the race in three hours 13 minutes. After the race, Fred was waiting to get his medal, and the spectators were cheering him loudly. Alice Roosevelt, the daughter of the US President, was in the crowd, and some journalist took a photo of Fred with her. But then suddenly somebody started shouting "cheater" and soon everybody was shouting the same thing. It was true. Fred had travelled 18 of the 42 kilometres in somebody's car! Fred didn't win the gold medal and he was banned from track and field.

Dishonischenko!

Fencing

Boris Onischenko, an army officer from the Soviet Union, was competing against Jim Fox from Britain in 1976 Montreal Olympics. Boris was winning and the electronic scoreboard was showing hit after hit from him. Jim Fox protested to the referee. Fox said that Boris was scoring points without hitting him. Olympic officials examined Boris's sword and they made a shocking discovery. Boris had changed the electronic part of his sword. He could turn on the hit light on the scoreboard even when he hadn't hit Fox. Boris went home the next day, in disgrace. The British newspapers called him "Dishonischenko"

The London Marathon

On Sunday 23 April, more than 30,000 people participated in the 2017 London Marathon, running 26.2 miles (42.2 kilometers) around the city centre.

The London Marathon has taken place every year since 1981, making this year's the 37th race. At the first one just 6,225 people completed the course, and since then it has grown into one of the biggest sporting events in the world. It is one of the 'World Marathon Majors', which is a competition made up of six global marathon events. The overall male and female winners receive $1 million in prize money between them. There are currently eleven people who have run every London Marathon since it began – they call themselves the 'Ever Presents'.

This year, the winners of the London Marathon were both from Kenya — Daniel Wanjiru, who finished in 2 hours and 5 minutes, and Mary Keitany, whose time was 2 hours and 17 minutes. In the wheelchair race, David Weir from the UK and Swiss athlete Manuela Schär were the champions.

However, most marathon runners are not professional athletes. They are amateurs who have trained for months to raise money for charity or just as a personal challenge. Many people choose to do fun things to make them stand out, like wear silly costumes or do things instead of just running. For example, Tom Harrison took three days just to reach the run's halfway point, as he is crawling the course on his hands and knees dressed as a gorilla, trying to raise money and attention for the charity 'The Gorilla Organization'.

Another highlight from this year's race is the story of Jackie Scully and Duncan Sloan, who got married on the morning of the marathon. They both then ran the race, with Jackie doing it in her wedding dress! They were running to raise money for cancer charities, as Jackie was diagnosed with breast cancer three years ago.

Matthew Rees was another runner to inspire people, as he stopped just before the finish line to help David Wyeth run the last 200m. They did not know each other before, but Matthew saw that David was exhausted and possibly close to collapsing and wanted to help him finish. Every year there are lots of inspiring stories like this from the London Marathon, and it is a huge achievement even to run the 26.2 miles!

Welcome to Folkstone

The plane

At 4.15 a.m. a taxi picked me up and took me 32 miles to Stansted airport. Although it was early morning, there was a lot of traffic and I arrived later than I had planned. I took my luggage to check-in and asked for a window seat, but the woman said there were no seat numbers. When I got to the security, I saw that there was a really long line. I began to worry that I might miss my flight because you have to board 40 minutes before take-off. I had to run to gate 48 and I arrived completely out of breath.

We boarded. Because there are no seat numbers on these flights, everybody tries to get on as quickly as they can. I sat next to a friendly Frenchman. We took off and soon I was looking down on London. There was no meal, not even coffee, but we landed 10 minutes ahead of schedule. I only had to wait 20 minutes for my luggage. Then I walked out into bright sunshine and waited for the bus to Avignon, about 40 kilometres away. I didn't have to wait long, and the bus took 45 minutes. It was only 11:00 and I had the whole day in front of me. My ticket cost £63, and I gave the trip 5/10 for comfort, and 5/10 for convenience.

The train

At 7.10 a.m. I arrived at Waterloo station by taxi. It took me just 30 minutes from home. I bought a newspaper and walked to the platform. I got on and found my seat. As soon as we started moving, I went to find the dining car and had a cup of coffee.

I looked out of the window. Although we were moving at 340 kilometres an hour, the trip was smooth and relatively quiet. The part where we travelled under the English Channel took 22 minutes. Soon I was looking at the fields and farmhouses of France. The sun was shining. I closed my eyes and went to sleep.

I arrived on time! I picked up my suitcase and followed the Exit signs. It was great not to have to wait for my luggage or to worry about getting a bus or taxi downtown. Just outside the station, I looked up and saw the medieval walls of Avignon's historic city canter. It was 2.20 in the afternoon and I was just in time for the late lunch! My ticket cost £65.80, and I gave the trip 8 /10 for comfort and 9/10 for convince.

When you hear the final whistle!

One of the hardest things for any professional athlete to do is to know when to retire. Do you retire when you are at your physical peak or do you wait until your body (or your coach) tells you that it is time to go? But even harder is finding the answer to the question "What am I going to do with the rest of my life?"

"There is a high risk of depression and people often find adjusting to a new way of life difficult." Says Ian Cockerill, a sports psychologist. "For athletes, there is an extra trauma- the loss of status, the loss of recognition, and the loss of glamour. That is the hardest part." As Eddie Acaro, the US jockey says, "When a jockey retires, he becomes just another little man."

Perhaps they just can't stand life without the high of playing professional sports. Michael Jordan, the greatest basketball player of all time, retired three times. He retired once from the Chicago Bulls, made a successful comeback with Bulls, and then retired again. His second comeback with an inferior team ended in failure, and he retired forever at the age of 40. Jordan said, "There will never be anything I do that will fulfil me as much as competing did."

Muhammad Ali needed the money, but his comeback fight at the age of 39, against Trevor Berbick, was one of the saddest spectacles in modern sports. After losing to Berbick, Ali retired permanently. Three years later, he developed Parkinson disease.

As Jimmy Greaves, a former soccer player for England, said "I think that a lot of players would prefer to be shot once their career is over." Many of them spend their retirement in a continual battle against depression, alcohol, or drugs.

Franz Beckenbauer is a classic example of a soccer player who won everything with his club, Bayern Munich. After retiring, he became a successful coach with Bayern and finally president of the club. Jon McEnroe, the infamous "bad boy" of tennis, is now a highly respected and highly paid TV commentator. But sadly, for most professional athletes these cases are the exceptions.

Appendix I – Audit form

UNIVERSITY of York

Education Ethics Committee
Ethical Issues Audit Form

This questionnaire should be completed for each research study that you carry out as part of your degree. Once completed, please email this form to your supervisor. You should then discuss the form fully with your supervisor, who should approve the completed form. **You must not collect your data until you have had this form approved by your supervisor (and possibly others - your supervisor will guide you).**

Surname / Family Name:	Ebrahimi
First Name / Given Name:	Sara
Programme:	PhD TESOL
Supervisor (of this research study):	Dr Zoe Handley
Topic (or area) of the proposed research study:	
The impact of semantic mapping technique on the organization of bilingual mental lexicon and L2 utterance fluency of Iranian EL learners	
Where the research will be conducted:	
Private language school in Iran	
Methods that will be used to collect data:	
Vocabulary and speaking tests	
If you will be using human participants, how will you recruit them?	
Personal and professional connections	

All supervisors, please read *Ethical Approval Procedures: Students*.
Taught programme supervisors. Note: If the study involves children, vulnerable participants, sensitive topics, or an intervention into normal educational practice, this form must also be approved by the programme leader (or Programme Director if the supervisor is also the Programme Leader)
Research student supervisors. The application is a joint one by the research student and supervisor(s). It should be submitted to the TAP member for initial approval and then to the Higher Degrees Administrator who will seek a second opinion from a designated member of Education Ethics Committee.
All students: forms may also require review by the full Ethics Committee (see below).

<u>First approval:</u> by the supervisor of the research study **(taught students);** or TAP member **(research students)**(after reviewing the form):

Please select one of the following options.

I believe that this study, as planned, meets normal ethical standards. I have checked that any informed consent form a) addresses the points as listed in this document, and b) uses appropriate language for the intended audience(s).	X
I am unsure if this study, as planned, meets normal ethical standards	☐
I believe that this study, as planned, does not meet normal ethical standards and requires some modification	☐

Supervisor/TAP member's Name (please type):	Nadia Mifka-Profozic (TAP member)
Date:	23 February 2018

Taught student supervisors - If the study involves children, vulnerable participants, sensitive topics, or an intervention into normal educational practice (see *Ethical Approval Procedures: Students*), please email this form for second approval to the Programme Leader (or Programme Director if the supervisor is also the Programme Leader). For this second approval, other documents may need to be sent in the same email e.g. the proposal (or a summary of it) and any informed consent and participant information sheets. If the study has none of the above characteristics, the supervisor should email this completed form to the Programme Administrator. This signals the end of the approval process and data collection can begin. The member of the EEC will notify the Programme Administrator only when the final outcome has been decided.

<u>Second approval:</u> by the Programme Leader; or Programme Director; or designated Ethics Committee member for research students:

Please select one of the following options:

I believe that this study, as planned, meets normal ethical standards. I have checked that any informed consent form a) addresses the points as listed in this document, and b) uses appropriate language for the intended audience(s).	☐
I am unsure if this study, as planned, meets normal ethical standards	☐
I believe that this study, as planned, does not meet normal ethical standards and requires some modification	☐

Name of Programme Leader; or Programme Director; or Ethics Committee member (please type):	
Date:	Click here to enter a date.

The supervisor should now <u>email</u> this completed form to the Programme Administrator, unless approval is required by the full Ethics Committee (see below).

Approval required by the full Education Ethics Committee

If the application requires review by the full Education Ethics Committee, please select one of the following options then forward the application to the Research Administrator (education-research-administrator@york.ac.uk).

The study involves deception	☐
The study involves an intervention and procedures could cause concerns	☐
The topic is sensitive or potentially distressing	☐
The study involves vulnerable subjects	☐
Other reason:	

Name of Programme Leader; or Programme Director; or TAP member (please type):	
Date:	Click here to enter a date.

FOR COMPLETION BY THE STUDENT

Data sources

1 If your research involves collecting secondary data only **go to SECTION 2**.

2 If your research involves collecting data from people (e.g. by observing, testing, or teaching them, or from interviews or questionnaires) **go to SECTION 1.**

SECTION 1: For studies involving people

3 Is the amount of time you are asking research participants to give reasonable? **YES**

4 Is any disruption to their normal routines at an acceptable level? **YES**

5 Are any of the questions to be asked, or areas to be probed, likely to cause anxiety or distress to research participants? **NO**

6 Are all the data collection methods used necessary? **YES**

7 Are the data collection methods appropriate to the context and participants? **YES**

8 Will the research involve deception? **NO**

9 Will the research involve sensitive or potentially distressing topics? (The latter might include abuse, bereavement, bullying, drugs, ethnicity, gender, personal relationships, political views, religion, sex, violence. If there is lack of certainty about whether a topic is sensitive, advice should be sought from the Ethics Committee.) **NO**

If YES, what steps will you take to ensure that the methods and procedures are appropriate, not burdensome, and are sensitive to ethical considerations?

10 Does your research involve collecting data from vulnerable or high risk groups? (The latter might include participants who are asylum seekers, unemployed, homeless, looked after children, victims or perpetrators of abuse, or those who have special educational needs. If there is a lack of certainty about whether participants are vulnerable or high risk, advice should be sought from the Ethics Committee. Please note, children with none of the above characteristics are not necessarily vulnerable, though approval for your project must be given by at least two members of staff; see above). **NO**

If YES, what steps will you take to ensure that the methods and procedures are appropriate, not burdensome, and are sensitive to ethical considerations?

11 Are the research participants under 16 years of age? **NO**
If NO, go to question 12.

If YES, and you intend to interact with the children, do you intend to ensure that another adult is present during all such interactions? Choose an item.

If NO, please explain, for example:
i) This would seriously compromise the validity of the research because [***provide reason***]

ii) I have/will have a full Disclosure and Barring Service check (formerly Criminal Records Bureau check). Choose an item.

iii) Other reasons:

Payment to participants

12 *If research participants are to receive reimbursement of expenses or any other incentives, including financial, before or after the study, please give details. You should indicate what they will receive and, briefly, the basis on which this was decided.*

I will send the participants thank you cards after the data collection

It is often considered good practice to consider what the researcher might offer the participants, in the spirit of reciprocity. Some ideas of what this might be include: materials at the end of the study, a workshop summarising the results of the study, a delayed treatment/intervention at the end of the study, an indication about where the findings might be accessed at a later date, a letter or token of thanks. Please ensure that you have considered the potential for reciprocity in your research.

If your study involves an INTERVENTION i.e. a change to normal practice made for the purposes of the research, go to question 13 (this does not include 'laboratory style' studies i.e. where ALL participation is voluntary):

If your study does not involve an intervention, go to question 20.

13 Is the extent of the change within the range of changes that teachers (or equivalent) would normally be able to make within their own discretion? **YES**

14 Will the change be fully discussed with those directly involved (teachers, senior school managers, pupils, parents – as appropriate)? **YES**

15 Are you confident that *all* treatments (including comparison groups in multiple intervention studies) will potentially provide some educational benefit that is compatible with current educational aims in that particular context? (Note: This is *not* asking you to justify a non-active control i.e. continued normal practice) **YES**

Please **briefly** describe this / these benefit(s):

The experimental groups will receive vocabulary training treatment with application of semantic maps, working on the elaboration of their semantic networks, which is expected to result in activating and retrieving words more quickly, helping them to speak more fluently; whereas, the control group is going to recxieve no treatment on semantic mapping.

16 If you intend to have two or more groups, are you offering the control / comparison group an opportunity to have the experimental / innovative treatment at some later point (this can include making the materials available to the school or learners)? **YES**

If NO, please explain:

17 If you intend to have two or more groups of participants receiving different treatment, do the informed consent forms give this information? **YES**

18 If you are randomly assigning participants to different treatments, have you considered the ethical implications of this? **YES**

19 If you are randomly assigning participants to different treatments (including non-active controls), will the institution and participants (or parents where participants are under 16) be informed of this in advance of agreeing to participate? **YES**

If NO, please explain:

General protocol for working in institutions

20 Do you intend to conduct yourself, and advise your team to conduct themselves, in a professional manner as a representative of the University of York, respectful of the rules, demands and systems within the institution you are visiting? **YES**

21 If you intend to carry out research with children under 16, have you read and understood the Education Ethics Committee's *Guidance for Ethical Approval for Research in Schools*? **N/A**

Informed consent

22 Have you prepared Informed Consent Form(s) which participants in the study will be asked to sign, and which are appropriate for different kinds of participants? **YES**

If YES, **please attach the informed consent form(s).**

If NO, please explain:

23 Please check the details on the informed consent form(s) match each one of your answers below. Does this informed consent form:

a) inform participants in advance about what their involvement in the research study will entail? **YES**

b) if there is a risk that participants may disclose information to you which you may feel morally or legally bound to pass on to relevant external bodies, have you included this within a confidentiality clause in your informed consent form? **N/A**

c) inform participants of the purpose of the research? **YES**

d) inform participants of what will happen to the data they provide (how this will be stored, who will have access to it, whether and how individuals' identities will be protected during this process)? **YES**

e) if there is a possibility that you may use some of the data publicly (e.g. in presentations or online), inform the participants how identifiable such data will be **and** give them the opportunity to decline such use of data? **YES**

f) give the names and contact details (e.g. email) of at least two people to whom queries, concerns or complaints should be directed? One of these people should be on the Education Ethics Committee (please use education-research-administrator@york.ac.uk) and not involved with the research. **YES**

g) in studies involving interviews or focus groups, inform participants that they will be given an opportunity to comment on your *written record* of the event? **N/A**

If NO, have you made this clear this on your consent form? Choose an item.

If NO, please explain why not:

h) inform participants how long the data is likely to be kept for? **YES**

i) inform participants if the data could be used for future analysis and/or other purposes? **YES**

j) inform participants they may withdraw from ACTUAL the study during data collection? **YES**

k) provide a date/timescale by which participants will be able to withdraw their data and tell the participants how to do this? (NB. If your data is going to be completely anonymized, any withdrawal of data needs to happen before this.) YES

**NA if your data will be anonymous at point of collection*

If your answer was NO to any of the above, please explain here, indicating which item(s) you are referring to (a-j):

24 Who will be asked to sign an Informed Consent Form? Please select all that apply:

CATEGORY	
Adult research participants	✓
Research participants under 16	☐
Teachers	☐
Parents	☐
Head/Senior leadership team member	✓

Other (please explain)	☐

25 In studies involving an **intervention** with under 16s, will you seek informed consent from parents?

N/A

If NO, please explain:

If YES, please delete to indicate whether this is 'opt-in' or 'opt-out'
If 'opt-out', please explain why 'opt-in' is not being offered:

SECTION 2

Data Storage, Analysis, Management and Protection

26 I am accessing data from a non-publicly available source (regardless of whether the data is identifiable) e.g., pupil data held by a school or local authority, learners' work. NO

If YES, I have obtained written permission, via an informed consent document, from a figure of authority who is responsible for holding the data. This informed consent a) acknowledges
responsibility for releasing the data and b) confirms that releasing the data does not violate any informed consents or implicit agreements at the point the data was initially gathered.

Choose an item.

27 I have read and understood the Education Ethics Committee's *Guidance on Data Storage and Protection* YES

28 I will keep any data appropriately secure (e.g. in a password protected file/ locked cabinet – delete as appropriate), maintaining confidentiality and anonymity (e.g. identifiers will be encoded and the code available to as few people as possible) where possible. YES

29 If your data can be traced to identifiable participants:

a) who will be able to access your data?

Sara Ebrahimi

b) approximately how long will you need to keep it in this identifiable format?

It will be identifiable for approximately two months, to the end of data collection session and two weeks after that. Afterwards names will be replaced by codes and will be anonymized for further analysis. The anonymized data will be stored with the Research Data York service for further use.

30 If working in collaboration with other colleagues, students, or if under someone's supervision, please discuss and complete the following:

We have agreed:
a) [*Sara Ebrahimi*] will be responsible for keeping and storing the data
b) [*Dr. Zoe Handley and Sara Ebrahimi*] will have access to the data
c) [*Dr. Zoe Handley and Sara Ebrahimi*] will have the rights to publish using the data

Reporting your research

31 In any reports that you write about your research, will you do everything possible to ensure that the identity of any individual research participant, or the institution which they attend or work for, cannot be deduced by a reader? YES

If NO please explain:

Conflict of interests

32 If the Principal Investigator or any other key investigators or collaborators have any direct personal involvement in the organisation sponsoring or funding the research that may give rise to a possible conflict of interest, please give details:

Potential ethical problems as your research progresses

33 If you see any potential problems arising during the course of the research, please give details here and describe how you plan to deal with them:

N/A

Student's Name (please type):	Sara Ebrahimi
Date:	26 January 2018

Please email this form to your supervisor. They must approve it, and send it to the Programme Administrator by email.

NOTE ON IMPLEMENTING THE PROCEDURES APPROVED HERE:

If your plans change as you carry out the research study, you should discuss any changes you make with your supervisor. If the changes are significant, your supervisor may advise you to complete a new 'Ethical issues audit' form.

For Taught Masters students, on submitting your MA dissertation to the programme administrator, you will be asked to sign to indicate that your research did not deviate significantly from the procedures you have outlined above.

For Research Students (MA by Research, MPhil, PhD), once your data collection is over, you must write an email to your supervisor to confirm that your research did not deviate significantly from the procedures you have outlined above.

Appendix J – Head teacher consent form

THE UNIVERSITY of York

DEPARTMENT OF EDUCATION
Heslington, York, YO10 5DD
Tel: (01904) 323460
Web: *http://www.york.ac.uk/education*

INFORMATION SHEET FOR HEAD TEACHER
The Effect of Vocabulary Training on EFL learners' Language proficiency

Researcher: Sara Ebrahimi

(Supervisor: Dr. Zoe Handley).

This document explains why we are conducting this research project and sets out what will be involved for the language school.

What is the purpose of the study?

This research project aims to investigate the effect of vocabulary training on EFL learners' language proficiency.

What sort of participants do we need?

For this research project, we are looking for adult language learners learning English as their foreign language, aged between 18 to 30, at intermediate level of language proficiency.

Who will give consent for a participant to take part?

I will get consent forms from the participants themselves (see attached information sheet and consent form). It will be made clear that the study is entirely voluntary, and it doesn't have any impact on their marks, and also, they are free to withdraw within the data collection period of time and two weeks after this period without giving a reason.

What will be involved?

I will take every care to reduce to a minimum disruption to the language school routine. The study does not need any special requirements since it is going to take place in the language classroom context. The intervention will take place during normal classes. Those participants who give consent to take part in the study are randomly allocated into four different classes and those who are not interested to participate, will be placed in another class.

Students whom are readily available and accessible at the time will form the participants of the study, and their level of language proficiency is checked via C-test. They will be assigned to two different groups. The process of placement in different groups will be totally random. This study involves three main stages;

- Speaking tests; There will be three different speaking tests, each taking only four minutes for each participant. The first will be administered prior to the second session, the second on the tenth session and finally the last one on the sixteenth session of the term.

- Test on vocabulary; There will be three different tests on participants' knowledge of vocabulary. Each test takes 9 minutes and will be administered on three different sessions, one on the third session of the term, the second on the eleventh session, and the last one on the seventeenth session of the term.

- Vocabulary training; There will be six sessions of vocabulary training, each taking forty-five minutes in six consecutive sessions (the fourth, fifth, sixth, seventh, eighth, nineth sessions of the term).

Who will run the research sessions?

Sara Ebrahimi will meet the participants as the teacher of the classes and run the sessions, and also take the tests.

Will all the participants' details and the assessment results be kept confidential?

Yes. All the information provided by the participants will be identifiable for only the period of data collection, and only the researcher, Sara Ebrahimi, will have access to the identifiable data. After data collection, all names will be replaced by codes. The anonymized data will be kept confidential indefinitely afterwards and stored securely with the Research Data York service, which is a professional University of York data management service.

I hope you allow me to conduct this research in this language school. Please keep this information sheet for your own records and thank you for taking the time to read this information.

Contact:

If you require any further information or have any questions about this study, please do not hesitate to contact Sara Ebrahimi by email (Sara.ebrahimi@york.ac.uk), Dr Zoe Handley as the research supervisor (Zoe.handley@york.ac.uk) or the Chair of Ethics Committee via email education-research-administrator@york.ac.uk

Address: Department of Education, The University of York, York, YO10 5DD

Department of Education, University of York

CONSENT FORM FOR HEAD TEACHER
Sara Ebrahimi

Declaration of Consent

I have been informed about the aims and procedures involved in the research project described above. I reserve the right to withdraw any participant at any stage during data collection, and the two weeks after that and also to terminate the project altogether if I think it necessary.

I understand that the participation in the study is voluntary and information gained will be anonymous and that participant's names will be removed from any materials used in the research.

Name:

Signed:

School:

Date:

Please email the signed form to Sara.ebrahimi@york.ac.uk

Appendix K – Participants' consent form

THE UNIVERSITY of York

DEPARTMENT OF EDUCATION

Heslington, York, YO10

5DD

Tel: (01904) 323460

Web: http://www.york.ac.uk/education

Information Page

The Effect of Vocabulary Training on EFL learners' Language proficiency

Dear Participant,

I, Sara Ebrahimi, am currently carrying out a research project to investigate the effect of vocabulary training on EFL learners' language proficiency. I am writing to ask if you are able to take part in the study.

What would this mean for you?

You will be asked to complete three different type of tests. The first one is an English proficiency test to see what level of English you are at, the second is a vocabulary task where you will be presented with a number of words and you will be asked to provide as many related words as you can remember. The third is a speaking task. You will be invited to tell a story while being presented with a six-frame picture. There are also six training sessions on vocabulary. The study is aimed at finding out if training you with vocabulary has a positive effect on your speaking. Your participation is totally voluntary in this study, and your allocation into different groups is done randomly.

Anonymity

The data that you provide (the task results and the audio recording) will be stored by code numbers. Any information that identifies you will be stored separately from the data.

Storing and using your data

During the data collection period of time, which will approximately take three months, all data will be named and identifiable. Two weeks after the end of data collection, all names will be replaced by codes, and participants will not be identified individually. They remain anonymous afterwards. The anonymized data may be used for further analysis and research. I am asking for permission to keep the data indefinitely for providing opportunities for the researcher and also other interested researchers to use this same data in their research studies. If you agree, the anonymized data will be stored with the Research Data York service, which is a professional University of York data management service.

If you do not want your data to be included in any information shared as a result of this research, please do not sign this consent form. You are free to withdraw from the study at any time during data collection and up to 2 weeks after the data is collected by sending an email to Sara Ebrahimi (Sara.ebrahimi@york.ac.uk).

Information about confidentiality

The data that I collect (task responses, transcripts from the speaking and vocabulary tests, and the audio recordings) may be used in *anonymous* format in different ways. Please indicate on the consent form attached with a ☑ if you are happy for this anonymized data to be used in the ways listed.

I hope that you will agree to take part. If you have any questions about the study that you would like to ask before giving consent or after the data collection, please feel free to contact Sara Ebrahimi by email (Sara.ebrahimi@york.ac.uk), Dr Zoe Handley as the research supervisor (Zoe.handley@york.ac.uk) or the Chair of Ethics Committee via email education-research-administrator@york.ac.uk
If you are happy to participate, please complete the form attached.

Please keep this information sheet for your own records.

Thank you for taking the time to read this information.

Yours sincerely

Sara Ebrahimi

The Impact of vocabulary training on EFL learners' language proficiency

Consent Form

I confirm that I have read and understood the information given to me about the above-named research project and I understand that this will involve me taking part as described above.

I understand that the purpose of the research is finding out the effect of vocabulary training on EFL learners' language proficiency.

I understand that data will be stored anonymously, and only Sara Ebrahimi will have access to any identifiable data. I understand that my identity will be protected by use of a code.

I understand that I may be assigned into different groups of the research and the process of assignment is totally random and by chance.

I understand that my data will not be identifiable, and the anonymous data may be used…

in publications that are mainly read by university academics

in presentations that are mainly attended by university academics

in publications that are mainly read by the public

in presentations that are mainly attended by the public

I understand that the data will only be identifiable for the period of data collection, which is approximately 3 (three) months, and up to 2 weeks after this point and will be replaced by codes afterwards ☐

☐

I understand that the identifiable data will be destroyed within two weeks after the end of data collection period of time ☐

I understand that the anonymized data will be securely stored indefinitely and made available to other interested researchers for use in their studies.

I understand that anonymous data could be used for future analysis or other research purposes. ☐

☐

I agree to voluntarily participate in this study.

I understand that I can withdraw my data at any point during data collection and up to two weeks after data is collected by sending an email to Sara Ebrahimi (Sara.ebrahimi@york.ac.uk) ☐

Name ________________________

Signature ________________________ Date ______________

Appendix L - List of all words selected from student textbooks

Sports					
Success	Champion	Blow	Catch	Dangerous	Knock out
Lose	Bulk up	Practice	Score	In shape	Throw
Coach	Fan	Injury	Prize	Race	Challenging
Compete	Cycle	Exciting	Hurt	Popular	Ride
Captain	Give up	Referee	Whistle	Diving	Cheat
Bruise	Medal	Opponent	Amateur	Speed	Exhausted
Spectator	Field	Win	Team	Track	Work out
Arena	Weightlifting	Slope	Tie	parasailing	Beat
Join up	Train	Bounce	Rule	Supporter	Applaud
Award	Sweat	sculling	Warm up	Slow	Fall

Traveling					
Flight	Aisle	Airline	Fascinating	Cabin crew	Luggage
Get off	Inn	Illegal	Set out	Boring	Take off
Turbulence	Drop off	Put on	Seatbelt	Stopover	Survive
Land	Pack	Expensive	Reasonable	Special	Wonderful
Enjoy	Rent	Reserve	Sail	Camp	Countryside
Forest	Sightseeing	Comfortable	Crowded	Disappointing	Peaceful
Adventure	Crash	Customs	Jet lag	Lake	Passenger
Pond	Tent	Ancient	Staff	Wild	Explore
Depart	Palace	Suitcase	Magnificent	Spectacular	Jungle
Fasten	Announcement	Cozy	Humble	Transit	Relaxing

Appendix M - Yes/No receptive test of vocabulary

Name:	Date:

Read through the lists of words carefully.
For each of the words;

- if you know what it means, place a cross (x) beside the word.
- if you don't know what it means, or if you are not sure, leave the box blank.

Do not make guesses. Mark the ones you are sure of the meaning since some of them are not even English words.

Comfortable		Compete	
Whistle		Skold	
Challenging		Champion	
Jamped		Score	
Suitcase		Diving	
Cycle		Micked	
Amateur		Reserve	
Yonze		Blow	
Land		Dangerous	
Slow		Race	
Captain		Lake	
Disappointing		Stypped	
Prize		Crash	
Referee		Shrist	
Vapse		Pack	
Magnificent		Camp	
Speed		Enjoy	
Depart		Success	
Zix		Ancient	
Fan		Jungle	
Opponent		Twans	
Plooled		Rule	
Wumps		Wild	
Slast		Practice	
Plood		Crowded	
Soys		Special	
Peaceful		Rent	

Yeud		Forest	
Bruise		Wonderful	
Plemp		Customs	
Reasonable		Boring	
Spectator		Relaxing	
Inn		Fascinating	
Fasten		Fall	
Coach		Brants	
Win		Maffed	
Teps		Sail	
Exhausted		Exciting	
Explore		Throw	
Countryside		Expensive	
Humble		Fowd	
Transit		Greaned	
Injury		Hurt	
Cozy		Medal	
Palace		Rawned	
Tent		Pond	
Aisle		Adventure	
Yocs		Ride	
Passenger		Catch	
Popular		Cheat	
Bloys		Spectacular	

Appendix N - List of cue words for productive word association tests

Pre-test				Immediate post-test				Delayed post-test			
Cue word	Freq.	Class	Ass.	Cue word	Freq.	Class	Ass.	Cue word	Freq.	Class	Ass.
Exhausted	K-2	Adj	3	Expensive	K-1	Adj	13	Special	K-1	Adj	20
Palace	K-3	Noun	14	Cheat	K-2	Verb	18	Reserve	K-2	Verb	16
Enjoy	K-1	Verb	14	Popular	K-2	Adj	22	Rent	K-1	Verb	15
Win	K-1	Verb	7	Jungle	K-4	Noun	21	Pack	K-1	Verb	23
Practice	K-2	Verb	18	Dangerous	K-1	Adj	22	Exciting	K-1	Adj	15
Cycle	K-3	Verb	17	Sail	K-1	Verb	7	Ancient	K-3	Adj	12
Lake	K-1	Noun	15	Prize	K-3	Noun	14	Tent	K-2	Noun	15
Depart	K-4	Verb	10	Amateur	K-4	Adj	15	Success	K-2	Verb	13
Slow	K-1	Adj	11	Land	K-1	Verb	25	Passenger	K-3	Noun	10
Wild	K-1	Adj	14	Hurt	K-1	Verb	12	Race	K-1	Noun	17

Note: Freq: level of frequency
Ass: Number of associations in the database
Class: Grammatical word class

Pre-test	Immediate post-test	Delayed post-test
Exhausted	Expensive	Special
Palace	Cheat	Reserve
Enjoy	Popular	Rent
Win	Jungle	Pack
Practice	Dangerous	Exciting
Cycle	Sail	Ancient
Lake	Prize	Tent
Depart	Amateur	Success
Slow	Land	Passenger
Wild	Hurt	Race

Appendix O - Choice of the corpora

In order to check the frequency of the responses to the cue words in word association task, decision was required to be made regarding which of the pre-existing corpora should be used. As the main focus of the study was the frequency of the spoken words rather than the written ones, the best choices were Spoken *British National Corpus (BNC)* and *Corpus of Contemporary American English (COCA),* as both of them contain large collection of spoken components. Although there are quite different number of corpora representing spoken English rather than mixing it with the written form, it was decided to use one of the following corpora due to two reasons. First and foremost, the spoken corpora, such as *London-Lund Corpus*, *Michigan Corpus of Academic Spoken English*, etc., are not large enough in comparison to the spoken part of these two corpora, which helped to lead to more representative samples. Secondly, some corpora such *Cambridge and Nottingham Corpus of Discourse in English,* is not available to all staff and researchers.

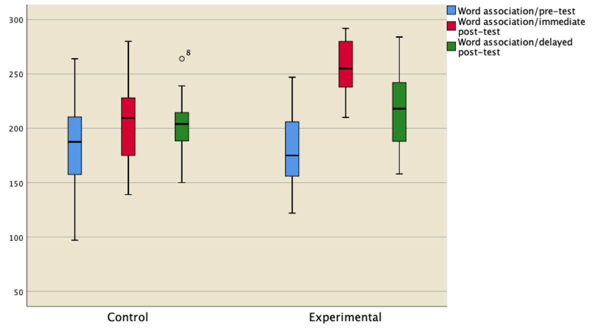
Spoken British National Corpus 2014 (Spoken BNC2014) contains 11,422,617 words of transcribed content, which was collected from 668 speakers in 1251 recordings. This corpus made use of PPSR (public participation in scientific research), which is considered to be the most innovative feature of this corpus. In comparison to the spoken component of the original BNC 1994, which had a spoken component of 10 million words, comprised of two parts; demographic (40%) and content governed (60%), the new BNC 2014 asked anyone interested to contribute by recording data. Hence, the spoken data collected was more naturalistic and involve spontaneous conversational interactions.

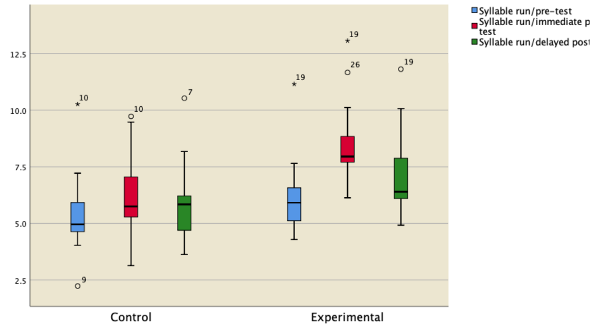
The second corpus, *Corpus of Contemporary American English (COCA),* presents the American variety of English language. It is a comprehensive, large, and updated corpus, including 385 million words. The spoken component of the *COCA*, containing 79 million words, is approximately eight times larger than the Spoken *BNC*, but it only includes the transcriptions of unscripted conversations from radio and TV programs, which minority group of people are regularly engaged in their daily lives. Although *COCA* contains larger number of spoken words, *BNC* was decided to be consulted for checking the frequency of the cue words, as its spoken sample was a better representation of informal and spontaneous speech, which was the main focus of this research. Accordingly, for the purpose of this study, it was decided to make use of Spoken BNC2014.

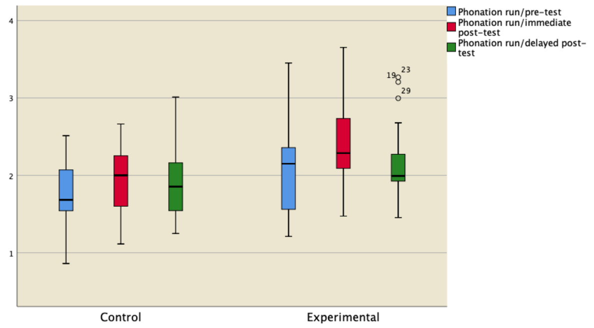
Appendix P - Test of Normality

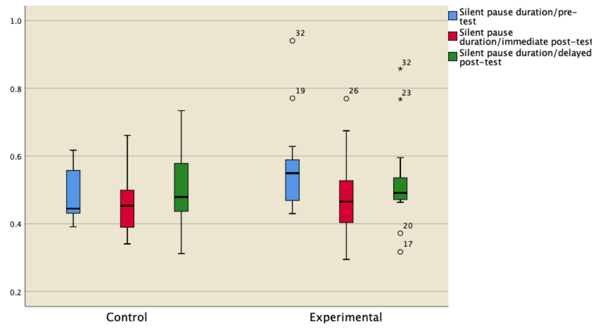
	Group	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
		Statistic	df	Sig.	Statistic	df	Sig.
Word association/Pre-test	Control	.14	16	.20*	.98	16	.96
	Experimental	.12	17	.20*	.94	17	.36
Word association/Immediate post-test	Control	.15	16	.20*	.96	16	.80
	Experimental	.11	17	.20*	.94	17	.36
Word association/Delayed post-test	Control	.11	16	.20*	.97	16	.88
	Experimental	.13	17	.20*	.94	17	.37
Syllable run/ Pre-test	Control	.22	16	.03	.86	16	.02
	Experimental	.18	17	.10	.82	17	.00
Syllable run/Immediate post-test	Control	.25	16	.00	.91	16	.12
	Experimental	.17	17	.19	.88	17	.03
Syllable run/Delayed post-test	Control	.23	16	.01	.90	16	.09
	Experimental	.19	17	.10	.89	17	.05
Phonation run/ Pre-test	Control	.16	16	.20*	.96	16	.67
	Experimental	.15	17	.20*	.91	17	.10
Phonation run/Immediate post-test	Control	.10	16	.20*	.97	16	.96
	Experimental	.17	17	.20*	.96	17	.69
Phonation run/Delayed post-test	Control	.14	16	.20*	.92	16	.16
	Experimental	.24	17	.00	.87	17	.02
Syllable duration/ Pre-test	Control	.21	16	.04	.88	16	.05
	Experimental	.11	17	.20*	.97	17	.81
Syllable duration/Immediate post-test	Control	.11	16	.20*	.95	16	.51
	Experimental	.10	17	.20*	.98	17	.99
Syllable duration/Delayed post-test	Control	.13	16	.20*	.94	16	.46
	Experimental	.15	17	.20*	.95	17	.45
Silent pause duration/Pre-test	Control	.23	16	.01	.84	16	.01
	Experimental	.18	17	.13	.80	17	.00
Silent pause duration/Immediate post-test	Control	.15	16	.20*	.93	16	.26
	Experimental	.19	17	.10	.93	17	.21
Silent pause duration/Delayed post-test	Control	.12	16	.20*	.97	16	.93
	Experimental	.27	17	.00	.81	17	.00

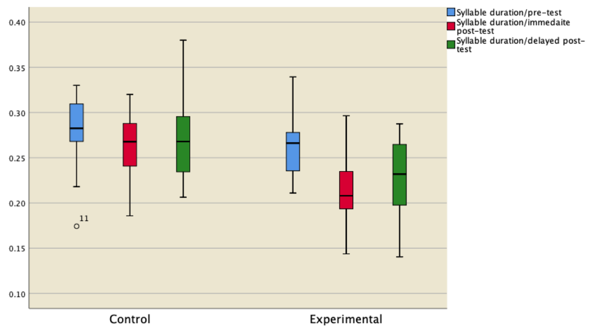
Appendix Q - Boxplots











Appendix R - Effect size

Effect size can be called "the most important parameter of all" (Murphy, Mayors, & Wolach, 2014, p. 162). It is a statistical measure showing the magnitude impact of the two variables on one another in numeric way. The most comprehensive definition of this concept can be "a quantitative reflection of the magnitude of some phenomenon that is used for the purpose of addressing a question of interest" (Kelley & Preacher, 2012, p. 140).

Accordingly, it measures the differences between two independent sample means, in order to express how large or small the differences are. The defaulted guideline in interpreting the effect size has been put forward by Cohen (1988), setting d = .2, r = .1 (small); .5, .3 (medium); and .8, .5 (large) benchmark, while empirical evidence confirm that this scale's underestimation of the variety and range of impacts more commonly obtained in L2 research (Plonsky & Oswald, 2014). Accordingly, more recent estimates of small, medium, and large, setting the *d* value in the approximation of 0.40, 0.70, and 1.00 respectively, were considered in the current study.

Appendix S – Mean number of responses in different categories of word association test for control and experimental groups

		Minimum			Maximum			Mean		
Category	**Group**	**Pre-test**	**Immediate Post-test**	**Delayed Post-test**	**Pre-test**	**Immediate Post-test**	**Delayed Post-test**	**Pre-test**	**Immediate Post-test**	**Delayed Post-test**
Unassociated responses	Control	3	3	4	21	16	17	11.18	8.78	8.81
	Experimental	6	0	5	23	7	19	12.47	1.64	10.05
Loosely associated responses	Control	1	1	0	11	8	12	4.56	3.75	3.81
	Experimental	2	0	1	9	7	11	5.58	3.52	5.00
Closely associated responses	Control	17	19	18	35	48	44	27.06	30.06	29.62
	Experimental	11	24	15	42	53	48	30.52	38.70	33.76
Canonical responses	Control	5	7	6	15	19	20	11.06	12.67	12.50
	Experimental	4	9	7	19	26	21	12.00	15.32	14.00
Noncanonical responses	Control	8	6	9	23	32	25	15.50	17.18	17.25
	Experimental	4	9	7	27	38	27	14.58	17.05	15.64
High frequency responses	Control	7	12	16	26	39	36	19.00	22.18	22.67
	Experimental	8	16	9	29	42	32	21.35	29.47	22.17
Low frequency responses	Control	2	2	1	15	16	12	7.43	7.50	6.87
	Experimental	3	3	3	16	15	20	8.64	9.11	11.29
Position based responses	Control	2	2	2	13	16	10	7.37	7.25	5.56
	Experimental	2	3	6	15	18	21	8.41	9.76	10.29
Semantically related responses	Control	4	4	9	19	22	24	12.68	14.18	15.00
	Experimental	5	10	5	24	25	20	12.76	17.41	12.76
Position/semantically related responses	Control	1	5	2	15	14	17	6.68	8.62	9.06
	Experimental	4	5	3	14	18	21	9.35	11.52	10.70

Appendix T - Dealing with the outliers

In the case of syllable run, firstly, the normality of the distribution of the data was assessed without the inclusion of the residuals and it was found out that the data does not still conform the normality of distribution assumptions. Hence, series of Wilcoxon and Mann-Whitney U tests were conducted to check if the training had any significant impact on the measure of the syllable run. A Wilcoxon signed rank test indicated that the training had a significantly great impact on experimental group at immediate post-test, $T = 0$, $z = -3.51$, $p = .00$, two tailed. The same great impact is also reported for the experimental group from the pre-test to the delayed post-test, $T = 13$, $z = -2.84$, $p = .00$, two tailed. In the case of the control group significant changes were found both from pre-test to immediate post-test, $T = 17$, $z = -2.42$, $p = .01$, two tailed, and from pre-test to delayed post-test phase, $T = 25$, $z = -1.98$, $p = .05$, two tailed. In the second phase of the analysis, comparing the performance of the groups, a Mann-Whitney *U* test indicated that the experimental group (*Mean Rank* = 21.19) was significantly higher than those of the control group (*Mean Rank* = 10.47) in pre-test to immediate post-test gains, $U = 37.00$, $z = -3.28$, $p = .00$, two-tailed. In addition to that, the pre to delayed post-test gains also revealed significant differences in the gains of the experimental (*Mean Rank* = 19.31) and control group (*Mean Rank* = 12.47), $U = 67.00$, $z = -2.09$, $p = .03$, two-tailed, on the direction of the experimental group.

Next series of Mann-Whitney *U* and Wilcoxon tests were conducted in order to check the measure of silent pause duration as the data was not normally distributed. The results were in line with the previously reported results in the Analysis chapter in chapter 5 (section 5.3.6), with the outliers included. The results of the Wilcoxon signed rank test on experimental group measure of silent pause duration indicates no significant difference in the phases of pre-test to immediate post-test, $T = 30$, $z = -1.70$, $p = .08$, also no significant difference was reported from pre-test to delayed post-test, $T = 31$, $z = -1.64$, $p = .10$, two tailed. The same trend could have been found in the control group from pre-test to immediate post-test, $T = 61$, $z = -.36$, $p = .71$, and pre-test to delayed post-test, $T = 56$, $z = -.621$, $p = .53$, two tailed.

The Mann-Whitney *U* results suggests that there was no significant difference on pre-immediate post gains in gaining shorter silent pause duration between the control and experimental groups, $U = 86.00$, $z = -1.34$, $p = .17$, and also so significance difference was reported on pre-test to delayed post-test phase between the two groups, U =97.00, z = -.90, p = .36, two-tailed. The results found in this section with the exclusion of the residuals were same as the results in the analysis chapter (chapter 5).

Appendix U – Correlation analysis of syllable run and phonation run in control and experimental groups

Group	Measures		Phonation run immediate	Phonation run delayed
Control	Syllable run immediate	Correlation Coefficient	.75**	
		Sig. (2-tailed)	.00	
		N	16	
	Syllable run delayed	Correlation Coefficient		.60*
		Sig. (2-tailed)		.01
		N		16
Experimental	Syllable run immediate	Correlation Coefficient	.85**	
		Sig. (2-tailed)	.00	
		N	17	
	Syllable run delayed	Correlation Coefficient		.82**
		Sig. (2-tailed)		.00
		N		17

Appendix V - Results of Mann-Whitney U test between experimental and control groups on pre-immediate post-test gains and pre-delayed post-test gains

Fluency measure	**Group**	**Test Comparison**	***U***	***z***	***p(two-tailed)***
Syllable run	Experimental vs. Control	Pre-test to immediate post-test gain	53.00	-2.99	.00*
	Experimental vs. Control	Pre-test to delayed post-test gain	82.00	-1.94	.05*
Phonation run	Experimental vs. Control	Pre-test to immediate post-test gain	77.00	-2.12	.03*
	Experimental vs. Control	Pre-test to delayed post-test gain	91.00	-1.62	.11
Syllable duration	Experimental vs. Control	Pre-test to immediate post-test gain	69.00	-2.41	.01*
	Experimental vs. Control	Pre-test to delayed post-test gain	80.50	-1.99	.04*
Silent pause duration	Experimental vs. Control	Pre-test to immediate post-test gain	86.00	-1.80	.07
	Experimental vs. Control	Pre-test to delayed post-test gain	97.00	-1.40	.16

Appendix W - Results of Wilcoxon signed rank test for different test time of pre-test, immediate post-test, delayed post-test within control and experimental groups

Group	Fluency measure	Testing times	*T*	*z*	*p*(two-tailed)
Experimental		Pre-test to immediate post-test	0	-3.62	.00
	Syllable run	Pre-test to delayed post-test	13	-3.00	.00
Control		Pre-test to immediate post-test	23	-2.32	.02
		Pre-test to delayed post-test	41	-1.39	. 16
Experimental		Pre-test to immediate post-test	10	-3.14	.00
	Phonation run	Pre-test to delayed post-test	68	-0.40	.68
Control		Pre-test to immediate post-test	31	-1.91	.06
		Pre-test to delayed post-test	33	-1.81	.07
Experimental		Pre-test to immediate post-test	0	-3.62	.00
	Syllable	Pre-test to delayed post-test	4.50	-3.19	.00
Control	duration	Pre-test to immediate post-test	6	-1.96	.06
		Pre-test to delayed post-test	7	-.62	.53
Experimental		Pre-test to immediate post-test	40	-1.72	.08
	Silent pause	Pre-test to delayed post-test	46	-1.44	.14
Control	duration	Pre-test to immediate post-test	61	-.36	.71
		Pre-test to delayed post-test	56	-.62	.53

References

Adolph, S. & Schmitt, N. (2003). Lexical coverage of spoken discourse. *Applied Linguistics, 24*(4), 425, 438.

Aghagolzadeh, F., & Davari, H. (2014). Iranian critical ELT: A belated but growing intellectual shift in Iranian ELT community. *Journal for Critical Education Policy Studies*, *14*(1), 391-410.

Ahmadian, M. J., & Tavakoli, M. (2011). The effects of simultaneous use of careful online planning and task repetition on accuracy, complexity, and fluency in EFL learners' oral production. *Language Teaching Research*, *15*(1), 35–59.

Ahmadian, M., Tavakoli, M., & Vercellotti, H. (2015). The combined effects of online planning and task structure on complexity, accuracy and fluency of L2 speech. *Language Learning Journal*, *43*(1), 41–56.

Aitchison, J. (2012). *Words in the mind*. Oxford: Wiley-Blackwell.

Al-Baili, M. A. (1988). *The effect of semantic mapping strategy on prose learning across time and ability*. Doctoral dissertation, Univesrity of Wisconsin, Madison, WI.

Anderson, J. R. (1983). *The architecture of cognition*. Lawrence Erlbaum Associates, Inc.

Anderson, J. R. (1995). *Learning and memory. An integrated approach*. John Wiley & Sons.

Anderson, J. R., & Lebiere, C. (1998). *The atomic components of thought*. Lawrence Erlbaum Associates Publishers.

Anderson, J. R. (1983). A spreading activation theory of memory. *Journal of Verbal Learning and Verbal Behavior*, *22*, 261–295.

Anderson, R. C., & Freebody, P. (1981). Vocabulary knowledge. In J. Guthrie (Ed.), *Comprehension and Teaching: Research Reviews* (pp. 77–117). Newark, DE: International Reading Association.

Babaii, E., & Ansary, H. (2001). The C-test a valid operationalization of reduced redundancy principle? *System*, *29*, 209–219.

Bachman, L. F. . (1985). Performance on cloze tests with fixed-ratio and rational deletions. *TESOL Quarterly*, *16*, 61–70.

Baddeley, A. D. (1990). *Human memory*. London: Lawrence Erlbaum Associates.

Baddeley, A. D. (1999). *Essentials of human memory*. Hove, UK: Psychology Press.

Badr, H. M., & Abu-Ayyash, E. A. S. (2019). Semantic mapping or rote memorisation: which strategy is more effective for students' acquisition and memorization of L2 vocabulary? *Journal of Education and Learning*, *8*(3), 158-174.

Baker-Smemoe, W., Dewey, D. P., Bown, J., & Martinsen, R. A. (2014). Does measuring L2 utterance fluency equal measuring overall L2 proficiency? Evidence from five languages. *Foreign Language Annals*, *47*(4), 707–728.

Balota, D. A. (1994). Visual word recognition: The journey from features to meaning. In M. A. Gernsbacher (Ed.), *Handbook of psycholinguistics* (pp. 303–348). Academic Press.

Barcroft, J. (2015). Can retrieval opportunities increase vocabulary learning during reading? *Foreign Lnaguage Annals, 48*(2), 236-249.

Bauer, L., & Nation, P. (1993). Word families. *International Journal of Lexicography, 6*(4). 253-279.

Blachowicz, C. L. Z., Fisher, P. J. L., Ogle, D., & Watts-taffe, S. (2006). Vocabulary: Questions from the classroom. *Reading Research Quarterly*, *41*(4), 524–539.

Bloem, I., Van den Boogaard, S., & La Heij, W. (2004). Semantic facilitation and semantic interference in language production: Further evidence for the conceptual selection model of lexical access. *Journal of Memory and Language*, *51*, 307–323.

Bock, K., & Levelt, W. (1994). Language production: Grammatical encoding. In M. A. Gernsbacher (Ed.), *Handbook of Psycholinguistics* (pp. 945–984). San Diego, CA,: Academic Press.

Boers, F., & Lindstromberg, S. (2008). How cognitive linguistics can foster effective vocabulary teaching. In F. Boers and S. Lindstromberg (Ed.), *Cognitive linguistic approaches to teaching vocabulary and phraseology* (pp. 1-65). De Gruyter Mouton.

Boersma, P., & Weenink, D. (2018). *Praat: doing phonetics by computer*. Retrieved from http://www.praat.org/.

Bosker, H. R., Pinget, A. F., Quené, H., Sanders, T., & De Jong, N. H. (2012). What makes speech sound fluent? The contributions of pauses, speed and repairs. *Language Testing*, *30*(2), 159–175.

Bosker, H. R., Quené, H., Sanders, T., & De Jong, N. H. (2014). The perception of fluency in native and nonnative speech. *Language Learning*, *64*(3), 579–614.

Boyle, M., & Coelho, C. A. (1995). Application of semantic feature analysis as a treatment for aphasic dysnomia. *American Journal of Speech-Language Pathology*, *4*(4), 94-98.

Brown, H.D. (1994). *Teaching by principles: an interactive approach to language pedagogy*. Englewood Cliffs, NJ: Prentice Hall Regents.

Bui, G., & Huang, Z. (2018). L2 fluency as influenced by content familiarity and planning: Performance, measurement, and pedagogy. *Language Teaching Research*, *22*(1), 94-114.

Burns, A. & Joyce, H. (1997*). Focus on speaking.* Sydney: National Centre for English Language

Teaching and Research.

Bygate, M. (2001). Effects of task repetition on the structure and control of orallanguage. In M. Bygate, P. Skehan, & M. Swain (Eds.), *Researching pedagogic tasks* (pp. 23–48). London, UK: Longman.

Carling, K. (2000). Resistant outlier rules and the non-Gaussian case. *Computational Statistics and Data Analysis*, *33*(3), 249–258.

Catalán, J. R., & Fitzpatrick, T. (2013). Frequency profiles of EFL learners' lexical availability. In I. Catalán (Ed.), *Lexical availability in English and Spanish as a second language* (pp. 83-100). Springer Nature.

Cenoz, J. (2000). Pauses and hesitation phenomena in second language production. *ITL - International Journal of Applied Linguistics*, *127*, 53–69.

Chambers, F. (1997). *What do we mean by fluency? 25*(4), 535–544.

Chaudron, C., Prior, M., & Kozok, U. (2005). Elicited imitation as an oral proficiency measure. *Paper presented at the 14th World Congress of Applied Linguistics,* Madison, WI

Clark, H. H., & Wasow, T. (1998). Repeating words in spontaneous speech. *Cognitive Psychology*, *37*(3), 201–242.

Clenton, J., De Jong, N. H., Clingwall, D., & Fraser, S. (2020). Investigating the extent to which vocabulary knowledge and skills can predict aspects of fluency for a small group of pre-intermediate Japanese L1 users of English (L2). In J. Clenton & P. Booth (Eds.), *Vocabulary and the four skills; pedagogy, practice, and implications for teaching vocabulary* (pp. 126–145). London, Routledge.

Cohen, A. D., & Aohek, E. (1981). Easifying second language learning. *Studies in Second Language Acquisition*, *3*(2), 221-236.

Cohen, J. D. (1988). *Statistical power analysis for the behavioral sciences*. Hillsdale, N.J. : L. Erlbaum Associates.

Cohen, J. D., Dunbar, K., & McClelland, J. L. (1990). On the control of automatic processes: A parallel distributed processing account of the Stroop effect. *Psychological Review, 97*, 332-361.

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

Collins, A. M., & Loftus, E. F. (1975). A spreading-activation theory of semantic processing. *Psychological Review, 82*, 407–28.

Costa, A. (2005). Lexical access in bilingual production. In J. Kroll and A.M.B. De Groot (Eds.) *Handbook of bilingualism: Psycholinguistics* approaches (pp. 308-325). Oxford: Oxford University Press.

Craik, F. I. M., & Lockhart, R. S. (1972). Levels of processing: A framework for memory research. *Journal of Verbal Learning and Verbal Behavior*, *11*, 671–684.

Craik, F., & Tulving, E. (1975). Depth of processing and the retention of words in episodic memory. *Journal of Experimental Psychology, 104*, 268–284.

Cramer, P. (1968). *Word association*. New York, NY: Academic Press.

Cremer, M., D., Dingshoff, M., de Beer, M., & Schoonen, R. (2011). Do word associations assess word knowledge? A comparison of L1 and L2, child and adult word associations. *International Journal of Bilingualism*, *15*(2), 187–204.

Cucchiarini, C., Strik, H., & Boves, L. (2002). Quantitative assessment of second language learners' fluency: Comparisons between read and spontaneous speech. *The Journal of the Acoustical Society of America*, *111*(6), 2862–2873.

Cucchiarini, C., Strik, H., Boves, L., (2000). Quantitative assessment of second language learners ' fluency by means of automatic speech recognition technology quantitative assessment of second language learners' fluency by means of automatic speech recognition technology. *The Journal of the Acoustical Society of America*, *107*(2), 989–999.

Daller, H., & Phelan, D. (2006). The C-test and TOEIC as measures of students' progress in intensive short courses in EFL; The test of English for international communication. *Der C-Test: Theorie, Emprie, Anwendungen/The C-Test: Theory, Empirical Research, Applications*, *6*, 101–119.

Daller, H., & Phelan, D. (2006). The C-test and TOEIC as measures of students' progress in intensive short courses in EFL. In R. Grotjahn (Ed.), *Der C-Test: Theorie, Empirie, Anwendungen. The C-Test: Theory, Empirical Research, Applications*. *Volume 6* (pp. 101-119). Peter Lang

Dang, T. N. Y. (2017). *Investigating vocabulary in academic spoken English: Corpora, teachers, and learners*. Victoria University, Wellington, New Zealand.

Dang, T. N. Y., & Webb, S. (2016). Evaluating lists of high-frequency words. *ITL. International Journal of Applied Linguistics*, *167*(2), 132–158.

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

De Bot, K & Lowie, W. (2010). Stability of representation in the multilingual lexicon. In M. Putz & L. Sicola (Ed.), *Cognitive Processing in Second Language Acquisition.* (pp. 117-134). Amsterdam: John Benjamins.

De Bot, K. (2004). The multilingual lexicon: Modelling selection and control. *International Journal of Multilingualism*, *1*(1), 17–32.

Deese, J. (1965). The structure of associations in language and thought. *Journal of Linguistics, 4*(1), 131-133.

De Deyne, S., Leopold, S., Navarro, D. J., Perfors, A., & Storms, G. (2012). Strong structure in weak semantic similarity: a graph based account. *Building Bridges Across Cognitive Sciences Around the World : Proceedings of the 34th Annual Meeting of the Cognitive Sciene Society* (pp. 1464–1469). USA; Cognitive Science Society.

De Deyne, S., & Storms, G. (2008). Word associations: network and semantic properties. *Behavior Research Methods*, *40*(1), 213–231.

De Deyne, S., & Storms, G. (2015). Word Associations. In J. R. Taylor (Ed.), *The Oxford Handbook of the Word* (pp. 465–480). Oxford: Oxford University Press.

De Deyne, Simon, Kenett, Y. N., Anaki, D., Faust, M., & Navarro, D. (2016). Large-scale network representations of semantics in the mental lexicon. In M. N. Jones (Ed.), *Big Data in Cognitive Science* (p. 174–202), Psychology Press: Taylor & Francis.

De Groot, A. M. B. (1989). Representational aspects of word imageability and word frequency as assessed through word association. *Ournal of Experimental Psychology: Learning, Memory, and Cognition*, *15*(5), 824–845.

De Groot, A. M. B. (2000). On the source and nature of semantic and conceptual knowledge. *Bilingualism: Language and Cognition*, *3*, 7–9.

De Jong, N. H. (2016). Fluency in second langauge assessment. In D. Tsagari & J. Banerjee (Eds.), *Handbook of Second Language Assessment* (pp. 203–218). Mouton de Gruyter.

De Jong, N. H. (2018). Fluency in second language testing: Insights from different disciplines. *Language Assessment Quarterly*, *15*(3), 237–254.

De Jong, N. H., & Bosker, H. R. (2013). Choosing a threshold for silent pauses to measure second language fluency. *Proceedings of the 6th Workshop on Disfluency in Spontaneous Speech*, 17–20.

De Jong, N. H., Groenhout, R., Schoonen, R., & Hulstijn, J. H. (2015). Second language fluency: Speaking style or proficiency? Correcting measures of second language fluency for first language behavior. *Applied Psycholinguistics*, *36*(2), 223–243.

De Jong, N. H., & Mora, J. C. (2019). Does having good articulatory skills lead to more fluent speech in first and second languages? *Studies in Second Language Acquisition*, *41*(1), 227–239.

De Jong, N. H., Schoonen, R., & Hulstijn, J. H. (2009). Fluency in L2 is related to fluency in L1. In *Paper presented at the Seventh International Symposium on Bilingualism (ISB7)*. Utrecht, Netherlands.

De Jong, N. H., Steinel, M. P., Florijn, A. F., Schoonen, R., & Hulstijn, J. H. (2012). Facets of speaking proficiency. *Studies in Second Language Acquisition*, *34*(1), 5–34.

De Jong, N. H., Steinel, M. P., Florijn, A., Schoonen, R., & Hulstijn, J. H. (2013). Linguistic skills and speaking fluency in a second language. *Applied Psycholinguistics*, *34*(5), 893–916.

De Jong, N. H., & Perfetti, C. A. (2011). Fluency training in the ESL classroom: An experimental study of fluency development and proceduralization. *Language Learning*, *61*(2), 533–568.

De Jong, N. H., & Vercellotti, M. Lou. (2015). Similar prompts may not be similar in the performance they elicit: Examining fluency, complexity, accuracy, and lexis in narratives from five picture prompts. *Language Teaching Research*, *20*(3), 387–404.

DeJong, N. H., & Wempe, T. (2009). Praat script to detect syllable nuclei and measure speech rate automatically. *Behavior Research Methods*, *41*(2), 385–390.

Deese, J. (1962). Form class and the determinants of association. *Journal of Verbal Learning and Verbal Behavior*, *1*, 79–84.

DeKeyser, R. (2001). Automaticity and automatization. In P. Robinson (Ed.), *Cognition and second language instruction* (pp. 125–151). New York, NY: Cambridge University Press.

DeKeyser, R. (2007). *Practicing in a second language: Perspectives from applied linguistics and cognitive psychology* . Cambridge: Cambridge University Press.

DeKeyser, R. (2017). Knowledge and Skill in ISLA. In S. Loewen & M. Sato (Eds.), *The Routledge Handbook of Instructed Second Language Acquisition* (pp. 15–33). Routledge.

Dell, G. S. (1986). A spreading activation theory of retrieval in sentence production. *Psychological Review*, *93,* 283–321.

Derwing, T. M. (2017). L2 fluency development. In S. Loewen & M. Sato (Eds.), *The Routledge Handbook of Instructed Second Language Acquisition* (pp. 246–259). Routledge.

Derwing, T. M., Munro, M. J., Thomson, R. I., & Rossiter, M. J. (2009). The relationship between L1 fluency and L2 fluency development. *Studies in Second Language Acquisition*, *31*(4), 533–557.

Derwing, T. M., Rossiter, M. J., Munro, M. J., & Thomson, R. I. (2004). Second language fluency: Judgments on different tasks. *Language Learning*, *54*(4), 655–679.

Dóczi, B., & Kormos, J. (2016). *Longitudinal developments in vocabulary knowledge and lexical organization*. Oxford, Oxford University Press.

Dufour, R. & Kroll, J. (1995) Matching words to concepts in two languages: A test of the concept mediation model of bilingual representation. *Memory & Cognition, 23*, 166-180.

Ellis, N. C. (2005). At the interface: Dynamic interactions of explicit and implicit language knowledge. *Studies in Second Language Acquisition*, *27*, 305–352.

Ellis, N. (2007). Implicit and explicit knowledge about language. In J. Cenoz & N. H. Hornberger (Eds.), *Encyclopedia of language and education*, *6*, 119–132. New York: Springer.

Ellis, N. C. (2015). Implicit and explicit language learning: Their dynamic interface and complexity. In P. Rebuschat (Ed.), *Implicit and explicit learning of languages*, (pp. 3–23). Philadelphia, PA: John

Benjamins.

Ellis, R. (1987). Interlanguage variability in narrative discourse: Style shifting in the use of the past tense. *Studies in Second Language Acquisition*, *9*(1), 1-19.

Ellis, R. (2005). Measuring implicit and explicit knowledge of a second language: A psychometric study. *Studies in Second Language Acquisition*, *27* , 141–172.

Ellis, R. (2006). Modelling learning difficulty and second language proficiency: The differential contributions of implicit and explicit knowledge, *Applied Linguistics*, 27(3), 431–463,

Ellis, R. (2009). The differential effects of three types of task planning on the fluency, complexity, and accuracy in L2 oral production. *Applied Linguistics*, *30*(4), 474–509.

Entwisle, D. R. (1966). *Word associations of young children*. Baltimore: John Hopkins University.

Farhady, H., Hezaveh, F. S., & Hedayati, H. (2010). The electronic journal for English as a second language. *The Electronic Journal for English as a Second Language*, *13*(4), 1–12.

Fay, D. & Culter, A. (1977). Malapropisms and the structure of the mental lexicon. *Linguistic Inquiry*, *8*(3), 505-520.

Field, A. (2009). *Discovering statistics using SPSS (Third ed.)*. London: Sage.

Fitzpatrick, T. (2009). Word association profiles in a first and second language: Puzzles and problems. In T. Fitzpatrick & A. Barfield (Eds.), *Lexical Processing in Second Language Learners* (pp. 38–52), Multilingual Matters.

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

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

Foroozandeh, E. & Forouzani, M. (2015). Developing school English materials for the new Iranian educational system. In C. Kennedy (Ed.), English language teaching in the Islamic Republic of Iran: Innovations, trends and challenges (pp. 49–59). London: British Council.

Tess, F., & Fitzpatrick, T. (2012). Word associations. In C. Chapelle (Ed.), *Encyclopedia of Applied Linguistics*. Oxford, John Wiley and Sons, Inc.

Fitzpatrick, T. & Izura, C. (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. & Munby, I. (2013). *Knowledge of word associations.* Milton, J. & Fitzpatrick, T. (Ed.), *Dimensions of Vocabulary Knowledge,* (pp. 92-105). Basingstoke: Palgrave Macmillan.

Fitzpatrick, T., Playfoot, D., Wray, A., & Wright, M. (2015). Establishing the reliability of word association data for investigating individual and group differences. *Applied Linguistics, 36*(1), 23-50.

Fitzpatrick, Tess, & Thwaites, P. (2020). Word association research and the L2 lexicon. *Language Teaching*, *53*(3), 237–274.

Foster, P., & Skehan, P. (1996). The influence of planning and task type on second language performance. *Studies in Second Language Acquisition*, *18*(3), 299-323.

Fraenkel, J. R., Wallen, N. E., & Hyun, H. H. (2012). *How to design and evaluate research in education.* New York: The McGraw-Hill Companies.

Francis, W. S. (2005). Bilingual semantic and conceptual representation. In J. F. Kroll & A. M. B. de Groot (Eds.), *Handbook of bilingualism: Psycholinguistic approaches* (pp. 251–267). Oxford: Oxford university press.

Francis, Wendy S. (1999). Cognitive integration of language and memory in bilinguals: Semantic representation. *Psychological Bulletin*, *125*(2), 193–222.

Freed, B. F. (2000). Is fluency, like beauty, in the eyes of the beholder? In H. Riggenbach (Ed.), *Perspectives on Fluency* (pp. 243–265). Michigan: University of Michigan Press.

Freed, B., Segalowitz, N., & Dewey, D. (2004). Context of learning and second language fluency in French: Comparing regular classroom, study abroad, and intensive domestic immersion programs. *Studies in Second Language Acquisition*, *26*, 275–301.

French, R. M., & Jacquet, M. (2004). Understanding bilingual memory: Models and data. *Trends in Cognitive Sciences*, *8*(2), 87–93.

Fulcher, G. (2003). *Testing second language speaking*. Harlow, North Dakota: Pearson Education.

Fulcher, G. (2014). *Testing Second Language Speaking* (2nd ed.). Harlow: Pearson Education.

Gairns, R. & Redman, S. (1986). *Working with words*. Cambridge: Cambridge University Press.

Galante, A., & Thomson, R. I. (2016). The effectiveness of drama as an instructional approach for the development of second language oral fluency, comprehensibility, and accentedness. *TESOL Quarterly*, *51*(1), 115–142.

García Amaya, L. (2009). New findings of fluency measures across three different learning contexts. In J. Collentine, B. A. Lafford, M. García, & F. M. Marín (Eds.), *Selected Proceedings of the 11th Hispanic Linguist* (pp. 68–80), Somerville, MA: Cascadilla Proceedings Project.

Ghonsooly ,B., & Hosienpour, A. (2009) .The effect of concept mapping on EFL speaking fluency. *IJAL, 12*(1), 87-114.

Gilabert, R. (2007). The simultaneous manipulation of task complexity along planning time and [+/-

Here-and-Now]: Effects on L2 oral production. *Investigating Tasks in Formal Language Learning*, *45*, 44–68.

Ginther, A., Dimova, S., & Yang, R. (2010). Conceptual and empirical relationships between temporal measures of fluency and oral English proficiency with implications for automated scoring. *Language Testing*, *27*(3), 379–399.

Goffman, E. (1981). Radio talk. In E. Goffman (Ed.), *Forms of talk* (pp. 197–327). Philadelphia, PA: University of Pennsylvania Press.

Goldman-Eisler, F. (1961). The significance of changes in the rate of articulation. *Language and Speech*, *4*(3), 171–174.

Goldman-Eisler, F. (1968). *Psycholinguistics: Experiments in spontaneous speech*. New York: Academic Press.

González-Fernández, B. & Schmitt, N. (2017). Vocabulary acquisition. In S. Loewen and M. Sato, *The Routledge Handbook of Instructed Second Language Acquisition* (pp. 280-298). New York and London: Routledge.

Graham, C. R., Lonsdale, D., Kennington, C., Johnson, A., & McGhee, J. (2008). Elicited imitation as an oral proficiency measure with ASR scoring. In *Proceedings of the 6th International Conference on Language Resources and Evaluation*, (pp.1604-1610), Paris, France. European Language Resources Association.

Hilton, H. (2009). Annotation and analyses of temporal aspects of spoken fluency. *CalicoJournal*, *26*(3), 644–661.

Hilton, H. (2014). Oral fluency and spoken proficiency: Considerations for research and testing. *Measuring L2 proficiency: Perspectives from SLA*, 27-53.

Hague, S. A. (1987). Vocabulary instruction: What L2 can learn from L1. *Foreign Language Annals*, *20*(3), 217-225.

Hamzah, M. S. G., Kafipour, R., & Abdullah, S. K. (2009). Vocabulary learning strategies of Iranian undergraduate EFL students and its relation to their vocabulary size. *European Journal of Social Sciences*, *11*(1), 39–50.

Handley, Z. (2017). *FLOW: Fluency in Oral interaction Workshop.* www.york.ac.uk/education/news-events/events/2017/flow/.

Hashimoto, N., & Frome, A. (2011). The use of a modified semantic features analysis approach in aphasia. *J Commun Disord. 44*(4), 459-69.

Hastings, A. J. (2002). In defense of C-testing. In R. Grotjahn (Ed.), *Der C-Test. Theoretische Grundlagen und praktische Anwendun (Vol. 4, pp. 11–29)gen*. Bochum: AKS-Verlag.

Heaton, J. B. (1966). *Composition through pictures*. Longman.

Henriksen, B. (2008). Declrative lexical knowledge. In D. Albrechtsen, K. Haastrup, & B. Henriksen (Eds.), *Vocabulary and writing in a First and second Language: Processes and development*. Palgrave Macmillan.

Henriksen, B. (1999). Three dimensions of vocabulary development. *Studies in Second Language Acquisition, 21*, 303–317.

Henriksen, B. & Haastrup, K. (1998). Vocabulary acquisition: acquiring depth of knowledge through network building. *International Journal of Applied Linguistics, 10*(2), 221-240.

Hermans, D., Bongaerts, T., DE Bot, K., & Schreuder, R. (1998). Producing words in a foreign language: Can speakers prevent interference from their first language? *Bilingualism: Language and Cognition, 1(*3), 213-229.

Hilton, H. (2008). The link between vocabulary knowledge and spoken L2 fluency. *Language Learning Journal*, *36*(2), 153–166.

Hirsh, K. W., & Tree., J. T. (2001). Word association norms for two cohorts of British adults. *Journal of Neurolinguistics*, *14*(1), 1–44.

Hoaglin, D. C., & Iglewicz, B. (1987). Fine-tuning some resistant rules for outlier labeling. *Journal of the American Statistical Association*, *82*(400), 1147–1149.

Housen, A., & Kuiken, F. (2009). Complexity, accuracy, and fluency in second language acquisition. *Applied Linguistics*, *30*(4), 461–473.

Huensch, A., & Tracy–Ventura, N. (2017). L2 utterance fluency development before, during, and after residence abroad: A multidimensional investigation. *Modern Language Journal*, *101*(2), 275–293.

Hulstijn, J. H. (1990). A comparison between the information-processing and the analysis/control approaches to language learning. *Applied Linguistics*, *11*(1), 30–45.

Hulstijn, J. H. (2001). Intentional and incidental second language vocabulary learning: A reappraisal of elaboration, rehearsal and automaticity. In P. Robinson (Ed.), *Cognition and second language instruction* (pp. 258-286). Cambridge: Cambridge University Press.

Hunt, A. & Beglar, D. (2005) A framework for developing EFL reading vocabulary. *Reading in a Foreign Language, 17*, 23-59.

Hunter, A. M. (2017). *Fluency development in the ESL classroom: The impact of immediate task repetition and procedural repetition on learners' oral fluency.* Doctoral dissertation, University of Surrey.

Iwashita, N., Brown, A., McNamara, T., & O'Hagan, S. (2008). Assessed levels of second language speaking proficiency: How distinct? *Applied Linguistics*, *29*(1), 24–49.

Jarema, G., & Libbe, G. (2007). *The mental lexicon: Core perspectives*. Emeral Group Publishing

Limited.

Jarvis, S. (2003). Probing the effects of the L2 on the L1: A case study. In V. Cook (Ed.), *Effects of the Second Language on the First* (pp. 81–102). Clevedon, UK: Multilingual Matters.

Jiang, N. (2000). Lexical representation and development in a second language. *Applied Linguistics*, *21*(1), 47–77.

Jiang, N. (2002). Form-meaning mapping in second language acquisition. *Studies in Second Language Acquisition*, *24*, 617–637.

Jiang, N. (2004). Semantic transfer and its implications for vocabulary teaching in a second language. *The Modern Language Journal*, *88*(3), 416–432.

Joe, A. (2010). The quality and frequency of encounters with vocabulary in an English for academic purposes programme. *Reading in a Foreign Langauge, 22*(1), 117-138.

Johnson, D. D., Pittelman, S. D., & Heimlich, J. E. (1986). Semantic mapping. *The Reading Teacher*, *39*(8), 778–783.

Jullian, P. (2000). Creating word-meaning awareness. *ELT Journal*, *54*(1), 37–46.

Kahneman, D. (1973). *Attention and effort*. Englewood Cliffs, NJ: Prentice Hall.

Kahng, J. (2014). Exploring utterance and cognitive fluency of L1 and L2 English speakers: Temporal measures and stimulated recall. *Language Learning*, *64*(4), 809–854.

Kahng, J. (2018). The effect of pause location on perceived fluency. *Applied Psycholinguistics*, *39*(3), 569–591.

Kahng, J. (2020). Explaining second language utterance fluency: Contribution of cognitive fluency and first language utterance fluency. *Applied Psycholinguistics*, *41*(2), 457–480.

Kang, S. H. K. (2010). Enhancing visuospatial learning: The benefit of retrieval practice. *Memory and Cognition*, *38*(8), 1009–1017.

Kasper, G. (1992). Pragmatic transfer. *Second Language Research*, *8*, 203–231.

Kroll, J. F., & Stewart, E. (1990). Concept mediation in bilingual translation. *Paper presented at the 31st annual meeting of the Psychonomic Society*. New Orleans, LA.

Kroll, J. F., & Stewart, E. (1994). Category interference in translation and picture naming: Evidence for asymmetric connections between bilingual memory representations. *Journal of Memory & Language, 33*, 149–174.

Kroll, J. F., and N. Tokowitz. (2009). Models of bilingual representation and processing: Looking back and to the future. In J. Kroll and A. M. B. de Groot (Eds.), *Handbook of bilingualism: Psycholinguistic perspectives* (531–554). New York: Oxford University Press.

Rastle, K., Harrington, J., & Coltheart, M. (2002). 358,534 nonwords: The ARC nonword database. *The Quarterly Journal of Experimental Psychology A: Human Experimental Psychology, 55A*(4), 1339–1362.

Kent, G.H. & Rosanoff, A.J. (1910). A study of association in insanity. *American Journal of Insanit*, *67*, 37-96 and 317-390.

Kersten, S. (2010). *The mental lexicon and vocabulary learning: Implications for the foreign language classroom.* Narr Francke Attempto Verlag.

Kelley, K., & Preacher, K. J. (2012). On effect size. *Psychological Methods*, *17*(2), 137–152.

Khodadady, E. (2014). Construct validity of C-tests: A factorial approach. *Journal of Language Teaching and Research*, *5*(6), 1353–1362.

Khoii, R., & Sharififar, S. (2013). Memorization versus semantic mapping in L2 vocabulary acquisition. *ELT Journal*, *67*(2), 199–209.

Klein-Braley, C. (1996). Towards a theory of C-Test processing. In R. Grotjahn (Ed.), *Der C-Test. theoretische grundlagen und praktische anwendungen* (pp. 23–94). Bochum: Brockmeyer.

Kogok, S. O., & Ahamed, M. A. (2017). The effects of using teaching vocabulary learning strategy through semantic mapping, *Journal of Humanities Sciences*, *18*(2), 1–13.

Koizumi, R., & In'nami, Y. (2013). Vocabulary knowledge and speaking proficiency among second language learners from novice to intermediate levels. *Journal of Language Teaching and Research*, *4*(5).

Koponen, M., & Riggenbach, H. (2000). Varying perspectives on fluency. In Riggenbach (Ed.), *Perspectives on fluency* (pp. 102–127). Michigan: University of Michigan Press.

Kormos, J. (2006a). *Speech production and second language acquisition*. Mahwah, New Jersey: Lawrence Erlbaum.

Kormos, J. (2006b). The structure of self-repairs in the speech of Hungarian learners of English. *Acta Linguistica Hungarica*, *53*(1), 53–76.

Kormos, J., & Dénes, M. (2004). Exploring measures and perceptions of fluency in the speech of second language learners. *System*, *32*(2), 145–164.

Kormos, J. (2011). Speech production and the Cognition Hypothesis. In *Second language task complexity researching the cognition hypothesis of language learning and performance* (pp. 39–58). John Benjamins Publishing Company.

Kruse, H. Pankhurst, J. & Sharwood S. M. (1987). A multiple word association probe in second language acquisition research, *Studies in Second Language Acquisition*, *9*, 141–154.

Lambert, C., Kormos, J., & Minn, D. (2017). Task repetition and second language speech processing. *Studies in Second Language Acquisition*, *39*(1), 167–196.

Lathem_koenig, Ch. & Oxenden, C.(2013). *American English file* (2nd Ed.). Oxford University Press. Oxford, UK.

Laufer, B., and J. Hulstijn. (2001). Incidental vocabulary acquisition in a second language. *Applied Linguistics, 22,* 1–26.

Laufer, B., & Nation, P. (2001). Passive vocabulary size and speed of meaning recognition: Are they related? *EUROSLA Yearbook*, *1*(1), 7–28.

Lee, M., Williams, J. N. (2001). Lexical access in spoken word production by bilinguals: evidence from the semantic competitor priming paradigm. *Bilingualism, 4*(3), 233-248.

Lennon, P. (1990). Investigating fluency in EFL: A quantitative approach. *Language Learning*, *40*(3), 387–417.

Lennon, P. (2000). The lexical element in spoken second language fluency. In H. Riggenbach (Ed.), *Perspectives on fluency* (pp. 25–42). Michigan: University of Michigan Press.

Levelt, W. J. M. (1989). *Speaking. From intention to articulation*. Cambridge, Mass: The MIT Press.

Levelt, W. J. M. (1999). Producing spoken language: A blueprint of the speaker. In C. Brown & P. Hagoort (Eds.), *Neurocognition of language* (pp. 88–122). Oxford, England: Oxford University Press.

Levelt, Willem J. M., Roelofs, A., & Meyer, A. s. (1999). A theory of lexical access in speech production. *Behavioral and Brain Sciences*, *22*, 1–75.

Levelt, Willem J.M. (1983). Monitoring and self-repair in L2. *Cognition*, *14*, 41–104. Lippman, M. Z. (1971). Correlates of contrast word associations: Developmental trends. *Journal of Verbal Learning and Verbal Behavior*.

Llanes, A., & Muñoz, C. (2009). A short stay abroad: Does it make a difference? *System*, *37*(3), 353–365.

Llanes, A., & Muñoz, C. (2013). Age effects in a study abroad context: children and adults studying abroad and at home, *Language Learning, 63*(1).

Logan, G. D. (1988). Automaticity, resources, and memory: Theoretical controversies and practical implications. *Human Factors*, *30*(5), 583–598.

Machalias, R. (1991). Semantic networks in vocabulary teaching and their application in the foreign language classroom. *Journal of the Australian Modern Language Teachers' Association*, *26*(3), 19–24.

MacKay, D. G. (1982). The problems of flexibility, fluency and speed-accuracy trade-off in skilled behaviour. *Psychological Review1*, *89*, 483–505.

Maclay, H., & Osgood, Ch., E. (1959). Hesitation phenomena in spontanous English speech. *Word, 15*(1), 19-44.

Margosein, C. M., Pascarella, E. T., & Plfaum, S. W. (1982). The effect of instrcution using semantic mapping on vocabulary and comprehension. *Journal of Early Adolescence*, *2*(2), 185–194.

Mayer, M., & Mayer, M. (1971). *A boy, a dog, a frog, and a friend.* New York: Puffin Pied Piper.

Mazdayasna, G., & Molaei, G. (2015). Needs analysis for General English courses: a model for setting priorities. In C. Kennedy (Ed.), *English language teaching in the Islamic Republic of Iran: Innovations, trends and challenges* (pp. 49–59). London: British Council.

McNeill, D. (1966). A study of word association. *Journal of Verbal Learning and Verbal Behavior*, *5*(6), 548–557.

Meara, P. (1978). Learners' word associations in French. *The Interlanguage Studies Bulletin, 3*(2), 192-211.

Meara, P. (1983). Word associations in a foreign language: A report on the Birkbeck vocabulary project. *Nottingham Linguistic Circular, 11*, 29–38.

Meara, P. (1992). Vocabulary in a second language. *Reading in a Foreign Language*, *3*(9), 761-810.

Meara, P. (1996). The dimensions of lexical competence. In G. Brown, K. Malmkjaer, and J. Williams (Eds.), *Performance and competence in second language acquisition*. Cambridge, UK: Cambridge University Press.

Meara, P. (2005). Lexical frequency profiles: A Monte Carlo analysis. *Applied Linguistics, 26*, 32–47.

Meara, P. (2009). *Connected words: Word associations and second language vocabulary acquisition*. Amsterdam: Philadelphia: John Benjamins.

Meara, P., & Fitzpatrick, T. (2000). Lex30: an improved method of assessing productive vocabulary in an L2. *System*, *28*, 19–30.

Meara, P. (1983). Word associations in a foreign language. *Nottingham Linguistics Circular, 11*(2), 17-34.

Meara, P. (1984). The study of lexis in interlanguage. In A. Davies, C. Criper, & A. P. R. Howatt (Eds.), *Interlanguage* (pp. 225–235). Edinburgh, Scotland: Edinburgh University Press.

Meara, P. (2004). Modelling vocabulary loss. *Applied Linguistics*, *25*(2), 137–155.

Meara, P. (2010). *EFL vocabulary tests*. ERIC Clearinghouse, Swansea, Wales.

Meara, P. & Wolter, B. (2004). V-Link: Beyond vocabulary depth. In D. Albrechtsen, K. Haastrup, &

B. Henriksen (Eds.), *Angles on the English-speaking world. Vol. 4* (pp. 85–96). Copenhagen, Denmark: Museum Tusculanum Press.

Mehnert, U. (1998). The effects of different lengths of time for planning on second language performance. *Studies in Second Language Acquisition, 20*(1), 83-108.

Mertens, D. M. (2010). *Research and evaluation in education and psychology: Integrating diversity with quantitative, qualitative, and mixed methods.* Los Angeles: Sage.

Michel, M. (2017). Complexity, accuracy, and fluency in L2 production. In Loewen, S. & M. Sato (Eds.), *The routledge Handbook of Instructed second language acquisition* (pp. 50-69). New Yoprk and London, Routledge.

Michael, E., & Gollan, T. (2005). Being and becoming bilingual: Individual differences and consequences for language production. In A. M. B. De Groot & J. Kroll (Eds.), *Handbook of Bilingualism: Psycholinguistic Approaches* (pp. 389–407). Oxford: Oxford University Press.

Miller, E. N., & Chapman., L. J. (1983). Continued word association in hypothetically psychosis-prone college students. *Journal of Abnormal Psychology*, *92*(4), 468–78.

Mirhosseini, S. A., & Khodakarami, S. (2015). A glimpse of contrasting de jure-de facto ELT policies in Iran. In C. Kennedy (Ed.), *English Language Teaching in the Islamic Republic of Iran: Innovations, Trends and Challenges* (pp. 23–34). London, UK: British Council.

Mochida, A & Harrington, M. (2006). The Yes/No test as a measure of receptive vocabulary knowledge. *Language testing. 23*(1). 73-98.

Mohammed, F. O., & Malo, S. S. (2020). Learning vocabularies through semantic mapping strategy. *Humanities Journal of University of Zakho*, *8*(2), 332–336.

Mora, J. C., & Valls-ferrer, M. (2020). Oral fluency , accuracy , and complexity in formal instruction and study abroad learning contexts. *TESOL QuarterlySchoonen*, *46*(4), 610–641.

Morin, R., & Goebel, J. (2001). Basic vocabulary instruction: Teaching strategies or teaching words? *Foreign Language Annals*, *34*(1), 8–17.

Muijs, D. (2004). *Doing quantitaive research in education with SPSS.* SAGE publication. London, UK.

Murphy, M. L. (2003). *Semantic relations and the lexicon; Antonymy, synonymy, and other paradigms*. Cambridge: Cambridge University Press.

Ellis. N. C. (2002). Frequency effects in language processing: A review with implications for theories of implicit and explicit language acquisition. *Studies in Second Language Acquisition*, *24*, 143–188.

Naeimi, M & Voo Foo, T. Ch., (2014). A Comparison of vocabulary learning strategies of Iranian EFL university students: Repeating versus cooperating with peers. *English Language Teaching, 7*(17), 102-110.

Nakagawa, S. (2004). A farewell to Bonferroni: the problems of low statistical power and publication bias, *Behavioral Ecology, 15*(6), 1044-1045.

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

Namei, S. (2002). *The Bilingual lexicon from a developmental perspective. A word association study of Persian-Swedish bilinguals*. Doctoral dissertation. Stockholm University, Sweden.

Nation, I. S. P. & Gu, P. Y. (2007). *Focus on vocabulary*. Sydney: NCELTR.

Nation, I. S. P. (1996). The four strands of a language course. *TESOL in Context*, *6*(2), 7-12.

Nation, I. S. P. (2001). *Learning vocabulary in another language*. Cambridge University Press.

Nation, I. S. P. (2006). How large a vocabulary is needed for reading and listening? *Canadian Modern Language Review*, *63*(1), 59–82.

Nation, I. S. P. (2008). *Teaching vocabulary: Techniques and strategies*. Boston: Heinle & Heinle.

Nation, I. S. P. (2020). The different aspects of vocabulary knowledge. In S. Webb (Ed.), *The Routledge Handbook of Vocabulary Studies* (pp. 15–30). Taylor & Francis.

Nation, I. S. P. & Meara, P. (2002). Vocabulary. In N. Schmitt (Ed.), *An introduction to applied linguistics* (pp. 35–54). London: Arnold.

Nation, P., & Newton, J. (1997). *Teaching vocabulary.* Cambridge: Cambridge University Press

Navracsics, J. (2007). Word classes and the bilingual mental lexicon. In Z. Lengyel & J. Navracsics (Eds.), *Second language lexical processes: Applied linguistic and psycholinguistic perspectivess*. Multilingual Matters.

Neff, K. J. E. (1991). *Neural net models of word representation*. Doctoral thesis, Ball State University.

Nelson, D. L., McEvoy, C. L., & Schreiber, T. A. (2004). The University of South Florida word association, rhyme, and word fragment norms. *Behavior Research Methods*, *36*, 402–407.

Nelson, K. (1977). The syntagmatic-paradigmatic shift revisited: A review of research and theory. *Psychological Bulletin*, *84*(1), 93–116.

Newton, R. R., & Rudestam, K. E. (1999). *Your statistical consultant : Answers to your data analysis questions*. SAGE publications.

Nissen, H. B., & Henriksen, B. (2006). Word class influence on word association test results. *International Journal of Applied Linguistics*, *16*(3), 389–408.

O'Brien, I., Segalowitz, N., Freed, B., & Collentine, J. (2007). Phonological memory predicts second language oral fluency gains in adults. *Studies in Second Language Acquisition*, *29*(04), 557–581.

Oldfield, R. (1966). Things, words, and the brain. *Quarterly Journal of Experimental Psychology, 18,* 340-353.

Orita, M. (2002). Word associations of Japanese EFL learners and native speakers: Shifts in response type distribution and the associative development of individual words. *Review of English Language Education in Japan*, *13*, 111–120.

Palermo, D. S. (1971). Characteristics of word association responses obtained from children in grades one through four. *Developmental Psychology, 5*(1), 118–123.

Paradis, M. (1994). Neurolinguistic aspects of implicit and explicit memory: Implications for bilingualism and SLA. In N. Ellis (Ed.), *Implicit and explicit learning of languages* (pp. 393–419). San Diego: Academic Press.

Paradis, M. (1997). The cognitive neuropsychology of bilingualism. In A. De Groot & J. Kroll (Ed.), *Tutorials in bilingualism. Psychologic perspectives* (pp. 331-354). Mahwah, NJ: Lawrence Erlbaum.

Paradis, M. (2004). *A Neurolinguistic theory of bilingualism*. Amsterdam: Benjamins.

Paradis, M. (2000). Cerebral representation of bilingual concepts. *Bilingualism: Language and cognition, 3,* 22–24.

Pavlenko, A. (2000). New approaches to concepts in bilingual memory. *Bilingualism: Language and cognition*, *3*(1), 1–4.

Pavlenko, A. (2003). Eyewitness memory in late bilinguals: Evidence for discursive relativity. *International Journal of Bilingualism*, *7*(3), 257–281.

Pavlenko, A. (2008). Structural and conceptual equivalence in the acquisition and use of emotion words in a second language. *The Mental Lexicon*, *3*(1), 91–120.

Pavlenko, A. (2009). *The bilingual mental lexicon; Interdisciplinary approaches*. Multilingual Matters.

Pittelman, S.D., Heimlich, J. E., Berglund, R. L., & French, M. P. (1991). *Semantic feature analysis: Classroom applications*. Newark, DE: International Reading Association.

Pittelman, S. D., Levin, K. M., & Johnson, D. D. (1985). An investigation of two instructional settings in the use of semantic mapping with poor readers. *Program Report, 85*(4), 1-28.

Playfoot, D., Balint, T., Pandya, V., Parkes, A., Peters, M., & Richards, S. (2016). *Applied linguistics*, *39*(5), 607–624.

Plonsky, L., & Oswald, F. L. (2014). How big is "Big"? Interpreting effect sizes in L2 research. *Language Learning*, *64*(4), 878–912.

Politzer, R. L. (1978). Paradigmatic and syntagmatic associations of first year French students. In V. Hausa (Ed.), *Papers in linguistics and child language* (pp. 203-210). Berlin: Mouton de Gru.

Potter, M. C., So, K., Von, Eckardt, B. & Feldman, L. B. (1984). Lexical and conceptual representation in beginning and more proficient bilinguals. *Journal of Verbal Learning and Verbal Behavior, 23*, 23-38.

Poulisse, N. (1997). Langauge production in bilinguals. In Annette M.B. de Groot & J. F. Kroll (Eds.), *Tutorials in bilingualism: Psycholinguistic perspectives* (pp. 201–225). New York: Psychology Press.

Préfontaine, Y. (2013). Perceptions of French fluency in second language speech production. *Canadian Modern Language Review*, *69*(3), 324–348.

Préfontaine, Y., Kormos, J., & Johnson, D. E. (2016). How do utterance measures predict raters' perceptions of fluency in French as a second language? *Language Testing*, *33*(1), 53–73.

Randall, M. (1980).Word association behaviour in learners of English as a second language. *Polyglot 2*,2, B4-D1.

Randall, M. (2007). *Memory, psychology, and second language learning*. John Benjamins Publishing Company.

Razmjoo, S. A., & Riazi, A. M. (2006). On the teaching methodology of Shiraz EFL institutes. *Journal of Social Sciences and Humanities of Shiraz University*, *23*(1), 58–70.

Read, J. (2000). *Assessing vocabulary*. Cambridge: Cambridge University Press.

Read, J. (2004). Plumbing the depths: How should the construct of vocabulary knowledge be defined? In P. Bogaards, and B. Laufer (Eds.), *Vocabulary in a second language: Selection, acquisition, and testing* (pp. 209–227). Amsterdam: John Benjamins.

Riazantseva, A. (2001). Second language proficiency and pausing;A Study of Russian Speakers of English. *Studies in Second Language Acquisition*, *23*(4), 497–526.

Riggenbach, H. (1991). Toward an understanding of fluency: A microanalysis of nonnative speaker conversations. *Discourse Processes*, *14*(4), 423–441.

Roberts, B., & Kirsner, K. (2000). Temporal cycles in speech production. *Language and Cognitive Processes, 15*(2), 129–157.

Robinson, P. (2001). Task complexity, task difficulty, and task production: Exploring Interactions in a Componential Framework. *Applied Linguistics*, *22*(1), 27–57.

Roelofs, A. (1992). A spreading-activation theory of lemma retrieval in speaking. *Cognition, 42*, 107_142.

Roelofs, A. (2000). Word meanings and concepts: what do the findings from aphasia and language specificity really say? *Bilingualism: Language and Cognition*, *3*(1), 25–27.

Rossiter, M. J. (2009). Perceptions of L2 fluency by native and non-native speakers of english. *Canadian Modern Language Review*, *65*(3), 395–412.

Rossiter, M. J., Derwing, T. M., Manimtim, L. G., & Thomson, R. I. (2010). Oral fluency: The neglected component in the communicative language classroom. *Canadian Modern Language Review*, *66*(4), 583-606.

Rothman, K. J. (1990). No adjustments are needed for multiple comparisons, *Epidemiology, 1*(1), 43-46.

Ruke-Dravina, V. (1971). Word associations in monolingual and multilingual individuals. *Linguistics, 74*, 66-84.

Russell, P. (1979). *The Brain book*. London: Routledge and Kegan Paul.

Sadeghi, K., & Richards, J. C. (2015). Teaching spoken English in Iran's private language schools: issues and options. *English Teaching: Practice & Critique*, *14*(2), 210–234.

Sag, I. A., & Wasow, T. (1999). *Syntactic theory: a formal introduction*. Stanford, CA: CSLI Publications.

Sagarra, N. & Alba, M. (2006). The key is in the keyword: L2 vocabulary learning methods with beginning learners of Spanish. *The Modern Language Journal, 90*(2), 228-243.

Saito, K., Ilkan, M., Magne, V., Tran, M. N., & Suzuki, S. (2018). Acoustic characteristics and learner profiles of low-, mid-and high-level second language fluency. *Applied Psycholinguistics*, *39*(3), 593–617.

Sajavaara, K. (1987). Speech production: Factors affecting fluency. In D. & M. Raupach (Ed.), *Psycholinguistic models of production* (pp. 45–65). Norwood, NJ: Ablex Publishing Corporation.

Sample, E. & Michel, M. (2014). An exploratory study into trade-off effects of complexity, accuracy, and fluency on young learners' oral task repetition. *TESL Canada Journal, 31*(8), 23-46.

Sánchez, M. J. (2002). Efecto de la instruccion conpatrones lingu¨ ı´sticos en el aprendizaje le´xico: camposema´ ntico shine. *Estudios Ingleses de La Universidad Complutense*, *10*, 183–199.

Sánchez, M. J. (2004). Effect of instruction with expert patterns on the lexical learning of English as a foreign language. *System*, *32*(1), 89–102.

Sangarun, J. (2005). The effects of focusing on meaning and form in strategic planning. In R. Ellis (Ed.), *Planning and task performance in a second language* (pp. 111–142). Philadelphia: John Benjamins.

Saragih, E. (2019). The effect of semantic mapping technique on technical vocabulary mastery for Midwifery students. *Studies in English Language and Education*, *6*(2), 333–342.

Saussure, F. D. (1916). *A Course in general linguistics*. London: Duckworth.

Schmidt, R. (1992). Psychological mechanisms underlying second language fluency. *Studies in Second Language Acquisition*, *14*(4), 357–385.

Schmidt, R. (2001). Attention. In Peter Robinson (Ed.), *Cognition and second language instruction* (pp. 1–32). New York: Cambridge university Press.

Schmitt, N. (1998a). Quantifying word association responses: What is native-like? *System*, *26*, 389–401.

Schmitt, N. (1998b). Tracking the incremental acquisition of second language vocabulary: A longitudinal study. *Language Learning*, *48*(2), 281–317.

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

Schmitt, N. (2010). *Researching vocabulary: A vocabulary research manual*. UK: Palgrave McMillan.

Schmitt, N., Ng, J. W. C., & Garras, J. (2011). The word associates format: Validation evidence. *Language Testing*, *28*, 105–126.

Schmitt, N., & Schmitt, D. (2014). A reassessment of frequency and vocabulary size in L2 vocabulary teaching. *Language Teaching, 47*(4), 484–503.

Schmitt, N., & Schmitt, D. (1995). Vocabulary notebooks: theoretical underpinnings and practical suggestions. *ELT Journal*, *49*(2). 133-143.

Schneider, W. (1985). Toward a model of attention and the development of automatic processing. In M. I. Posner & O. S. Marin (Eds.), *Attention and performance XI* (pp. 474-492), Hillsdale, NJ., Erlbaum.

Schneider, W., Dumas, S. T., & Shiffrin, R. M. (1984). Automatic and controlled processing and attention. In R. Parasuraman & D. R. Davies (Eds.), *arieties of attention* (pp. 1–27). New York: Academic Press.

Shiffrin, R. M., & Czerwinski, M. P. (1988). A model of automatic attention attraction when mapping is partially consistent. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 14,* 562-569.

Segalowitz, N. (2003). Automaticity and second languages. In C. Doughty & M. Long (Eds.), *The handbook of second langauge acquisition* (pp. 382–408). Oxford, UK: Blackwell.

Segalowitz, N., & Hulstijn, J. (2005). Automaticity in bilingualism and second language learning. In J. Kroll & A. De Groot (Eds.), *Handbook of bilingualism: Psycholinguistic approaches* (pp. 371–388). Oxford University.

Segalowitz, N. (2010). Cognitive bases of second language fluency. In *Cognitive bases of second* *language fluency*.

Segalowitz, N. (2016). Second language fluency and its underlying cognitive and social determinants. *IRAL - International Review of Applied Linguistics in Language Teaching*, *54*(2), 79–95.

Segalowitz, N., & Freed, B. F. (2004). Context, contact, and cognition in oral fluency acquisition: Learning Spanish in at home and study abroad contexts. *Studies in second language acquisition*, *26*(02), 173-199.

Segalowitz, N., French, L., Guay, J., Segalowitz, N., & French, L. (2018). What features best characterize adult second language utterance fluency and what do they reveal about fluency gains in short-term immersion ? *Canadian Journal of Applied Linguistics Revue Canadienne de Linguistique Appliquée*, *20*(2), 90–116.

Seifoori, Z., & Vahidi, Z. (2012). The impact of fluency strategy training on Iranian EFL learners' speech under online planning conditions. *Language Awareness*, *21*(1-2), 101–112.

Seyfeddinipur, M., Kita, S., & Indefrey, P. (2008). How speakers interrupt themselves in managing problems in speaking: Evidence from self-repairs. *Cognition*, *108*, 837–842.

Shademan, S. (2002). *Epenthetic vowel harmony in Farsi*. Los Angeles: University of California.

Shatzz, C. J. (1992). The developing brain. *Scientific American*, *267*(3), 60–67.

Singleton, C. (2009*) Intervention for dyslexia*. Bracknell: Dyslexia-SpLD Trust.

Singleton, D. (1999). *Exploring the second language mental lexicon*. Cambridge, UK: Cambridge University Press.

Singleton, D. (2007). How integrated is the integrated mental lexicon? In Zs. Lengyel & J. Navracsics (Eds.), *Second language lexical processes: Applied linguistic and psycholinguistic perspectives* (pp. 3–16). Clevedon, UK: Multilingual Matters.

Singleton, D. (1999). *Exploring the second language mental lexicon*. Cambridge: Cambridge University Press.

Skehan, P. (2003). Task-based instruction. *Language Teaching*, *36*(1), 1–14.

Skehan, P. (2009). Modelling second language performance: Integrating complexity, accuracy, fluency, and lexis. *Applied Linguistics*, *30*(4), 510–532.

Skehan, P., & Foster, P. (2005). Strategic and on-line planning: The influence of surprise information and task time on second language performance. In R. Ellis (Ed.), *Planning and task performance in a second language* (pp. 193–216). Amsterdam: John Benjamins.

Skehan, P. (2014). Limited attentional capacity, second language performance, and taskbased pedagogy. In Peter Skehan (Ed.), *Processing perspectives on task performance* (pp. 211–260). Amsterdam: John Benjamins Publishing Company.

Skehan, P., & Foster, P. (1999). The influence of task structure and processing conditions on narrative retellings. *Language Learning*. *49*(1), 93-120.

Skehan, P., Foster, P., & Shum, S. (2016). Ladders and snakes in second language fluency. *IRAL - International Review of Applied Linguistics in Language Teaching*, *54*(2), 97–111.

Söderman, T. (1993). Word associations of foreign language learners and native speakers: Different response types and their relevance to lexical development. In *Problem, process, product in language learning: Papers from the Stockholm-Abo Conference,* B. Hammarberg (ed.) Abo: AfinLA (pp. 157–169).

Sökmen, A. (1993). Word association results: a window to the lexicons of ESL students. *JALT Journal*. *15*(5), 135-150.

Sökmen, A. (1997). Current trends in teaching second language vocabulary. In & N. Schmitt (Ed.), *Vocabulary: Description, acquisition and pedagogy* (pp. 237–257). Cambridge: Cambridge University Press.

Sommer, R. (1910). *Diagnostik der geisteskrankheiten für Praktische arzte und studierende [Diagnosis of mental illness for practical physicians and students].* Berlin/Wien: Urban & Schwarzenberg.

Stahl, S. A., & Vancil, S. J. (1986). Discussion semantic vocabulary is what makes in work maps instruction. *The Reading Teacher*, *40*(1), 62–67.

Stoller, F. L., & Grabe, W. (1993). Implementations for L2 vocabulary acquisition and instruction from L1 vocabulary research second language reading and vocabulary learning. In T. Huckin, M. Hayynes, & J. Body (Eds.), *Second language reading and vocabulary learning* (pp. 24–45). Norwood, New Jersey: Ablex.

Stubbs, M. (2001). *Words and phrases: Corpus studies of lexical semantics*. Blackwell.

Suzuki, S., & Kormos, J. (2019). Linguistic dimensions of comprehensibility and perceived fluency: an investigation of complexity, accuracy, and fluency in second language argumentative speech. *Studies in Second Language Acquisition*, *42*(1), 1–25.

Svenconis, D. J., & Kerst, S. (1994). Investigating the teaching of second language vocabulary through semantic mapping in a hypertext environment. *CALICO Journal*, *12*(2–3), 33–57.

Tavakoli, P. (2009). Investigating task difficulty: Learners' and teachers' perceptions. *International Journal of Applied Linguistics*, *19*(1), 1–25.

Tavakoli, P. (2011). Pausing patterns: Differences between L2 learners and native speakers. *ELT Journal*, *65*(1), 71–79.

Tavakoli, P. (2016). Fluency in monologic and dialogic task performance: Challenges in defining and measuring L2 fluency. *International Review of Applied Linguistics in Language Teaching*, *54*, 133–150.

Tavakoli, P. (2020). Making fluency research accessible to second language teachers: The impact of a training intervention. *Language Teaching Research*.

Tavakoli, P, Campbell, C., & Mccormack, J. (2016). Development of speech fluency over a short period of time: Effects of pedagogic intervention. *TESOL Quarterly*, *50*(2), 447–471.

Tavakoli, P., & Foster, P. (2011). Task design and second language performance: The effect of narrative type on learner output. *Language Learning*., *58*(2),439-473.

Tavakoli, P., & Hunter, A. M. (2018). Is fluency being ‘neglected’ in the classroom? Teacher understanding of fluency and relatead classroom practices. *Language Teaching Research*, *22*(3), 330–349.

Tavakoli, P., Nakatsuhara, F., & Hunter, A. M. (2020). Aspects of fluency across assessed levels of speaking proficiency. *Modern Language Journal*, *104*(1), 169–191.

Tavakoli, P., & Skehan, P. (2005). Strategic planning, task structure, and performance testing. In R. Ellis, *Planning and task performance in a second lanugage*. Amsterdam: John Benjamins.

Teijlingen, E. V., & Hundley, V. (2002). The importance of pilot studies. *Nursing Stardard*, *16*(40), 33–36.

Thai, C., & Boers, F. (2016). Repeating a monologue under increasing time pressure: Effects on fluency, complexity, and accuracy. *TESOL Quarterly*, *50*(2), 369–393.

Thomson, R. I., & Isaacs, T. (2011). Speaking in a drone: Are listeners attuned to pitch-related differences when rating L2 speech samples for personality attributes? *American Association of Applied Linguistics Conference,* Chicago, IL.

Towell, R., Hawkins, R., & Bazergi, N. (1996). The development of fluency in advanced learners of French. *Applied Linguistics*, *17*(1), 84–119.

Tracy-Ventura, N., McManus, K., Norris, J., & Ortega, L. (2014). “Repeat as much as you can”: Elicited imitation as a measure of oral proficiency in L2 French. In P. Leclercq, H. Hilton, & A. Edmonds (Eds.), *Measuring L2 proficiency: Perspectives from SLA* (pp. 143-166). Multilingual Matters.

Uchihara, T. & Clenton, J., (2020). Investigating the role of vocabulary size in second language speaking ability. *Language Teaching Research*, *24*(4), 540-556.

Uchihara, T., & Saito, K. (2019). Exploring the relationship between productive vocabulary knowledge and second language oral ability. *Language Learning Journal*, *47*(1), 64–75.

Uchihara, T., Saito, K., & Clenton, J. (2020). Re-examining the relationship between productive vocabulary and second language oral ability. In J. Clenton & P. Booth (Eds.), *Vocabulary and the four skills; pedagogy, practice, and implications for teaching vocabulary* (pp. 146–165). Taylor & Francis.

Van Hell, J. G., & De Groot, A. M. B. (1998). Conceptual representation in bilingual memory: Effects of concreteness and cognate status in word association. *Bilingualism: Language and Cognition*, *1*(3), 193–211.

Varonis, E. M., & Gass, S. (1982). The comprehensibility of non-Native speech. *Studies in Second Language Acquisition*, *4*(2), 114–136.

Wang, Z. (2014). On-line time pressure manipulations: L2 speaking performance under five types of planning and repetition conditions. In P. Skehan (Ed.), *Processing Perspectives on Task Performance* (pp. 27–62). John Benjamins.

Warning, R & Nation, P. (2004). Second language reading and incidental vocabulary learning, *Angles on the English Speaking World, 4,* 11-23.

Webb, S. (2007). The effect of repetition on vocabulary knowledge. *Applied Linguistics, 28*, 46–65.

Webb, S. (2020). Incidental vocabulary learning. In S. Webb (Ed.), *The Routledge Handbook of Vocabulary Studies* (pp. 225–240). Routledge.

West, M. (1953). *General service list of English words*. London, UK: Longman, Green.

Wigglesworth, G. (1997). An investigation of planning time and proficiency level on oral test discourse. *Language Testing*, *14*(1), 85-106.

Wilcox, R. (2017). *Modern statistics for the social and behavioral sciences*. New York: Chapman and Hall/CRC.

Wilkins, D. A. (1972). *Linguistics in language teaching.* Australia: Edward Arnold.

Wilks, C., & Meara, P. (2002). Untangling word webs: Graph theory and the notion of density in second language word association networks. *Second Language Research*, *18*(4), 303–324.

Wilks, C., & Meara, P. (2007). Implementing graph theory approaches to the exploration of density and structure in L1 and L2 word association networks. In H. Daller, J. Milton, & J. Treffers-Daller (Eds.), *Modeling and assessing vocabulary knowledge* (pp. 167–181). Cambridge:Cambridge University Press.

Wilks, C., Meara, P., & Wolter, B. (2005). A further note on simulating word association behaviour in a second language. *Second Language Research*, *21*(4), 359–372.

Witton-Davies, G. (2014). *The study of fluency and its development in monologue and dialogue*. Doctoral Dissertation, Lancaster University, Lancaster, UK.

Wolter, B. (2001). Comparing the L1 and L2 Mental Lexicon. *Studies in Second Language Acquisition*, *23*(01), 41–69.

Wolter, B. (2002). Assessing proficiency through word associations: Is there still hope? *System*, *3*, 315-329

Wood, D. (2010). *Formulaic language and second language speech fluency*. Continuum.

Yang, Y. I. J. (2014). *Is speaking fluency strand necessary for the college students to develop in the EFL class ?, 4*(2), 225–231.

Yokokawa, H., Yabuuchi, S., Kadota, S., Nakanishi, Y., & Noro, T. (2002). Lexical Evidence, networks in L2 mental lexicon: Japanese, from a word-association task for Learners. *Language Education and Technology*, *39,* 21–39.

Yuan, F., & Ellis, R. (2003). The effects of pre-tasl planning and on-line planning on fluency, complexity and accuracy in L2 monologic oral production. *Applied Linguistics*, *24*, 1–27.

Zahedi, Y., & Abdi, M. (2012). The effect of semantic mapping strategy on EFL learners' vocabulary learning. *Procedia-Social and Behavioral Sciences*, *69*, 2273–2280.

Zandian, S. (2015). Migrant literature and teaching English as an international language in Iran. In C. Kennedy (Ed.), *English language teaching in the Islamic Republic of Iran: Innovations, trends and challenges* (pp. 113–123). British Council.

Zareva, A. (2016). Effects of lexical class and word frequency on the L1 and L2 English-based lexical connections. *The Journal of Language Learning and Teaching*, *1*(2), 1–17.

Zareva, A., Schwanenflugel, P., & Nikolova, Y. (2005). Relationship between lexical competence and language proficiency: Variable sensitivity. *Studies in Second Language Acquisition*, *27*(4), 567-595.

Zareva, A., & Wolter, B. (2012). The "promise" of three methods of word association analysis to L2 lexical research. *Second Language Research*, *28*(1), 41–67.

Zeng, T., Branigan, H. & Pickering, M., (2020). Do bilinguals represent between-language relationships beyond the word level in their lexicon? *Journal of neurolinguistics. 55.*