

**An Investigation on Productive Vocabulary, Fluency and their Interaction  
in Study Abroad Context.**

Vladyslav Gilyuk

Master 2 Langues, Littératures et Civilisations Étrangères et Régionales: Études  
Anglophones, Université Paul Valéry

W411/2AN5: Mémoire de recherche

Prof. Amanda Edmonds and Prof. Elisa Sneed German

29 June 2021



## Abstract

The present study investigated the development of productive vocabulary, fluency and their interaction in the study abroad (SA) context. The goal of the study was to find out how productive vocabulary and fluency changed over the SA period and to determine whether these changes showed parallel developmental trajectories that may signal the interaction between the two constructs. A pretest-posttest design was used to qualitatively study the changes for five intermediate/advanced learners of English, who spent one academic year in an English-speaking country. For the fluency investigation, the interviews were conducted before and after the SA experience and analyzed in terms of utterance fluency. Similarly, a Lex30 test was conducted before and after the SA experience in order to analyze productive vocabulary. The results showed that fluency significantly developed over the SA period, whereas the development of productive vocabulary was less evident. Finally, due to the differences in the results the interaction between fluency and productive vocabulary in the SA context was not found.

*Keywords:* Productive Vocabulary, Fluency, Study Abroad, Second Language Acquisition, Lex30

## Résumé

Dans le présent mémoire, j'ai cherché à examiner le développement du vocabulaire productif, la fluence orale et leur interaction en contexte de séjour linguistique.

L'idée pour ce mémoire est en partie basée sur les travaux de Foster (2020), qui a souligné le manque d'études sur les relations entre la fluence orale et d'autres construits et a suggéré d'étudier la relation entre la fluence orale et le vocabulaire productif. De plus, le contexte du séjour linguistique a été choisi pour étudier et comparer les changements de fluence orale et de vocabulaire productif car il permet des opportunités d'acquisition de la langue plus naturelles. L'objectif de ce mémoire est par conséquent d'étudier comment le vocabulaire productif et la fluence orale ont changé au cours de la période de séjour linguistique et de déterminer si ces changements montrent des trajectoires de développement parallèles qui peuvent signaler l'interaction entre les deux compétences langagières.

La première partie du mémoire présente la revue de la littérature pour chacun des trois grands thèmes : fluence orale, vocabulaire productif et séjour linguistique. Pour la revue de la littérature sur la fluence orale, les définitions aux sens large et étroit ont été présentées en mettant l'accent sur la définition de Lennon (1990) : “ an impression on the listener's part that the psycholinguistic processes of speech planning and speech production are functioning easily and efficiently” [une impression de la part de l'auditeur que les processus psycholinguistiques de la planification du discours et la production du discours fonctionnent facilement et efficacement] (p. 391) qui a été développée par Segalowitz en trois constructions de la fluence orale : la fluence cognitive (processus de planification de la parole), la fluence énoncée (processus de production de la parole) et la fluence perçue (processus d'acceptabilité de la parole produite par l'interlocuteur). Cependant, seule la fluence de l'énoncé est généralement utilisée dans la plupart des études pour mesurer la fluence orale, car elle est basée sur des mesures strictes qui sont indépendantes de la perception humaine et garantissent ainsi plus d'objectivité. Skehan (2003) et Tavakoli et Skehan (2005) ont proposé 3 aspects cruciaux pour la mesure de la fluence de l'énoncé : 1) la fluence de la vitesse, qui est responsable de la mesure de la vitesse, du débit et de la densité de la parole produite, 2) la fluence de rupture, qui mesure les phénomènes de pause et d'hésitation, 3) la fluence de réparation qui mesure les reformulations, les corrections et les répétitions. La revue des études a montré que les 3 mesures de fluence de la vitesse (les taux d'articulation avec hésitations, taux d'articulation sans hésitation, et longueur moyenne des énoncés) sont parmi les réflecteurs de fluence les plus fiables. De même, l'identification du nombre, de la durée et de l'emplacement de la pause a également été considérée comme un

réflecteur fiable de la fluence. Il est inutile de mentionner qu'une attention particulière a été attribuée à l'emplacement de la pause, qui était connecté au modèle de Levelt (1989) composé de 3 étapes : conceptualisateur, formulateur et articulateur. Sur la base du modèle de Levelt (1989), il a été suggéré que les pauses entre les clauses sont liées à l'étape de conceptualisation, ce qui signifie qu'elles surviennent à la suite de la planification de la clause suivante ou de la recherche de nouvelles idées de discussion (macro-planification) tandis que les pauses dans les clauses sont liées aux étapes de formulation et/ou d'articulation, ce qui signifie qu'elles surviennent à la suite de la recherche d'un certain mot, d'une forme grammaticale ou d'une orthographe (micro-planification) (Skehan, Foster, & Shum, 2016). De plus, le choix de l'unité pour la division de la parole a été discuté comme un facteur important pour l'identification de l'emplacement de la pause, en se concentrant sur l'AS-unit par Foster, Tonkyn et Wigglesworth (2000). Enfin, le seuil de pause a été discuté sur la base de l'étude de De Jong et Bosker (2013) qui a trouvé qu'une pause de 250 ms était une pause optimale.

L'examen de la littérature sur le vocabulaire a montré que le vocabulaire dans les études plus antérieures a généralement été divisé en deux catégories : passif et actif. Cependant, une description plus détaillée du vocabulaire a été proposée par Henriksen (1999) qui avait présenté un modèle d'acquisition de vocabulaire composé de trois dimensions : une dimension de connaissance partielle-précise, une dimension de profondeur de connaissance et une dimension réceptive-productive. Par exemple, la dimension connaissance partielle-précise explique comment un mot est acquis. Il a été soutenu que le mot n'est pas connu sur la base du « tout ou rien », mais passe par des phases de connaissance partielle du mot. La taxonomie développée plus tard par Nation (2001) a montré ce qu'il faut savoir pour estimer que l'apprenant a une maîtrise complète ou partielle d'un mot. La dimension de profondeur et de taille montre à quel point les mots sont connus en mesurant la profondeur de la connaissance par une approche dimensionnelle ou une approche de développement et combien de mots sont connus en mesurant leur nombre respectivement. Enfin, la dimension réceptive à productive explique qu'il existe une différence entre le vocabulaire réceptif (les mots qu'une personne reconnaît et comprend lorsqu'ils sont rencontrés dans le texte ou le discours mais ne peut pas les produire) et le vocabulaire productif (les mots qu'un apprenant peut produire à l'oral ou à l'écrit). De plus, le vocabulaire productif est de deux sortes : contrôlé (quand un apprenant produit un mot à l'aide d'un stimulus ou d'un contexte) et libre (quand un apprenant produit un mot sans aucun stimulus supplémentaire). L'examen des mesures de vocabulaire a montré que les recherches ont inventé et utilisé une variété de tests

différents. Par exemple, les premiers tests portaient sur le vocabulaire réceptif : Oui/Non tests, Vocabulary Level Test de Nation (1983), Eurocentres Vocabulary size test (Meara et Jones, 1987) tandis que les tests sur le vocabulaire productif ont été développés plus tard : le Lexical Frequency Profile de Laufer et Nation (1995) et le Productive Vocabulary Levels Test développé par Laufer et Nation (1999). Pour les besoins de cette étude, une attention particulière a été portée au test Lex30 de Meara et Fitzpatrick (2000). Lex30 représente une tâche d'association de mots, dans laquelle le candidat se voit présenter une liste de 30 mots stimulus et il lui est demandé de fournir les quatre premiers mots qui lui viennent à l'esprit lorsqu'il pense aux mots stimulus. Un point est attribué pour chaque réponse en dehors des 1 000 premiers mots les plus fréquents, les autres réponses reçoivent 0 point. Le score final peut être représenté par la somme des points peu fréquents et/ou le pourcentage de réponses peu fréquentes. L'apprenant avec un plus grand nombre de points ou un pourcentage plus élevé est considéré comme ayant un vocabulaire productif plus large. Finalement, après la description de la fluence orale et du vocabulaire, les quatre études sur leurs relations (De Jong, Steinel, Florijn, Schoonen & Hulstijn, 2012; Koizumi et In'nami, 2013; Uchihara et Clenton, 2018; Uchihara et Saito, 2016) ont été décrites. Chaque étude a utilisé des mesures différentes mais a trouvé une relation entre les deux compétences langagières.

La dernière partie de la revue de littérature concernait le séjour linguistique. Une fois que les définitions de séjour linguistique (i.e., “combination of immersion in the native speech community, integrated with formal classroom learning” [une combinaison d'immersion dans la communauté linguistique native, intégrée à un apprentissage formel en classe] (Freed, 1995a, p.5)) et des conclusions générales ont été brièvement discutées, les études sur la fluence et le vocabulaire en séjour linguistique ont été examinés en détail. Par exemple, la revue des études de Freed (1995c) et de Segalowitz et Freed (2004) qui étaient basées sur une comparaison des changements de fluence dans le contexte plutôt naturel (le séjour linguistique) et le contexte plutôt formel (la salle de classe) et des études de Freed et al. (2004) et Serrano, Llanes et Tragant (2011) basé sur une comparaison des changements de fluence en séjour linguistique, contexte formel semi-intensif et contexte formel intensif, a montré que le séjour linguistique était plus bénéfique pour le développement de la fluence la fluence que le contexte formel (à l'exception du contexte formel intensif dans Freed et al. (2004)). L'examen des études de Towell, Hawkins et Bazergui (1996) et de Llanes et Muñoz (2009) qui se sont concentrés sur le développement de la fluence au cours de la période de séjour linguistique avec l'aide de différentes mesures, a montré la fiabilité de mesures de la fluence de la vitesse et la fluence de rupture. De plus, la nécessité d'identifier l'emplacement

des pauses en tant que mesures de la fluence a été soulignée dans les études de Mora et Valls-Ferrer (2012), Leonard et Shea (2017) et Huensch et Tracy-Ventura (2017) qui ont également mesuré et trouvé le développement de fluence pendant la période de séjour linguistique. Les conclusions des études sur le vocabulaire dans le contexte de séjour linguistique étaient moins cohérentes. Par exemple, l'étude d'Ife, Vives Boix et Meara (2000) n'ont montré aucune différence dans le développement du vocabulaire au cours de la période de séjour linguistique, tandis que les études de Milton et Meara (1995), Fitzpatrick (2012) et Serrano, Tragant et Llanes (2012) ont signalé le développement du vocabulaire après une période de séjour linguistique. De même, les études de Collentine (2004) et Freed, So et Lazar (2003) qui ont comparé le développement dans le contexte de séjour linguistique et le contexte formel ont indiqué qu'un développement plus grand était trouvé dans le contexte formel, tandis que les études de Dewey (2008) et Jiménez-Jiménez (2010) ont montré les résultats inverses. Enfin, seules deux études sur l'interaction entre vocabulaire et fluence en contexte de séjour linguistique ont été présentées. Les études de Leonard et Shea (2017) et de McManus, Mitchell et Tracy-Ventura (2021) n'ont trouvé que partiellement l'interaction entre les deux compétences langagières.

La deuxième grande partie du mémoire présente l'étude. La présente étude a suivi une conception prétest-posttest qui examine qualitativement les changements dans le vocabulaire productif, la fluence orale et leur interaction en contexte de séjour linguistique dans les entretiens de 5 participants qui ont passé une année universitaire dans un environnement anglophone. L'étude visait à répondre aux questions de recherche suivantes : RQ.1. Dans quelle mesure la fluence orale a-t-elle changé au cours d'une année universitaire dans le contexte de séjour linguistique ? RQ.2. Dans quelle mesure le vocabulaire productif a-t-il changé au cours d'une année scolaire dans le contexte de séjour linguistique ? RQ.3. Les changements de vocabulaire productif semblent-ils montrer des trajectoires de développement parallèles avec une fluence orale au cours d'une année scolaire dans le contexte de séjour linguistique?

Tout d'abord, la méthodologie de l'étude a été présentée. Les participants étaient 5 français, apprenants de l'anglais de niveau intermédiaire à avancé qui ont passé 9 mois en Irlande ou en Angleterre. Les données pour l'analyse de la fluence ont été recueillies au moyen d'un entretien oral avant le départ vers le pays anglophone et après le retour en France. Les données de vocabulaire productif ont été obtenues au moyen de la tâche d'association de mots productifs contrôlée Lex30 avant et après le séjour linguistique. En ce qui concerne le codage des données de fluence, premièrement, les entretiens collectés ont été transcrits

manuellement dans CLAN, deuxièmement, à la suite de De Jong, Groenhout, Schoonen, et Hulstijn (2015) et Huench et Tracy-Ventura (2017) les transcriptions ont été divisées en unités AS et enfin, le codage phonétique a été réalisé en PRAAT. Le codage des données de vocabulaire productif a consisté en la correction des fautes d'orthographe et des erreurs de morphologie dans tous les mots, la suppression des noms propres et la lemmatisation des réponses. En ce qui concerne l'analyse des données, la fluence a été mesurée par trois mesures de fluence de la vitesse (les taux d'articulation avec hésitations, taux d'articulation sans hésitation, et longueur moyenne des énoncés) et quatre mesures de fluence de répartition (nombre de pauses silencieuses au sein de l'ASU, nombre de pauses silencieuses entre les ASU, durée moyenne des pauses silencieuses au sein de l'ASU et durée moyenne des pauses silencieuses entre les ASU), tandis que le vocabulaire productif était représenté par la somme des points peu fréquents (les mots qui ne se trouvaient pas dans les 1 000 mots les plus fréquents) et également le pourcentage de réponses peu fréquentes.

Une fois la méthodologie décrite, les résultats ont été présentés et discutés. Tout d'abord, les résultats de la fluence ont montré un grand développement dans toutes les mesures pour tous les participants, ce qui a permis de répondre à la question de recherche N°1 : la comparaison des entretiens pré-test et post-test dans les 7 mesures de fluence de l'énoncé montre une augmentation de fluence sur une période passée en contexte de séjour linguistique. Deuxièmement, les résultats du vocabulaire productif ont indiqué un léger développement pour 4 (sur 5) participants dans deux mesures, ce qui a permis de répondre à la question de recherche N°2 : la comparaison des entretiens pré-test et post-test dans les deux mesures de notation de Lex30 montre une augmentation de vocabulaire productif sur une période passée dans le contexte de séjour linguistique. Enfin, la comparaison du vocabulaire productif et des résultats de fluence obtenus auprès de cinq participants n'ont trouvé que de légers changements dans le vocabulaire productif tandis que la fluence orale a fait l'objet de changements plus grands au cours d'une année scolaire dans le contexte de séjour linguistique, ce qui a permis de répondre à la question de recherche N°3 : la comparaison des changements de vocabulaire productif et de fluence dans les entretiens post-test ne montre pas de trajectoires de développement parallèles au cours d'une année scolaire dans le contexte de séjour linguistique. Les résultats de cette étude indiquent que plus de recherches utilisant différentes méthodes doivent être effectuées afin de trouver des résultats plus approfondis sur l'interaction entre le vocabulaire productif et fluence dans le contexte de séjour linguistique.

## Contents

|  |    |
|--|----|
| 1. Introduction.....   | 3  |
| 2. Literature Review .....   | 4  |
| 2.1. Fluency.....  | 4  |
| 2.1.1. Definition of Fluency .....                                       | 4  |
| 2.1.2. Measurement of Fluency .....                                      | 7  |
| 2.1.2.1. Speed Fluency.....  | 9  |
| 2.1.2.2. Repair Fluency.....   | 10 |
| 2.1.2.3. Breakdown Fluency. ....   | 10 |
| 2.1.2.3.1. A Unit for all Reasons. ....                                  | 13 |
| 2.1.2.3.2. Pause Threshold.....  | 14 |
| 2.2. Vocabulary .....  | 15 |
| 2.2.1. Definition of Vocabulary .....                                    | 15 |
| 2.2.2. Model of Vocabulary Acquisition.....                              | 16 |
| 2.2.2.1. The Partial-Precise Knowledge Dimension.....                    | 17 |
| 2.2.2.2. Depth vs. Size Dimension. ....                                  | 19 |
| 2.2.2.2.1. Depth.....  | 20 |
| 2.2.2.2.2. Breadth.....  | 21 |
| 2.2.2.3. Receptive to Productive Dimension.....                          | 22 |
| 2.2.3. Measurement of Vocabulary.....                                    | 24 |
| 2.2.3.1. Lex30.....  | 26 |
| 2.2.4. Studies on the Relationships Between Vocabulary and Fluency ..... | 29 |
| 2.3. Study Abroad.....   | 30 |
| 2.3.1. Definitions and General Findings in the SA Context.....           | 31 |
| 2.3.2. Studies on Fluency in the Study Abroad Context .....              | 32 |
| 2.3.3. Studies on Vocabulary in the Study Abroad Context.....            | 36 |



|  |    |
|--|----|
| 2.3.4. Studies on the Interaction Between Vocabulary and Fluency in the Study Abroad Context ..... | 38 |
| 3. The Study.....  | 39 |
| 3.1. Method Section.....   | 40 |
| 3.1.1. Participants.....   | 40 |
| 3.1.2. Data Collection.....  | 41 |
| 3.1.2.1. Fluency. ....   | 41 |
| 3.1.2.2. Vocabulary. ....  | 42 |
| 3.1.3. Data Coding .....   | 42 |
| 3.1.3.1 Fluency. ....  | 42 |
| 3.1.3.2 Vocabulary. ....   | 45 |
| 3.1.4. Data Analysis .....   | 45 |
| 3.1.4.1. Fluency. ....   | 45 |
| 3.1.4.2. Vocabulary. ....  | 46 |
| 4. Results .....   | 47 |
| 4.1. Fluency in the Study Abroad Context.....  | 47 |
| 4.2. Productive Vocabulary in the Study Abroad Context .....                                       | 51 |
| 4.3. Interaction Between Productive Vocabulary and Fluency in the Study Abroad Context .....       | 52 |
| 5. Discussion.....   | 57 |
| 6. Limitations of the Study and Implications for Future Research .....                             | 60 |
| 7. Conclusion .....  | 60 |
| References .....   | 62 |
| Appendix A.....  | 78 |
| Appendix B.....  | 80 |

## 1. Introduction

The research in second language acquisition (SLA) has mainly been conducted on separate constructs and less on the relationships between them. For example, the majority of studies on oral fluency, which is one of the most studied constructs in the SLA, sought to find out how fluency may be developed by focusing only on the various measurements (Llanes & Muñoz, 2009; Towell, Hawkins & Bazergui, 1996) or on a comparison of different contexts (Freed, 1995c; Freed, Segalowitz & Dewey, 2004; Segalowitz & Freed, 2004). Fewer studies have examined whether other constructs may contribute to fluency development, but a number of studies investigating the interaction between formulaic sequences and fluency, found out the development in formulaic sequences leads to the development in fluency (Boers, Eyckmans, Kappel, Stengers, & Demecheleer, 2006; Boers and Lindstromberg, 2012; Chen, 2019; Cordier, 2013; McGuire & Larson-Hall, 2017; Wood, 2006, 2009). The lack of studies on the relationships between fluency and other constructs was underlined by Foster (2020). She argued that the research on fluency's relationship with other constructs may reflect how fluency changes because of the changes in the other constructs, which may help the researchers and learners to understand how fluency can be developed by the improvement of these other construct/s (i.e., the relationship between vocabulary and reading indicates that one needs to read more in order to enlarge the vocabulary size (Qian, 2002; Van Gelderen, Schoonen, De Glopper, Hulstijn, Simis, Snellings, & Stevenson 2004)). Moreover, Foster (2020) suggested a research agenda on fluency relationships with other constructs for the next decade. The present thesis, based on one of the Foster's (2020) research tasks suggesting studying "the relationship between productive vocabulary size and fluency performance measures" (p. 8), will examine the interaction between productive vocabulary and fluency.

In order to study the interaction between productive vocabulary and fluency, one needs to create the same conditions for possible changes in both constructs. The SA context was chosen to study and compare the changes in fluency and productive vocabulary because it allows more natural language acquisition opportunities. Moreover, a number of the studies on fluency and productive vocabulary showed that the SA context contributes to their development (see sections 2.3.2 and 2.3.3, respectively) which also means that it may be useful to study their interaction in the SA context. Therefore the present thesis aims to examine the changes in productive vocabulary, fluency and their interaction in the SA context.

The paper's thesis is organized in the following way: section 2 presents an overview of the literature. Section 3 presents the research questions and explains methodology. Section

4 presents and describes the obtained results. Section 5 provides the discussion of the obtained results. Section 6 presents the limitations of the study and suggests ideas for further research. The conclusions are presented in Section 7.

## **2. Literature Review**

For the purposes of this study, the literature review will be divided into three parts: the first one will present the literature on fluency, the second one will focus on the literature on vocabulary, and the third one will explore the literature on research in the SA context.

### **2.1. Fluency**

Fluency is one of the most important components of speaking ability. Every learner of a foreign language seeks to become a fluent speaker. According to Freed (1995b), fluency is one of the most studied topics in first and second language acquisition. Nevertheless, fluency has never been a simple construct to define and measure. Therefore, the various definition of fluency found in the literature will be presented in this section and then the various methods of the fluency measurement will be described.

#### ***2.1.1. Definition of Fluency***

Fluency is a “complex phenomenon” (Freed et al., 2004, p. 279) to define. According to Lennon (1990), it may be understood in two different senses: broad and narrow. Fluency in a broad sense refers to global speaking proficiency (i.e., someone may be fluent in a foreign language in general), while fluency in a narrow sense is referred as a component of oral proficiency which is used for its measurement (i.e., someone may speak fluently which means only the smoothness and rapidity in speaking ability).

One of the most prominent fluency definitions in the broad sense was provided by Fillmore (1979). He suggested four different dimensions of fluency, which taken together define fluency in terms of global oral proficiency. First of all, fluency was defined as “the ability to talk at length with few pauses, the ability to fill time with talk” (p. 93). This definition underlines the ability to talk with fluidity in speech and without hesitation (more interestingly, this definition, when treated not as one of four dimensions, but as a separate one may be considered as a definition in a narrow sense, since it refers only to the smooth speaking abilities). Secondly, fluency was defined in terms of the correct use of language as “the ability to talk in coherent, reasoned, and semantically dense sentences (p. 93) which should show “a mastery of the semantic and syntactic resources of the language” (p. 93). The third dimension of fluency emphasized the ability to have and to apply the appropriate knowledge “in a wide range of contexts” (p. 93), meaning that the speaker should always

have something appropriate to say in any situation. The last aspect of fluency dealt with “the ability some people have to be creative and imaginative in their language use” (p. 93), which allows people “to express their ideas in novel ways, to pun, to make up jokes, to attend to the sound independently of the sense, to vary styles, to create and build on metaphors” (p. 93). According to Fillmore (1979), a speaker who has all four abilities may be referred to as fluent. However, Riazantseva (2001) argued that “fluency in this sense encompasses the notion of accuracy or grammatical correctness in language use that manifests itself in adherence to the phonological, syntactic, semantic, prosodic, and pragmatic rules of the language” (p.499) and therefore fluency in this broad sense refers to global speaking proficiency since it is “unclear how this conceptualization differs from the definition of global oral proficiency” (Kormos, 2006, p. 155).

The definition of fluency in the narrow sense (which we are concerned with in this paper) is smoothness and rapidity of speech which are closely associated with nativelikeness (Lennon, 1990). One of the first fluency definitions in the narrow sense was proposed by Pawley and Syder (1983) who defined “native-like fluency” as “the native speaker’s ability to produce fluent stretches of discourse” (p. 191). Richards, Platt and Weber (1985) also defined fluency with an emphasis on the smoothness and native-like use of the speech: “the features which give speech the qualities of being natural and normal, including native-like use of pausing, rhythm, intonation, stress, rate of speaking, and use of interjections and interruptions” (p. 108). Rehbein (1987) tackled fluency from a different, more procedural perspective. He defined it in terms of two stages of fluency production: “fluency means that the activities of [1] planning and [2] uttering can be executed nearly simultaneously by the speaker of the language” (p.104) and also highlighting the importance of the “[3] speaker’s evaluation of the hearer’s expectations” (p. 104). This definition was later developed by Lennon (1990) who also made an emphasis on three salient processes: “[3] an impression on the listener’s part that the psycholinguistic processes of [1]speech planning and [2]speech production are functioning easily and efficiently” (p.391) and defined fluency as a “performance phenomenon” for the listener: “fluency reflects the speaker’s ability to focus the listener’s attention on his or her message by presenting a finished product, rather than inviting the listener to focus on the working of the production mechanisms” (pp. 391–392). Although, these definitions were helpful in understanding the phenomenon of fluency, it was difficult to apply them in practice, therefore a practical general definition of fluency in the narrow sense was proposed by Lennon (2000): “a working definition of fluency might be the

rapid, smooth, accurate, lucid, and efficient translation of thought or communicative intention into language under the temporal constraints of on-line processing” (p. 26).

Chambers (1997) argued that an investigation of fluency development of foreign language learners and an investigation of the differences in fluency between them would lead towards better understanding of fluency phenomena. For example, Brown (2003) stated that “fluency is probably not an absolute characteristic that students either have or do not have. If, in fact, fluency is a matter of degrees, students at any level of proficiency can probably achieve some degree of fluency” (p. 7), however, fluency is not an easy skill and “may take the adult learner of a foreign language years to achieve” (Pawley & Syder, 1983, p. 199) which results in differences between the learners. Leo Tolstoy (1878) wrote, “Все счастливые семьи похожи друг на друга, каждая несчастливая семья несчастлива по-своему” [All happy families are alike; each unhappy family is unhappy in its own way] (p. 16), in a similar way Riggenbach (1991) suggested that all fluent learners “resemble each other” (p. 439) while non-fluent ones may differ in various ways. The possible explanation of differences in fluency among non-native speakers were proposed by two approaches: the Individual Differences approach and the Universalist approach. The Individual Differences approach, supported by Kormos (1999) and Skehan (2002), proposed that the limitations in individual learner’s abilities (i.e., less developed memorization skill or the lack of learning strategies) may cause the differences among non-native learners. While the Universalist Approach supported by Towell et al. (1996), DeKeyser (1997), and Segalowitz (2000), in contrast, stated that limitations in the automatization of the encoding processes and production of speech (i.e., in order to automatize the speech and be more fluent one needs more practice in oral production) may cause differences among non-native learners.

Nevertheless, all these definitions and approaches were considered as not encompassing all aspects of fluency and not providing a reliable fluency measurement, by Segalowitz (2010). He further developed the definitions of Rehbein (1987) and Lennon (1990) which focused on 3 components of fluency, by categorizing the phenomena into cognitive fluency, utterance fluency, and perceived fluency. Segalowitz (2010) suggested that each of these dimensions is responsible for one particular aspect of fluency. For example, Goldman-Eisler (1968) was the first researcher to argue that fluency depends on speech planning in one’s mind: “the complete speech act is a dynamic process, demanding the mobilization in proper sequence of a series of complex procedures and is the temporal integration of serial phenomena” (p.6). This process of speech planning (i.e., [1] for Rehbein (1987) and Lennon (1990)) was referred as cognitive fluency by Segalowitz (2010) and

defined as “the efficiency of operation of the underlying processes responsible for the production of utterances” (p.156). He argued that this is the construct of fluency which plays a role first, since one needs to use her or his cognitive abilities to organize and plan what to say before speech production or perception take place. The second construct of fluency which is responsible for speech production was named utterance fluency (i.e., [2] for Rehbein (1987) and Lennon (1990)) and defined as “the features of utterances that reflect the speaker’s cognitive fluency” (Segalowitz, 2010, p.165). Perceived fluency (i.e., [3] for Rehbein (1987) and Lennon (1990)) is the final construct of fluency which refers to the acceptability of the two previous constructs by the interlocutor: “the inferences listeners make about speakers’ cognitive fluency based on their perceptions of their utterance fluency”. Segalowitz (2010) argued that when working on fluency one should use these three constructs “that are reliable indicators of how efficiently a speaker is able to mobilize and temporally integrate, in a nearly simultaneous way, the underlying processes of planning and assembling an utterance in order to perform a communicatively acceptable speech act” (p.165).

### ***2.1.2. Measurement of Fluency***

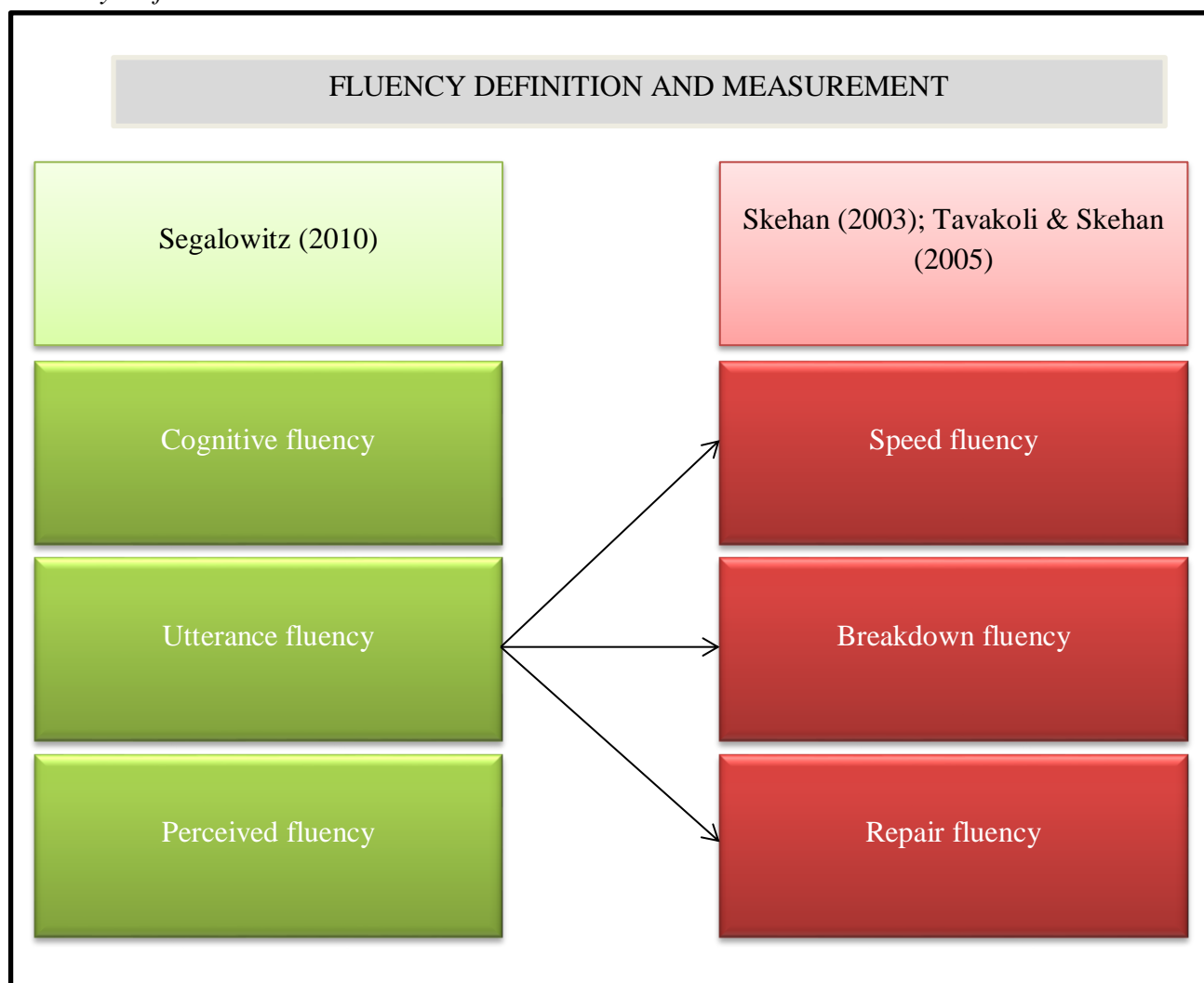
However, although Segalowitz (2010) encouraged researchers to look at the three constructs of fluency together, in practice there are some concerns about their measurement. For instance, when it comes to cognitive fluency, there is no reliable instrument to track the changes in the thinking processes in one’s mind when planning what to say. Therefore, fluency has been measured either on “speakers’ productions [utterance fluency] (Raupach, 1980; Towell, Hawkins & Bazergui, 1996; Riggensbach, 2000; Freed, Segalowitz & Dewey, 2004; [...]), or to a lesser extent, on listeners’ perceptions [perceived fluency], using judgments from raters (Ejzenberg, 1992; Trofimovich & Baker, 2006)” (Valls-Ferrer, 2011, p.66). However, it has been argued that measurement of perceived fluency may be subjective, because the perception of speech by judges with different backgrounds (individual differences) may highly influence the results (Bosker, Pinget, Quené, Sanders & De Jong, 2012; Kahng, 2014; Segalowitz, 2010, 2016). The measurement of utterance fluency, in contrast, was considered to be a more objective one (Skehan, 2003, 2009; Tavakoli & Skehan 2005), because it is based on the strict measures which are independent from human perception and thus guarantee more objectivity. Moreover, the measurement of utterance fluency may also “provide a window into cognitive fluency [...] [by enabling] researchers to gather valuable empirical evidence on psycholinguistic mechanisms at stake in speech

production since processes of language production themselves are not directly accessible to observation (Chambers, 1997)” (Cordier, 2013, p.40).

But which measures should be used in order to better measure utterance fluency? Skehan (2003) and Tavakoli and Skehan (2005) proposed 3 crucial aspects for the utterance fluency measurement which are presented in Figure 1.

**Figure 1**

*Fluency Definition and Measurement*



They argued that each of these aspects should allow the measurement of some aspect of utterance fluency, for example, speed fluency is responsible for measuring the speed, flow and density of the produced speech, breakdown fluency measures pauses and hesitation phenomena, while repair fluency measures reformulations, corrections and repetitions. For the purposes of the present study, these 3 aspects and their measurements will be presented in detail in the following subsections.

**2.1.2.1. Speed Fluency.** One of the major differences between native and non-native speakers characterized as the “fluency gap” (Segalowitz, 2010, p. 2) was argued to consist in the speed of speech which is generally slower for the non-natives (Raupach, 1984; Wiese, 1984). In order to measure the differences or longitudinal changes in speed fluency one should use the temporal variables which provide “objective measurements of the output of the productions which must lie behind language processing (Towell, 2002, p.119). However, there are too many various measures that one may use. For instance, García-Amaya (2018, pp.182–183) presented 11 possible temporal measures that were previously used in different studies: Total speaking-time duration, mean length of run in seconds, mean length of run in syllables, phonation time, speech rate, adjusted (or pruned) speech rate, articulation rate, phonation-time ratio, number of syllables, number of pruned syllables, mean syllable duration. However not all of them are reliable and necessary for speed fluency measurement. In order to find out the reliable temporal variables, the researchers used one of three ways:

They have compared speech from fluent and nonfluent speakers (Ejzenberg, 2000; Riazantseva, 2001; Riggenbach, 1991; Tavakoli, 2011), investigated the longitudinal development of fluency (Derwing, Munro, & Thomson, 2007; Freed, 1995, 2000; Lennon, 1990, Mora & Valls-Ferrer, 2012; Towell, Hawkins, & Bazergui, 1996; Wood, 2010), and related utterance fluency to perceived fluency by correlating fluency ratings with temporal variables (Bosker et al., 2013; Cucchiarini, Strik, & Boves, 2002; Derwing, Rossiter, Munro, & Thomson, 2004; Fulcher, 1996; Kormos & Denes, 2004; Rossiter, 2009). (Kahng, 2014, p.811)

There were only a few temporal measures which were found reliable by most of the studies. For the purpose of this study, only three such measures will be discussed in detail. For example, the speech rate (SR), measured as syllables uttered per minute/second, including pause time, has been found to be one of the most reliable temporal measures (Ejzenberg, 2000; Freed, 1995c, 2000; Kahng, 2014; Kormos & Dénes, 2004; Lennon, 1990; O’Brien, Segalowitz, Freed, & Collentine, 2007; Riggenbach, 1991; Segalowitz & Freed, 2004; Towell et al., 1996). Also it was the only measure which was able to capture the changes in all three aspects of fluency (because it includes speech (utterance fluency) which contains repairs (repair fluency) and pauses (breakdown fluency)). Despite the fact that SR was found to be a reliable measure, Kormos and Dénes (2004) argued that it could not represent speed fluency alone. The second temporal measure, the mean length of runs (MLoR), which corresponds to the average length (in syllables) of a run, where runs are strings of speech between pauses, was also considered reliable by most of the studies (Ejzenberg, 2000; Freed, 1995c, 2000;



Kormos & Dénes, 2004; Lennon, 1990; O'Brien et al., 2007; Riggenbach, 1991; Segalowitz & Freed, 2004; Skehan, 2009; Towell et al., 1996; Towell & Dewaele, 2005) and along with SR were considered as composite measures of utterance fluency. The last measure, the articulation rate (AR) indicates the number of syllables articulated within a given period of phonation ((AR) and reflects the changes “of proceduralization within the articulator” (Towell et al., 1996, p. 92). AR is usually measured as the number of syllables per minute or second, excluding pause time. Some researchers (Cucchiari, Strik & Boves, 2002; Kormos & Dénes, 2004) doubted its usefulness; however others (including Kormos herself in later studies) argued it to be also one of the most reliable measures (De Jong, Steinel, Florijn, Schoonen & Hulstijn, 2012; Ginther, Dimova, & Yang, 2010; Mora & Valls-Ferrer, 2012; Préfontaine, Kormos & Johnson, 2016; Suzuki & Kormos, 2019; Towell et al., 1996).

Researchers have found out more contradicting results on the other measures. For example, Lennon, (1990) and Freed (1995c) opted for phonological rate, while De Jong et al. (2012) and De Jong, Groenhout, Schoonen, and Hulstijn, (2015) argued that mean syllable duration (inverse articulation rate) should be used. The controversies in choosing reliable and informative measures were summed up by Kormos (2006). She stated that any of these measures may be used depending on the methodology and the goals of a study, and suggested that one does not need to choose all the measures because they may overlap with each other.

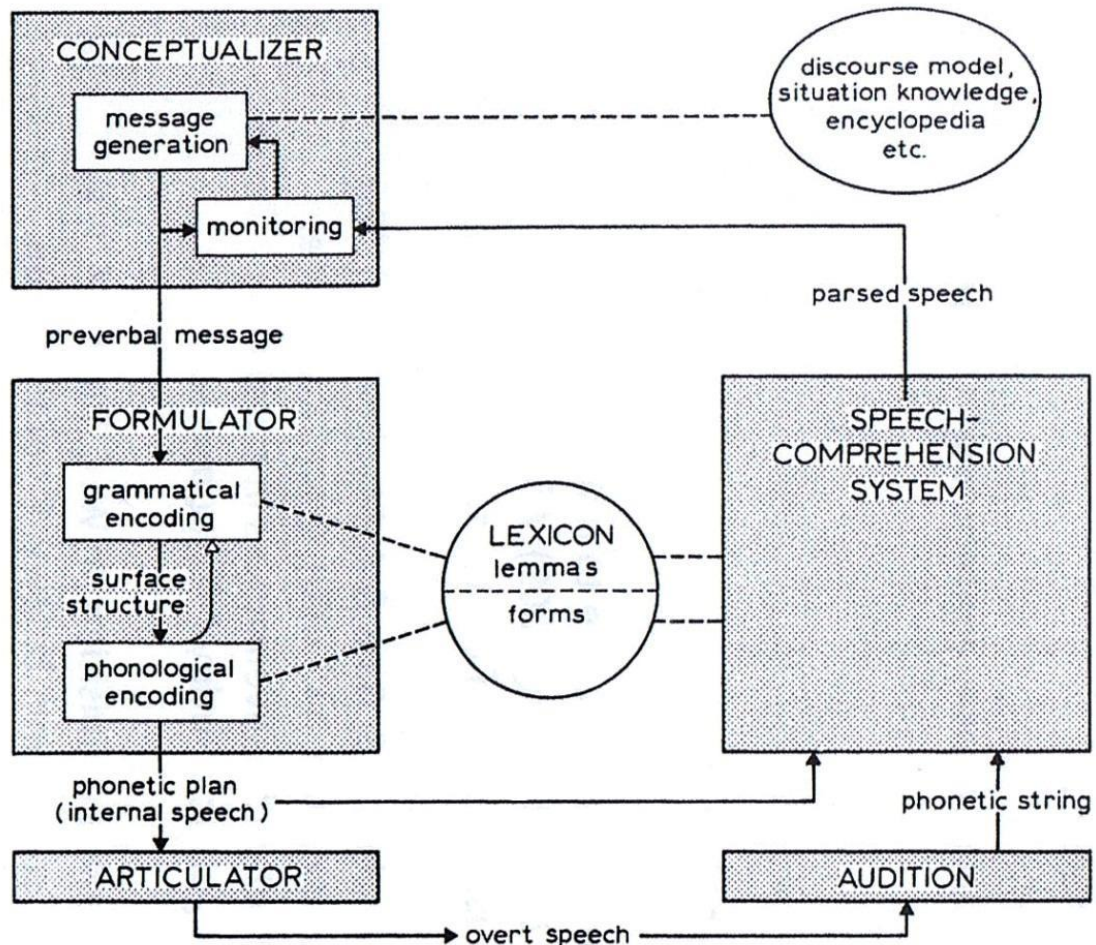
**2.1.2.2. Repair Fluency.** According to Skehan (2003) and Tavakoli and Skehan (2005), repair fluency relates to number and length of the repetitions, hesitations, false starts replacements and reformulations “that are used to repair speech during the production process” (Tavakoli & Wright 2020, p.47). The differences in repair fluency were hypothesized to show the difference between native and non-native speakers (i.e., number of reformulations was hypothesized to be larger for non-native speakers). However, the research on repair fluency (Baker-Smemoe, Dewey, Bown, & Martinsen, 2014; Kormos, 2006; Kormos & Dénes, 2004; Tavakoli & Skehan, 2005) has shown that repair fluency does not necessarily differentiate native and non-native speakers and may rather be affected by individual differences (i.e, personality, native language, memory capacity and L2 learning experience (Baker-Smemoe et al., 2014, Kahng, 2014)). Moreover, Ellis and Barkhuizen (2005) argued that repair fluency “indicates the extent to which the learner is oriented towards accuracy” (pp. 149–150), more so than indicating the degree of fluency. Therefore, repair fluency may not be regarded as a reliable measurement of fluency.

**2.1.2.3. Breakdown Fluency.** Breakdown fluency reflects the amount of pausing in speech production. Chafe (1985) argued that pausing in native speaker’s speech is

Good evidence that speaking is not a matter of regurgitating material already stored in the mind in linguistic form, but it is a creative act, relating two media, thought and language, which are not isomorphic but require adjustments and readjustments to each other. (as cited in Chambers, 1997, p. 538)

However, pauses may not only be good evidence of a “creative act” but also occur “as the result of psycholinguistic or cognitive processes” (Cordier, 2013, p.64). Chambers (1997) argued that the presence of pauses in speech is one of the “normal features of the interaction” (p. 538), but at the same time it may signal the “poor functioning of mental processes” (p. 538), especially for non-native speakers. The comparison of native and non-native speakers allowed the researchers to argue that they differ in the use of pauses, for example, native speakers were using fewer pauses which were also shorter while non-natives were using more pauses which also were longer (Hilton, 2008, 2014; Skehan and Foster, 2008). In order to measure the pause and hesitation phenomena, the two types of pauses (silent and filled) were identified and tested for measuring pause frequency and length in relation to fluency. In general, the number (frequency) of silent pauses and their length were found to positively correlate with fluency scores (Cucchiaroni et al., 2002; Foster & Skehan, 1999; Ginther et al., 2010; Kormos & Dénes, 2004; Lennon, 1990; Préfontaine et al., 2016) while the studies on filled pauses showed less or no correlation at all (Lennon, 1990; Foster & Skehan, 1999; Bosker et al., 2012).

However, the measurement of fluency with frequency and length of silent pauses only does not allow us to show the differences in the psycholinguistic and cognitive processes, while the measurement of pause location does (Lennon, 1984, 1990; Pawley & Syder, 2000; Towell et al., 1996; Riggensbach, 1991). It has been argued that more fluent speakers pause more at clause boundaries (Lennon, 1990; Towell et al., 1996) while less fluent ones make more pauses within clauses (Freed, 1995c; Riggensbach, 1991). Pausing within clauses (“clause” here is used as a generalized term, pausing was also measured in various kinds of units (see section 2.1.2.3.1. A Unit for all Reasons, for more details) was argued to reflect the problems with speech planning. The explanation of difficulties with speech planning has been based on Levelt’s (1989) three-stage model of speaking (Figure 2).

**Figure 2***Levelt's Model of Speech Production*

Note: Reprinted from *Speaking: From intention to articulation* (p.9), by W. Levelt, 1989, MIT Press.

Levelt's model consists of 3 stages: conceptualizer, formulator, and articulator. At the conceptualization stage, the process of speech planning and generation of a pre-verbal message takes place. The formulation stage is responsible for encoding thoughts into a grammatically and phonologically acceptable message. The final, articulation stage converts the planned message into speech. When all stages are carefully completed fluent speech production should take place.

Based on Levelt's model, it was suggested that pauses at clause boundaries are connected to the conceptualization stage which means that they occur as a result of planning of the next clause, or searching for new ideas for discussion (macro-planning) while pauses within clauses are related to the formulation and/or articulation stages which means that they

occur as a result of searching for certain words (lexical retrieval), grammatical forms or spelling (micro-planning) (Skehan, Foster, & Shum, 2016). Recent studies (Hilton, 2008, 2014; Kahng, 2014; Skehan & Foster, 2008; Skehan et al., 2016; Tavakoli, 2011; Tavakoli, Nakatsuhara, & Hunter, 2020) having compared the pause location of native and non-native speakers, all shown that natives paused more between clauses whereas non-natives paused more within clauses, and concluded that the measurement of pause location reflects processing difficulties in speech production and may be used to measure fluency.

**2.1.2.3.1. A Unit for all Reasons.** Measuring pause location in oral speech production has not uniquely used clauses as a unit for analysis; researchers have segmented oral productions into various units of speech for better analysis. Foster, Tonkyn, and Wigglesworth (2000) argued that there was no unique commonly established unit that was used by the researchers, and thus all units which were used varied from researcher to researcher and were based on different criteria. They analyzed 87 units and grouped them due to their definition criteria into three categories: mainly semantic, mainly intonational, and mainly syntactic. The idea behind mainly semantic units (i.e., C-unit (Pica, Holliday, Lewis, & Morgenthaler, 1989), Idea unit (Kroll, 1977)) consisted in dividing speech into units which conveyed ideas. In other words, one idea equals one unit, as for example one of mainly semantic units called C-unit was defined as “utterances, for example, words, phrases and sentences, grammatical and ungrammatical, which provide referential or pragmatic meaning” (Pica et al., 1989, p.72). When it comes to mainly intonational units (i.e., Idea unit (Chafe, 1980), Tone unit (Crystal & Davy, 1975), their identification was based on listening to intonation (rises and falls) of a speaker which marks the boundaries of a unit, as for the Tone unit which was based on “distinctive configuration of pitches, with a clear centre, or nucleus. [...]The nucleus is the syllable (or, in some cases, series of syllables) which carries the greatest prominence within the tone-unit” (Crystal & Davy, 1975, p.16). Finally, mainly syntactic units are based on simple syntactic elements. For example, the sentence (Quirk, Greenbaum, Leech, & Svartvik, 1985) or a clause (Kroll, 1977) may be used as a unit for analysis. However the most popular among mainly syntactic units (and also among all three categories of units) was the T-unit (Hunt, 1970) which was more complex in comparison to other syntactic measures: “a main clause plus any other clauses which are dependent upon it” (Foster et al., 2000, p. 360).

Nevertheless, despite the various methods of unit identification, all three categories were strongly criticized by Foster et al. (2000). For example, all mainly semantic units were criticized for not providing clear instructions of idea identification and sometimes for

impossibility of such identification. Similarly, the intonational units were also considered not applicable as it was not always possible to segment the intonation of non-native speakers into such units. The main criticism of mainly syntactic units consisted in the limitations of applying such units in spoken data which “is not nearly so tidy and clear cut” (p.262) as a written one and which does not always allow for division into simple clauses since the oral production of a speaker may go beyond the level of clause, because it may contain words, combinations of words, reformulations and other features of oral speech which cannot be classified as a clause.

Having criticized the existing units, Foster et al. (2000) suggested a new “unit for all reasons” (p.372) called the AS-unit (henceforth ASU). An ASU is a mainly syntactic unit defined as a “single speaker's utterance consisting of an independent clause, or sub-clausal unit, together with any subordinate clause(s) associated with either” (p.365). It was based on the T-unit (Hunt, 1970) and developed “to deal with the features characteristic of spoken data” (p.365). Moreover, Foster et al. (2000) provided detailed instructions for ASU identification (see section 3.1.3.1 Fluency). ASU had been used by various researchers and was argued to be one of the best units for oral data analysis (De Jong, 2016; De Jong et al., 2015; Huensch & Tracy-Ventura, 2017; Kahng, 2014, Tavakoli 2011).

**2.1.2.3.2. Pause Threshold.** Another crucial element for breakdown fluency analysis consists in deciding of what can be counted as a pause. An optimal pause for the fluency measurement should not capture the natural short physical pauses (i.e., swallowing, breathing pauses (Zellner, 1994)), but the pauses which occur as a result of cognitive process (macro/micro-planning pauses). The thresholds from 100ms to 400ms have been the most popular among researchers, each supporting their own one. For instance, Freed (2000) argued that “only those unfilled pauses [(.4 a second or larger)]... [are] heard as dysfluent” (p.248), while Towell et al. (1996) opted for 300ms because “anything less than 0.3 seconds is easily confused in a spectrogram with other speech phenomena such as the stop phase of a plosive sound, and anything longer can omit significant pause phenomena” (p.46).

However, De Jong and Bosker (2013) argued that none of the previous studies has deliberately focused on the pause thresholds and compared them to establish the most optimal pause. Therefore, De Jong and Bosker (2013) compared the correlation between cut-off points from 50 to 400 milliseconds with fluency scores. The results showed that the threshold should be set at least at 250ms and at 300ms at most, while the highest correlation was obtained with 250ms which was established as an optimal threshold and recommended for use when measuring fluency.

## 2.2. Vocabulary

According to Meara (1980), vocabulary has been considered as a “neglected aspect of language learning” (p.1) for a long period of time, however nowadays more and more researchers contribute their studies to research in that direction. In order to study vocabulary, I will present relevant definitions first, then a model of vocabulary acquisition will be presented and finally different vocabulary measurements will be described. Additionally, for the purposes of the current study, the studies on the relationship between vocabulary and fluency will be presented in the last subsection.

### 2.2.1. Definition of Vocabulary

The importance of vocabulary knowledge has always been underlined by researchers, writers and philosophers, for example, the American journalist and philosopher Henry Hazlitt (1993) wrote:

A man with a scant vocabulary will almost certainly be a weak thinker. The richer and more copious one’s vocabulary and the greater one’s awareness of fine distinctions and subtle nuances of meaning, the more fertile and precise is likely to be one’s thinking. Knowledge of things and knowledge of the words for them grow together. If you do not know the words, you can hardly know the thing. (p.51)

But what does “knowing the words” mean? Actually, “knowing the word” is one of the most used definitions of vocabulary. For example, the Cambridge dictionary (n.d) defined “vocabulary” as “all the words known and used by a particular person”, a similar definition was suggested by the MacMillan (n.d) dictionary: “all the words that a person knows”. However, just knowing words is not as crucial as the ability to use them when they are needed; as Laufer and Nation (1995) suggested, “Vocabulary is not usually learned for its own sake. An important aim of a vocabulary program is to bring learners’ vocabulary knowledge into communicative use” (p.308). The importance of vocabulary has been recognized in all four language skills (reading, writing, listening and speaking), and has even been placed in a higher rank than grammar, at first by Wilkins (1972): “Without grammar very little can be conveyed, without vocabulary nothing can be conveyed.” (p.111), and later by Schmitt (2010) who made an illustrative observation on this topic: “Learners carry around dictionaries and not grammar books” (p.4). Our digital age showed the same trend in underling the superiority of vocabulary:

The importance of vocabulary knowledge in foreign language learning has often been illustrated with reference to learners taking dictionaries and not grammar books, especially when they go abroad. Although this observation may seem outdated with

the advent of digital age and global communication technologies, dictionaries and online translators are the most popular reported applications for language learning on students' mobile devices, according to a survey by Simon and Fell (2012). (Zaytseva, Pérez-Vidal & Miralpeix, 2018, p.210)

Despite the salience of vocabulary development in first language acquisition and second language acquisition, the studies on vocabulary massively appeared only in the 1980s–1990s, starting with an article by Meara (1980), called “Vocabulary acquisition: A neglected aspect of language learning” which drew the attention of many other researchers to this field. Since then, vocabulary has gained in weight and according to Schmitt (2014), has become one of the major topics in SLA.

The concept of vocabulary is a complex one: “While we are gaining an increasing understanding of the development of some isolated aspects of vocabulary, the overall acquisition system is far too complex and variable for us to comprehend in its entirety, and so it still eludes description” (Schmitt, 2010, p.36) . There is no unique study which covers all aspects of vocabulary, therefore, there is a need to look at a large number of studies in order to fully explore the theme of vocabulary.

### **2.2.2. Model of Vocabulary Acquisition**

In a well-known article by Bialystok and Sharwood-Smith (1985), the authors established two basic dimensions of language proficiency: “knowledge” and “control”. In other words, they distinguished between “competence” and “performance”, “passive” and “active”, “receptive” and “productive”, “declarative” and “procedural”. They also suggested that these dimensions are interrelated using “the library metaphor” (p.105), proposing that a human's mind is like a library where a great number of books on different subjects are arranged in a certain system and stored (“knowledge” that a person gained through life) while the human should know where books on each subject are stored and be able to easily and quickly find and use the book on the subject that is needed for a particular situation (“control” over knowledge, being able to use the knowledge in practice). This “passive-active continuum” (Laufer & Paribakht, 1998, p.367) was largely used in the field of vocabulary in 1980s–1990s, as it was the most evident and useful distinction that allowed for a new wave of studies. However, this distinction was not enough to cover all dimensions of vocabulary and some researchers started to develop new more advanced theories. One of the main studies was conducted by Henriksen (1999), who developed the “passive-active continuum”, and established a Model of Vocabulary Acquisition (Henriksen, 1999). She described the 3 new dimensions which aimed to explain the process of vocabulary acquisition at different levels:

I see a need for being more specific and suggest three dimensions as a balanced position between the global and the separate trait view. I propose that we distinguish among three separate but related vocabulary dimensions: (a) a “partial-precise knowledge” dimension, (b) a “depth of knowledge” dimension, and (c) a “receptive-productive” dimension. (p.305)

Each of these dimensions is crucial and central to the analysis of vocabulary acquisition, and therefore each one will be presented and analyzed separately in detail in the following subsections.

**2.2.2.1. The Partial-Precise Knowledge Dimension.** Before dividing vocabulary into types and categories, one should at first understand the basic process of vocabulary acquisition, that is to say, how a word is acquired. At the beginning of the research on vocabulary “the criterion for knowing the word has been quite liberal” (Read, 1998, p.16), certain researchers checked knowledge of a word “as the ability to translate the lexical item into the L1, to find the right definition in a multiple-choice task, or to paraphrase in the target language” (Henriksen, 1999, p.305). However, all these formats (including Yes/NO tests) were criticized for providing only two possibilities – knowing and not knowing, thus representing “vocabulary knowledge as precise comprehension” (p.305), which was not the case. According to Harley (1995), “learners do not know a word on an ‘all-or nothing’ basis, but go through phases of ‘partial word knowledge.’” (p.3). This idea was developed by Henriksen (1999) suggesting that there are some stages in acquiring a word:

In the process of acquiring word meaning, the learner’s knowledge of a certain lexical item moves from mere word recognition (i.e., acknowledging that the word exists in the target language) through different degrees of partial knowledge (Brown, 1994) toward precise comprehension. (p.312)

Although the first attempts to describe the degrees from partial to precise knowledge were made even earlier by Richards (1976), Brown (1994) and Miller (1996), they had not fully described all aspects. For instance, Miller (1996) argued that knowing a word means knowing “its own sound, its own spelling, its own meaning, its own role, its own use, its own history” (p.5). More detailed studies were performed by Laufer (1990, 1993) who proposed her own “taxonomy of components of word knowledge, consisting of form (phonological, graphic, morphological), syntactic behavior, meaning (referential, associative, pragmatic) and relations with other words (paradigmatic and syntagmatic)” (Laufer & Paribakht, 1998, p.368) and Nation (1990) who suggested 4 main types of knowledge (form, position, function



and meaning). However the most popular, useful and detailed taxonomy was done by Nation in 2001, which is presented in Table 1.

**Table 1**

*What is involved in knowing a word?*

|         |                        |   |   |
|---------|------------------------|---|---|
| Form    | Spoken                 | R | What does the word sound like?                          |
|         |                        | P | How is the word pronounced?                             |
|         | Written                | R | What does the word look like?                           |
|         |                        | P | How is the word written and spelled?                    |
|         | Word parts             | R | What parts are recognizable in this word?               |
|         |                        | P | What word parts are needed to express meaning?          |
| Meaning | Form and meaning       | R | What meaning does this word form signal?                |
|         |                        | P | What word form can be used to express this meaning?     |
|         | Concepts and referents | R | What is included in the concept?                        |
|         |                        | P | What items can the concept refer to?                    |
|         | Associations           | R | What other words does this word make us think of?       |
|         |                        | P | What other words could we use instead of this one?      |
| Use     | Grammatical functions  | R | In what patterns does the word occur?                   |
|         |                        | P | In what patterns must we use this word?                 |
|         | Collocations           | R | What words or types of words occur with this one?       |
|         |                        | P | What words or types of words must we use with this one? |
|         | Constraints on use     | R | Where, when and how often would we meet this word?      |
|         |                        | P | Where, when and how often can we use this word?         |

Note: R = Receptive vocabulary knowledge; P = Productive vocabulary knowledge

Adapted from *Learning vocabulary in another language* (p.27), by I. S. P. Nation, 2001, Cambridge University Press.

Nation's taxonomy shows what one needs to know in order to say that he/she has a complete command of a word. However, even native speakers may not know all the presented categories of knowing a word for all words they are using, they "are likely to have only

mastery of a limited number of word knowledge categories for a large percentage of words in their lexicon” (Schmitt & Meara, 1997, p.18). The taxonomy was created not to show that native speakers have full command of a word and non-native speakers most likely not, but to show that there are different stages of word acquisition and that it is not “knowing” or “not knowing” dichotomy but rather a process of interconnected incremental process of a word acquisition. A person may know only one category but with time develop and acquire more categories thus partially discovering different aspects of the word and getting a more complete command of a word (Schmitt, 1998). Nation’s taxonomy confirms Henriksen’s (1999) idea that vocabulary acquisition is not an all-or-nothing process but rather “from zero to partial to precise” one (Schmitt, 2010, p.21), and what is more important, it allows us “to study the incremental acquisition of individual words” (Schmitt, 1998, p.283) in terms of depth, or in other words, how well these words are known.

**2.2.2.2. Depth vs. Size Dimension.** The depth of knowledge is the second dimension in Henriksen’s (1999) model, however for the purposes of the present thesis depth will be analyzed in connection with breadth, as two opposing but at the same time interrelated concepts. The concept of depth means the quality of vocabulary knowledge while the concept of breadth means quantity of vocabulary knowledge or more simply vocabulary size. In other words, depth means “how well words are known” and breadth means “how many words are known” (Anderson & Freebody, 1981, pp.92–93). Although those concepts seem to go in the completely opposite directions, they are deeply interrelated: “The unidimensionality tests, involving comparisons of different pairs of person measures derived from separate analyses of the different strength modalities, confirm that size and strength are related constructs” (Laufer, Elder, Hill, & Congdon, 2004, p.222). According to Schmitt (2014), “the size–depth relationship depends on various factors such as the size of the learner’s lexicon, the frequency level of the target words measured, and the learner’s L1” (p.941). He argues that there is a weaker relationship between breadth and depth for higher frequency words and for someone who has a small vocabulary, and a greater relationship for low frequency words and for someone who has a large vocabulary size. At the same time, Read (2000) doubted the usefulness of depth in vocabulary representation and underlined the one of breadth: “Despite the fact that the [size] tests may seem superficial, they can give a more representative picture of the overall state of the learner’s vocabulary than an in-depth probe of a limited number of words” (p.18). Laufer et al. (2004) also supported this statement arguing that “knowing many words is more important than knowing few words in depth. Hence, a good vocabulary test should test how many words are known, or try to provide a picture of the learner’s overall

vocabulary” (p.209). However, Tseng and Schmitt (2008) demonstrated that both breadth and depth are important components of lexical knowledge. Ishii and Schmitt (2009) used a battery of 4 size-depth vocabulary tests and concluded that depth of knowledge is a key component of overall lexical competency, and thus should be included in lexical assessment along with vocabulary size.” (p.19). According to Schmitt (2014), “Depth typically adds unique explanatory power compared to size alone.” (p.913), moreover he argued that large vocabulary size “by itself is insufficient, as learners need to know words well in order to use them productively, appropriately, and fluently.” (p.942), therefore both concepts are worthy of attention.

**2.2.2.2.1. Depth.** Cronbach (1942) was among the first researchers who underlined that knowing just what the word means is insufficient, and one also needs to know “paradigmatic (antonymy, synonymy, hyponymy, gradation) and syntagmatic relations (collocational restrictions)” (Henriksen, 1999, pp.305–306). Read (2004) also argued that learners should have more knowledge about a word: “They should develop a rich and specific meaning representation as well as knowledge of the word’s format features, syntactic functioning, collocational possibilities, register characteristics, and so on” (p. 155). Ishii and Schmitt (2009) argued that not all lexical problems are “caused by a small vocabulary [breadth]” (p.6), some are caused by insufficient depth knowledge, for example by “limited knowledge of secondary meaning senses” (p.6), or “limited awareness of the different derivative forms of a word (e.g., silly, silliness)” (p.6). In general there are two approaches to the dimension of depth:

Some researchers (Richards, 1976; Ringbom, 1987; Nation, 1990; 2001) claim that knowing a word involves a range of inter-related ‘subknowledges’ such as morphological and grammatical knowledge and knowledge of word meanings. Others assume that lexical knowledge consists of progressive levels of knowledge, starting with a superficial familiarity with the word and ending with the ability to use the word correctly in free production (Faerch et al., 1984; Palmberg, 1987). (Laufer et al., 2004 p.203)

These two directions were described by Schmitt (2010) as dimension and developmental approaches.

A dimension approach involves specifying “some of the types of word knowledge one can have about lexical items, and then quantifying participants’ mastery of those types.” (p.224). In other words, it consists of separate components or inter-related subknowledges as in the above-mentioned studies (1976), Ringbom (1987) and Nation (1990; 2001). For

example, Nation's taxonomy (Table 1) may be used in order to measure how well the word is known, counting how many categories the learner knows.

The developmental approach implies incremental word development along the following scale: "From no knowledge at all to complete mastery" (Schmitt, 2010, p.16). For example, the Vocabulary Knowledge Scale is one of the most popular developmental scales which consists of 5 stages:

- I. I don't remember having seen this word before.
- II. I have seen this word before, but I don't know what it means.
- III. I have seen this word before, and I think it means ——. (synonym or translation)
- IV. I know this word. It means ——. (synonym or translation)
- V. I can use this word in a sentence: ——. (Write a sentence.) (If you do this section, please also do Section IV). (pp. 218–219)

At the beginning, a learner starts with not knowing a word at all, then having only some ideas about it but not knowing it. In the third stage one is able to produce some basic meaning of a word, and in the next a learner knows the word well and is ready to use it. In the final stage one knows how to use the word in a sentence with "semantic appropriateness and grammatical accuracy" (Fitzpatrick & Clenton, 2017, p.859) This approach was reported to be especially helpful at the early development of word knowledge (Schmitt 2010).

Nevertheless, even having both of the approaches, one cannot fully capture and measure the development of vocabulary depth and the research in this direction still continues (Schmitt 2010).

**2.2.2.2.2. Breadth.** It is much easier to conceptualize breadth than depth, as breadth of vocabulary or vocabulary size "is basically counting known lexical items" (Schmitt, 2014, p.915). When talking about counting, one needs to specify what exactly is counted. In general, most researchers argued that counting words may not be the best idea when measuring vocabulary size. For instance, a learner who knows a simple word "look" may be able to produce a lot more similar words with the same root (i.e "looking", "looked", "overlooked"), because the "mind stores only the base form of a lemma and then attaches inflectional suffixes" (Schmitt, 2010, p.189). Therefore, it has been argued that lemmas was "probably the best unit overall" (Schmitt, 2010, P.193) because in that way one may exclude the repetition of words with the same base.

It has been widely assumed that a large vocabulary is associated with better general language knowledge. For example Meara (1996) argued that the learners who have larger vocabularies "are more proficient in a wide range of language skills than learners with

smaller vocabularies, and there is some evidence to support the view that vocabulary skills make a significant contribution to almost all aspects of L2 proficiency” (p.37). Therefore, the following question arises: how large a vocabulary is needed? According to the studies by Goulden, Nation and Read, (1990), D’Anna, Zechmeister, and Hall (1991) and Schmitt (2010), a vocabulary ranging between 15,000 and 20,000 word families is established as a norm for an educated English native speaker. However, when it comes to non-native-speakers, a number of the researchers (Hirsh & Nation, 1992; Hu & Nation, 2000; Laufer, 1992; Nation, 2006; Schmitt, 2008) suggested that one needs to know around 2,000 – 3,000 word families in order to be able to understand 95% of everyday spoken English, and around 6,000 – 7,000 word families to understand 98% of spoken English. Nation (2006) put an emphasis on the difference between high and low-frequency words, and argued that non-native speakers should learn high-frequency words in the very first place since they are the most used ones and constitute the mentioned 95%. Needless to mention that knowing word families is not knowing words: “Each word family includes several individual word forms, including the root form (stimulate), its inflections (stimulated, stimulating, stimulates), and regular derivations (stimulation, stimulative)” (Schmitt, 2010, p.8), thus, 6,000 word families may entail around 28, 000 words (Schmitt, 2010).

**2.2.2.3. Receptive to Productive Dimension.** The last but not the least third “receptive to productive” dimension is the most widely known one, not only by researchers but also by ordinary learners. Almost any learner may say that their vocabulary may be divided into two groups: the words which they recognize and understand when they are met in text or someone’s speech but cannot produce them themselves (which is an example of receptive vocabulary) and words which the learners can produce in speech or writing (which is an example of productive vocabulary). There is a typical saying by learners with big receptive and small productive vocabularies: “I am like a dog. I understand everything that is said, but I can’t say anything in response”. The researchers have delved deeper into the distinction between receptive and productive knowledge suggesting that a word is known productively only when one understands how to use it appropriately in a given context and not just being able to produce it:

In our opinion, mere memorization of a word form in a given context without understanding the word's meaning cannot be called productive knowledge. If the learner can repeat the memorized word with its context in a test situation without understanding it, this is mechanical reproduction, not production. (Laufer, 1998, p.257)

Moreover, Laufer (1998) went even further and distinguished the two types of productive vocabulary: controlled and free. When a learner is producing a word with help of some stimulus or context (for instance by providing synonyms, antonyms or words that are associated with the given words in the task or by filling a gap in a sentence) it is called controlled productive knowledge, whereas when a learner is producing a word without any additional stimulus and “at one’s free will” (He, 2019, p.130) (for instance, when one is asked to write a text or to speak for a certain amount of time) it is called free productive knowledge. However, Meara and Fitzpatrick (2000) argued that, in fact, the free production may be also limited to certain stimuli. For instance, when learners are asked to write an essay which implies the use of free productive vocabulary, they may be limited to the topic of the essay (for instance, if the topic is about football, it will limit and direct the free productive knowledge towards football-related words) as well as the topic of oral conversation at an exam may influence the kind of words that will be used by a speaker (see section 2.2.3. Measurement of Vocabulary, for further analysis of controlled and free productive knowledge).

The question of the relationships between productive and receptive vocabulary has also been raised by researchers: “The relationship between an L2 learner’s passive and active vocabularies remains interesting but unexplored; statements about this relationship have been vague and unsubstantiated” (Laufer & Paribakht, 1998, p.369). Melka-Teichroew (1997) stated that it had not been discovered yet when exactly receptive knowledge becomes productive: “At what point familiarity is such that one could say that knowledge is no longer receptive, but is productive, or at which point receptive knowledge can be converted into productive knowledge.” (p.313). The “point” is yet to be established, however it was suggested that a word moves from receptive to productive vocabulary through certain degrees of word knowledge (as discussed in section 2.2.2.2.1. Depth).

Not only has the relationship been explored by researchers but also the difference in size of receptive and productive vocabularies. It has been assumed that receptive vocabulary is a way larger than the active one (Aitchison, 1989; Channell, 1988), however, the studies which actually checked the difference show contradicting results. Takala (1984) suggested that the difference is rather small, claiming that about 90% of receptive vocabulary is known productively. Laufer and Paribakht (1998) showed lower scores of 62% to 77% and similar results were provided by Fan (2000) who found 53% to 81% of common receptive-productive knowledge. In contrast to previous studies, Laufer (2005) found a great difference between two vocabularies: the results showed that only 16% to 35% of passive vocabulary is

known productively. The difference between receptive and productive results provided by these studies does not allow for establishing the concrete level that shows how larger passive is over active, as highlighted by Schmitt (2014), “the inconsistency of these figures highlights the difficulties and confusion involved in dealing with the receptive–productive issue.” (p. 923); and deserve more attention from the researchers’ side. Nevertheless, in all these studies receptive knowledge was definitely found to be larger, moreover certain researchers (Laufer, 1998; Laufer & Paribakht, 1998) studied how the difference between the two vocabularies changed when new words were acquired and found that “an increase in one’s passive vocabulary will, on the one hand, lead to an increase in one’s controlled active vocabulary, but at the same time lead to a larger gap between the two” (Laufer, 1998, p.267). Therefore, it was argued that the difference between receptive and productive vocabularies may be much larger for the learner with greater vocabularies.

### ***2.2.3. Measurement of Vocabulary***

The need for the vocabulary measurement reappeared along with the studies on vocabulary in the 1980s–1990s (Meara, 1987). Vocabulary was commonly measured by means of various vocabulary tests, which in most cases measured vocabulary size. One of the most popular tests at that time was Eurocentres Vocabulary size test invented by Meara and Jones (1987). The test was originally designed as a student placement test for Eurocentres language schools in the UK, and was constructed as a Yes/No checklist format which was designed to measure absolute size of one’s vocabulary. The idea of using “the absolute simplest” (Read, 1993, p.355) Yes/No test format was not necessarily new, as according to Melka-Teichroew (1982, p.7), the tradition of using Yes/No tests may go back as far as 1890. However, traditional Yes/No tests were not considered reliable and trustworthy due to their format (Meara, 1992). The traditional Yes/No test presents a learner with a list of words; the learner in turn should indicate whether they recognize them (Yes) or not (No). The main concern with such tests was the fact that the learners may not always be honest and mark words that are unknown as known: “Unfortunately, people aren’t 100% reliable, or 100% honest” (Meara, 1992, p.9). The solution to this problem was found by Anderson and Freebody (1983) who suggested using “nonwords” characterized by Meara and Jones (1988) as “imaginary words which are very carefully constructed so that they share the physical characteristics of the real words in the same set.” (p.85), and which were used in order to calculate the probability of the learner’s dishonest answer and excluding it from results analysis. This solution was applied by Meara and Jones (1987) in their Eurocentres test, and later developed by Meara (1990) in the same test by including 1 non-existing word for every

2 existing. Moreover, Meara (1990) computerized the test; the computer provided the person who was tested with the words from each thousand of the most used 10,000 English words and non-existing words and automatically checked for the reliability of test takers.

The other test which had an immense influence on the development of vocabulary measurement was Vocabulary Level Test (VLT) by Paul Nation (1983). VLT assesses vocabulary size of the learner by measuring the knowledge of words at four frequency levels: 2,000 words, 3,000 words, 5,000 and 10,000 words. The test is organized in a word-definition matching format, where one needs to match one of 36 words (at each level) to 18 proposed definitions. Similar to Eurocentres Test, VLT was designed for “diagnostic testing” (Read, 1988, p.17) which “can be utilized by teachers and administrators in a pedagogical context to inform decisions concerning whether an examinee is likely to have the lexical resources necessary to cope with certain language tasks” (Schmitt N., Schmitt D. & Clapham, 2001, p.56). However, in contrast to the Eurocentres test, it was found to be more reliable and valid (Beglar, 2009; Nation & Beglar, 2007; Schmitt et al., 2001) and argued to be one of the best tests of that time: “Nearest thing we have to a standard test in vocabulary” (Meara, 1996, p.38). Moreover, the first version of VLT by Nation (1983) had given a rise to its various iterations, the most successful of which was made by Schmitt et al. (2001) who developed the test and used it in a study with 801 participants with different L1 languages (the original version was not supposed to be used for learners with Romance L1s). This iteration was also found valid and reliable and nowadays is still considered to be a useful tool of vocabulary measurement: “The Vocabulary levels test (Schmitt et al., 2001) is still a well-used standard vocabulary measurement, yet the authors have not updated it at all since it was launched over 17 years ago” (Schmitt, Nation & Kremmel, 2019, p.110); even despite the appearance of the more recent versions (i.e. New Vocabulary Levels Test (McLean & Kramer, 2015) and Updated Vocabulary Levels Test (Webb, Sasao, & Ballance, 2017). Nevertheless, there is no ideal vocabulary test: “No vocabulary test (or any language test) can be suitable for all learners in all contexts” (Schmitt et al., 2019, p.111). The above-mentioned tests (and many others) has been classified as passive vocabulary ones which measure only passive vocabulary size, whereas it was argued that for measuring such skills as oral fluency which is a productive skill the application of productive vocabulary tests, “which focus on vocabulary use instead of vocabulary knowledge” (González & Píriz, 2016, p.34), are considered to provide more representative results (De Jong, 2016). Therefore, for the purposes of the present study, one should take a closer look at the existing productive vocabulary tests.



The productive vocabulary tests are of two sorts: controlled and free. “Controlled indicates that the test is designed to elicit specific, predetermined vocabulary items, and free indicates that vocabulary produced by the test taker in a relatively unconstrained task will be measured” (Fitzpatrick & Clenton, 2017, p.846). The example of a free productive test is Lexical Frequency Profile (LFP) designed by Laufer and Nation (1995). When taking the LFP test, the learner is asked to produce a written response (essay) of 300 words to a question or topic. The produced words are then categorized according to four frequency categories: words from 1000 of most frequent words, words from 2000, words from the University Word List and words which are not on any of the other lists. The researchers suggested that the learner who has a larger vocabulary would use more words from lower-frequency category while one who has a smaller vocabulary would rely on using more higher-frequency words. Moreover, the test was designed for a test-retest procedure which also allows one to measure potential changes in one’s vocabulary over time.

An example of a controlled productive test is the Productive Vocabulary Levels Test (PVLТ) developed by Laufer and Nation (1999) on the basis of VLT. The test requires the learners to fill in a gap in a sentence with one word, knowing only its initial letter. The learners should understand the given context to be able to retrieve the word they are asked for. Moreover, the words are classified according to the same frequency bands as for the VLT, thus learners with larger vocabularies are expected to provide more correct answers for lower-frequency words.

The main difficulty for researchers who want to use a vocabulary test is the choice of the right one: “But which one to use?” (Schmitt et al., 2019, p.110). The above-presented tests may be described in a much more detailed way and also there are a lot of other tests which may be described here. However, as suggested by Schmitt et al. (2019, p.111), one should choose and focus on the test which would suit the purpose of one’s study. Therefore, for the purposes of the present study only the Lex30 test will be presented and subjected to particular scrutiny.

**2.2.3.1. Lex30.** Lex30 was designed by Meara and Fitzpatrick (2000) as a response to the existing productive tests of that time. For instance, Meara and Fitzpatrick (2000) criticized LFP for being “effective mainly at low levels; when, for example, testees are expected to have a limited vocabulary size” (p.21) and not being effective at high levels because it may be difficult to interpret the results on the basis of only 18 examples per frequency band. At the same time, the free productive tests like PVLТ were criticized for being context-limited (mainly because of a topic which limits and directs towards the use of

topic-related words), but even more importantly, the learners in such tests had a tendency to produce many more highly frequent words and fewer low frequent ones, which were considered to be the only “true indicators of a large productive vocabulary” (Meara & Olmos, 2010, p. 223). “Therefore, Meara and Fitzpatrick (2000) proposed their own test which aimed to exclude the above-mentioned weaknesses and take only the best from both types of productive tests: “Lex30 aims to combine the conceptual advantages of a context-independent test of ‘free productive knowledge’ (Laufer, 1998) with the scoring and administrative advantages of a quantifiable measuring tool (Fitzpatrick, 2007, p. 119).

Lex30 represents a word association task, where the test taker is presented with the list of 30 stimulus words and asked to provide the first four words (or three words in earlier versions) that come to their mind when they think of the stimuli words. The 30 stimulus words (cues) should be selected following the three criteria: “1. All the stimulus words are highly frequent — In our experiment, the words were taken from Nation’s first 1000 wordlist (Nation, 1984) [...] 2. None of the stimulus words typically elicits a single, dominant primary response [...] 3. Each of the stimulus words typically generates responses which are not common words” (Meara & Fitzpatrick, 2000, pp. 22–23). The responses to the stimuli (120 max = 30 cues x 4 responses) are lemmatized and subjected to frequency analysis. One point is given for each response (except for proper nouns) outside the first 1,000 most frequent words, the other responses are not given anything or given 0 points. The final score may be represented by the sum of the “infrequent” points and/or the percentage of infrequent responses as suggested by, Fitzpatrick and Meara (2004) and Fitzpatrick (2007), because “it minimizes the influence of varying corpus size derived from the Lex30 task” (Uchihara & Saito, 2017). The one with a larger number of points or higher percentage is considered to have a larger productive vocabulary.

Meara and Fitzpatrick (2000) argued that their test was much better than the existing tests, for example, Lex30 “uses 30 prompts, each activating a different semantic field” (Fitzpatrick & Clenton, 2017, p.861) in comparison to LFP which narrows down towards one direction; or produces more low frequent responses in comparison to PVLTL, which allows for a more precise vocabulary measurement. Moreover, it is relatively simple in preparation and easy to administer requiring only 15 minutes to complete the task (which allows for not losing a learner’s attention) and includes the best features of both controlled and free tests:

There is no predetermined set of response target words for the subject to produce, and in this way, Lex30 resembles a free productive task. However, the stimulus words

tend to impose some constraints on the responses, and Lex30 thus shares some of the advantages of context-limited productive tests. (Meara & Fitzpatrick, 2000, p.22)

However, Baba (2002) argued that despite the advantages of the test, Meara and Fitzpatrick (2000) have not confirmed the validity and reliability of the test which are the real markers of its adequacy and appropriateness for measuring productive vocabulary. Baba (2002) in the same paper found Lex30 to be a reliable test and encouraged other researchers to test it: “It would perhaps be beneficial to use other approaches to estimate Lex30’s reliability such as the test-retest approach or the parallel forms approach” (p. 69). Although no test may be considered as completely validated (Fulcher & Davidson, 2007), the results from 11 studies summed up by González and Píriz (2016, p.35-36), showed that Lex30 was found to be both valid and reliable, when tested with different methods (Test-retest, Parallel forms, Comparing performance on different tests, Comparing NS and NNS scores), different numbers of test takers with different language backgrounds.

Baba (2002) also hypothesized that there may be a significant difference in results between oral and written versions of Lex30. He argued that the use of a written version “assesses the learners’ written performances but does not assess their spoken vocabulary knowledge” (p.70), because a learner may know the word but have difficulties in writing it down, thus limiting the number of words which could have been said aloud when taking an oral version but are lost in writing when using a written one. Clenton (2010) devoted a separate study to that problem. He compared the written and spoken versions of Lex30 with the same cues and found no significant difference in the results:

An examination of the scores attained on the spoken and the written test formats of Lex30 reveals no significant difference between the two. Both elicited broadly similar scores, so we may claim that Lex30 seems to elicit spoken and written productive vocabulary knowledge in broadly the same way. Thus, Lex30 appears to tap subjects’ productive vocabulary knowledge regardless of the task format (spoken or written).

p.114

Therefore one may use a written version (which should be more easy to administer) of a test for measuring both written and spoken productive vocabulary.

Finally, Lex30 may be quite flexible in terms of corpora. One may use different word lists or corpora as the basis for the test, for example Clenton (2010) tested 3 version of Lex30 using 3 different set of cues selected from 3 data lists (Lexorig, JClk and JC2k) and found no difference in the results: “Mean scores from each of the three sets of cues (Lexorig, JClk, and JC2k) were similar, [which] suggests that regardless of the cues we test with, Lex30 appears

to elicit similar scores that elicit similar proportions of infrequent vocabulary items” (Clenton, 2010, p.134).

#### ***2.2.4. Studies on the Relationships Between Vocabulary and Fluency***

The study by De Jong et al. (2012) was among the first to explore the relationships between vocabulary and speaking proficiency, although only partially, since it was not their only and primary goal (they were also measuring 5 other aspects). The study was carried out on 181 advanced Dutch learners, using controlled productive PVLТ (Laufer & Nation 1999) for the vocabulary measurement and 8 role-play speaking tasks (judged by native speakers) for the measurement of speaking proficiency. As a result, productive vocabulary knowledge was found to be a good predictor of speaking proficiency, but no detailed discussion on that finding was provided.

A more detailed study was performed by Koizumi and In'nami (2013). The researchers also measured productive knowledge, but through the means of L1-L2 translations, while a 15 minute monologue (which included the introduction of the learner, description and comparison of several pictures) was chosen to represent speaking proficiency, which was coded using ASU (Foster et al., 2000) and conceptualized as consisting of fluency (measured by speed and repair), accuracy, and syntactic complexity. The participants were 224 intermediate English learners. The findings suggested that vocabulary knowledge contributes to all three above-mentioned concepts of speaking proficiency. Koizumi & In'nami (2013) went further and tried to explain why vocabulary knowledge and fluency may be connected with each other:

Vocabulary knowledge and fluency may be associated with each other because L2 learners with larger and deeper vocabulary knowledge, and faster access to it, can perform lexical searches more easily and quickly. Learners with greater vocabulary knowledge can recall adequate words and use them for speaking through knowledge of antonyms and collocations (Aitchison, 2003). Consequently, processing will be smoother for them than for those with a smaller lexicon, although the speed of intermediate-level learners with greater vocabulary knowledge is still much slower than for high-proficiency learners, and their processing is far from automatic. On the other hand, learners with poorer vocabulary knowledge may not be able to find appropriate words, or may take longer to search for words at the formulation stage, resulting in reduced speed fluency. (p.911)

Uchihara and Clenton (2018) decided to conduct a study on vocabulary knowledge and speaking ability using a receptive vocabulary test. The updated version of “Yes/No” test

was used to measure vocabulary and an oral picture narrative judged by native speakers was used to measure oral abilities. The study was carried out with 46 advanced English learners. The results, similar to the previous studies (although they used productive tests), showed that vocabulary knowledge “predicts L2 oral ability for advanced learners albeit to a lesser extent than when compared to studies with low proficiency learners (Koizumi & In’nami, 2013)” (p.552).

Finally, the study by Uchihara and Saito (2016) was based on 39 English learners of different degrees (advanced, intermediate and beginners). The vocabulary measurement was conducted through the means of productive Lex30 test while a picture description task was used for eliciting spontaneous speech, which was analyzed for fluency (using only speech rate) comprehensibility and accentedness by natives. The results suggested that vocabulary knowledge “significantly correlated with L2 fluency, but not with comprehensibility or accentedness” (p.1). Moreover, the researchers argued that the observed correlation between vocabulary knowledge and fluency “might indicate that more proficient L2 learners, as indicated by their productive vocabulary scores, might be able to speak spontaneously without too many pauses and repetitions, and at a faster tempo.” (p.2). Uchihara and Saito (2016) also argued that their methods might be improved by using more detailed data, for example, by “employing longer speech samples (> 3 min), [in] future studies of this kind” (p.9) and underlined the need for further research in this direction.

### **2.3. Study Abroad**

Historically, studying abroad has always been an integral part of the educational process, although it was not always accessible for all. For example, in the 17<sup>th</sup>-19<sup>th</sup> centuries, it was accessible only for a privileged class of powerful and flourishing countries. More precisely, the young aristocrats at the end of their studies in their native country took a journey called the Grand Tour that “typically involved three or four years of travel around Europe and included an extensive sojourn in Italy [(cultural center of that time)]” (Encyclopædia Britannica. (n.d.)). Nowadays, studying abroad is no longer considered to be a limited to the very privileged and almost every student may participate in various exchange programs. For example, in the last three decades, around 10 million participants took part in the most popular program in Europe, called “Erasmus+” which allows students (and other people associated with educational institutions i.e., teachers, apprentices, etc.) to experience a period abroad for educational purposes and supports them financially (Erasmus+, 2018). For instance, in 2018 “with a budget of €2.8 billion euros, Erasmus+ supported more than

850,000 mobilities” (Erasmus+, 2018, p.7). Moreover, according to the Organisation for Economic Co-operation (*OECD*, 2017), around 5 million students per year experienced an SA around the globe from 2010 to 2015; while the number of students that will take part in the exchange programs should rise to 8 million participants per year by 2025.

The SA experience is extremely attractive for students due to the five reasons stated by themselves:

le désir de vivre une expérience à l'étranger, celui d'apprendre une langue étrangère, celui de nouer de nouvelles relations sociales, celui de développer la capacité d'adaptabilité et la prise de perspective et, enfin, celui d'améliorer les perspectives de carrière (Commission Européenne, 2014). (Arvidsson, 2019, pp.5–6)

However, students are not the only ones who have become interested in SA. The rise of students' interest in SA opened a vast field for research which resulted in the rise of interest for researchers. Firstly, I will present the definitions and general findings in SA studies, and then more detailed studies on fluency and vocabulary in the SA contexts will be discussed.

### ***2.3.1. Definitions and General Findings in the SA Context***

SA for researchers constitutes a particular interest as a context of learning. The importance and influence of a learning context was underlined by Llanes (2011):

The learning context plays a decisive role for several reasons, and it is one of the crucial variables in becoming bilingual/multilingual: the quality and quantity of the input, the opportunities the learners have to practice the L2, and the type of instruction in the L2 all vary according to the context of learning. (p.189)

According to Freed (1995a), SA in the 1990s, defined as a “combination of immersion in the native speech community, integrated with formal classroom learning” (p.5) was assumed to be “the best environment [context] for learning a second language” (p.5). Segalowitz and Freed (2004) argued that the SA context should be more beneficial for language learners mainly because of being fully immersed in the target language, having more opportunities to interact with native speakers and use a target language (TL) every day. The SA was generally considered as “a unique learning context for evaluating changes” (Leonard & Shea, 2017, p.179) and raised a lot of questions for researchers to explore:

What precisely is gained linguistically from living in the country? Which language skills? To what levels of proficiency? What kinds of experience and programs in-country are the most effective in building students' language skills? What is the minimal duration an effective program must have? How much language must a

student have in order to take maximum advantage of the experience? (Brecht Davidson & Ginsberg, 1995, p.38)

According to Collentine (2009), SA research may be divided into two periods: before Freed's (1995b) publication of "Second language acquisition in a study abroad context" and after. The period before publication mainly encompasses the studies which examined the SA too broadly "with instruments that sought to assess learners' overall L2 abilities" (Collentine, 2009, p.219). For example, Carroll (1967) found that even short three-month SA contributes to the development of linguistic skills such as metalinguistic knowledge, Opper, Teichler and Carlson (1990) used learners' self-reported assessments and found development in all 4 main language skills (listening, reading, speaking, and writing), while Krashen and Seliger (1976) using teacher rankings did not find SA to be necessarily beneficial for the learners. The second wave of studies was caused by Freed's (1995b) book, where having summarized previous research on the SA, she stated that "numerous questions remain to be answered by carefully-controlled empirical studies" (p.16) and encouraged other researchers to study the SA context. The following studies tested the changes in various language skills of learners over the SA context, for instance the biggest development was found in oral fluency (Llanes & Muñoz, 2009; Mora & Valls-Ferrer, 2012; Towell, et al.,1996) and lexical skills (Foster & Tavakoli, 2009; Milton & Meara, 1995; Ife, Vives Boix & Meara, 2000). The positive impact of the SA context was also found, to a lesser extent, in reading (Lapkin, Hart & Swain, 1995), listening (Evans & Fisher, 2005) and writing (Sasaki, 2004). In order to establish which of these language skills were more likely to be better developed particularly in the SA context, researchers compared the SA context with at home (AH) context. The results of such studies indicated that the SA may not necessarily be better than AH for the development of listening (Cubillos, Chieffo & Fan 2008), reading (Dewey, 2004) and writing (Freed, So & Lazar, 2003) skills or for grammar (Collentine, 2004), but might be better for lexical skills (Dewey, 2008; Fitzpatrick, 2012; Foster, 2009; Jiménez-Jiménez, 2010) and especially for oral fluency development (Freed, 1995c; Segalowitz & Freed, 2004; Segalowitz, Freed, Collentine, Lafford, Lazar, & Díaz-Campos, 2004)(see sections 2.3.2 and 2.3.3. for the detailed description of studies on vocabulary and fluency (respectively) in the SA context).

### ***2.3.2. Studies on Fluency in the Study Abroad Context***

In order to demonstrate that the changes in learner's fluency were the result of SA impact, a number of researchers measured fluency using different types of measurements, means of data collection, numbers of participants with different language level and background, and also compared the changes in SA to changes in other contexts. Since the

discussion of all such studies is not the goal of this paper, only the most well-known key studies which are relevant to the present study will be discussed below.

To begin with, the study by Freed (1995c) was based on a comparison of changes in fluency after 1 semester in the SA and AH contexts. The data from thirty English-speaking learners of French (from novice to intermediate) were collected through the means of Oral Proficiency Interview (OPI) at the beginning and at the end of the semester. At first, fluency was assessed by means of six trained native speakers of French who marked learners “from 1 (‘not at all fluent’) to 7 (‘extremely fluent’)” (p.129) on the basis of four forty-five second speech segments from the interviews of each learner. The results of judges’ assessment were summarized by Freed (1995c) in the following way: “Students who have lived and studied abroad were found to speak more and at a significantly faster rate” (p.141). Additionally, fluency was measured by more objective linguistic measurements: SR, frequency of unfilled pauses, frequency of filled pauses, length of fluent speech runs, frequency of repairs. However, in that type of analysis, only SR was found to have developed after SA experience.

Towell, et al. (1996) performed a longitudinal study which compared 12 advanced English-speaking learners of French before and after six months in the SA context. To collect the data, the participants were shown a film and asked to retell it orally. The retellings were analyzed in terms of SR, AR, MLoR and phonation/time ratio (PTR) to track changes in fluency. The results in AR and PTR generally showed insignificant changes towards fluency development, while SR and especially MLoR significantly increased over the SA experience and were argued to represent the development in fluency in general.

The year 2004 was extremely fruitful in fluency research in the SA context. For example, Segalowitz and Freed (2004) investigated the role of learning context on fluency in SA and AH settings. Participants were 40 advanced English-speaking learners of Spanish who spent a semester in one of the contexts. Four minutes from pretest and posttest interviews (OPI) were selected for analysis of temporal and hesitation measures of fluency (SR, MLoR without silent pauses (400ms), MLoR without filled pauses, and longest fluent run (without silent and filled pauses)). The results indicated that the SA context was found to be better than AH for fluency development since the learners from the SA context showed significantly greater gains in three fluency measures: SR, MLoR without filled pauses, and longest fluent run. In the same year a similar study was conducted by the same two authors and two other ones (Collentine, Lafford, Lazar & Díaz-Campos). Segalowitz et al. (2004), using the same methods and measures, based their study on 46 advanced English-speaking



learners of Spanish who spent 1 semester either in the SA or in AH contexts. They found similar results in greater fluency gains in the SA context.

Freed et al. (2004) went further and compared the development of fluency in three different learning contexts: SA, AH and IM (immersion program). The study focused on 28 advanced university students of French who spent 7 (IM) or 12 (SA and AH) weeks in one of the contexts. Data was collected by means of oral interviews before and after each of the above-mentioned learning contexts. Freed et al. (2004) decided to analyze fluency on the basis of 2 minutes from pretest and 2 minutes from posttest interviews, using 9 measures (SR, hesitation-free speech runs, filler-free speech runs, fluent runs, repetition-free speech runs, grammatical-repair-free speech runs, total words spoken, duration of speaking time), however only a few of them were found to be helpful. For example, the advantage of the SA context over AH was found only in three measures: SR, filler-free speech runs, and fluent runs, thus SA was argued to be a better context for fluency development. However, the IM showed much better results in comparison to both SA and AT in total words spoken, SR, hesitation-free and repetition-free speech and was argued to be a much better context for fluency development than the other two. Similar results were demonstrated by Serrano et al. (2011) who also compared three contexts: the learners from intensive AH had greater fluency gains than SA and semi-intensive AH, while the learners from the SA context had greater fluency gains in comparison to semi-intensive AH. However, their study was limited to the use of only one fluency measure (SR).

Llanes and Muñoz (2009) studied fluency development after a short SA experience. They investigated fluency development in twenty-four Catalan/Spanish-speaking students of English that participated in a 3–4 week SA program. The data collection consisted of 10–15 minute interviews which began with the question/answer part and led to a description of a series of 6 pictures (“The Picnic Story” (Heaton, 1966)). The results indicated the significant development in 4 measures (SR, AR, other language word ratio, and longest fluent run) out of 6 (not in filled pauses per minute and silent pauses per minute); therefore it was argued that short SA experience may also be beneficial for fluency development.

The lack of stable pause measures reflecting fluency development (as in e.g. above-mentioned Llanes and Muñoz (2009)) was highlighted by Mora and Valls-Ferrer (2012) who suggested that pauses should not be measured per minute only but their location within clauses should also be considered. For example, in their study on fluency development in the SA context, apart from the usual temporal measures (SR, AR, MLoR, PTR, dysfluency ratio) they also used two measures which considered pause location: pause frequency (number of

clause-internal pauses per minute) and pause time ratio (percentage of clause-internal pause time with respect to total time). The study was based on 30 upper intermediate/advanced Catalan/Spanish-speaking learners of English who spent three months in the SA context. The guided interviews in pairs which consisted of participant-participant and participant-interviewer interactions were used to obtain the data for analysis. The results were argued to provide strong evidence for the positive impact of the SA period on the development of fluency, by indicating the increase of SR, AR, MLoR and decrease of pause frequency (less pauses) and pause time ratio (shorter pauses).

The importance of pause location and the pause measures which reflect fluency in general was even more explicitly studied by Leonard and Shea (2017). Their study was conducted on 39 English-speaking learners of Spanish who spent three months in an SA context. Oral data were gathered via 3 two minutes monologues in Spanish (and also in English), while only 30 seconds of each monologue were used for analysis. In order to measure fluency only two speed fluency measures were used: AR and MLoR, while breakdown fluency was measured by rate of: all pauses, mid-clause pauses, end-of-clause pauses, long pauses, short pauses, unfilled pauses, filled pauses and percent of pausing time. The finding showed that the SA period led to positive changes in two speed measures and also in rate of: all pauses, mid-clause pauses, long pauses, unfilled pauses, filled pauses and percent of pausing time that reflected fluency development. Moreover, Leonard and Shea (2017) compared the use of mid-clause and end-clause pauses in Spanish and English monologues, and stated that the participants stopped far more in mid-clauses in their L2 in pretest monologues which indicates the problems with formulation and articulation processes (Hilton, 2008, 2014). Additionally, in posttest monologues participants' "pattern of mid-clause and end-of-clause pauses in the L2 approximated the L1 pattern, indicating important changes from pre- to post-SA" (p.188), therefore pause location was argued to be an important measure for explaining difficulties in fluency development and tracking fluency changes in general.

Huensch and Tracy-Ventura (2017) also explored the theme of fluency development in the SA context considering pause location. The researchers collected data over a 2 year period at 6 different times in order to investigate how fluency changes during the SA period and a year after it, however here this study will be presented in pretest and posttest form in order to discuss the final impact of the SA. The study was based on 24 English-speaking learners of Spanish who spent nine months in the SA context. The data was collected by means of the three picture-based narration tasks at the beginning and the end of SA

experience and were subjected to analysis. The three measurements of speed fluency: MLoR, SR and mean syllable duration and four measurements of breakdown fluency: number of silent pauses; number of filled pauses, mean silent pause duration within ASU and mean silent pause duration between ASU were used for the analysis of fluency changes. The results indicated positive changes in SR, MLoR, number of silent pauses, filled pauses and mean silent pause duration within ASU which were argued to reflect the positive impact of SA on fluency development.

### ***2.3.3. Studies on Vocabulary in the Study Abroad Context***

The measurement of vocabulary knowledge has long been mainly a feature of language schools, which used vocabulary tests to place learners according to their level of knowledge. In SLA, vocabulary measurement became popular in the 1990s, while the works of Freed (1995b) and Milton and Meara (1995) drew attention to lack of studies on vocabulary development in the SA context, which resulted in the increase of such studies. The studies presented below will be described with an emphasis on the measurement methods.

The study by Milton and Meara (1995) is “one of the first oft-cited and groundbreaking studies on vocabulary growth during SA” (Zaytseva et al., 2018, p.212). The study was based on fifty-three exchange students at different proficiency levels from four countries (Germany, France, Italy and Spain) who spent six months in the UK. Students’ receptive vocabulary knowledge was measured by means of the Yes/No Eurocentres Vocabulary Size Test at the beginning and at the end of the SA period. Results showed that SA positively impacted vocabulary which showed gains “at a rate of over 2,500 words per year” (Milton & Meara, 1995, p.31). Additionally, the researchers noticed that the students with smaller vocabularies at the beginning of SA experience improved the most, while the students with larger vocabularies improved less.

Ife et al. (2000) conducted a study on thirty-six intermediate/advanced English-speaking learners of Spanish who spent 1 or 2 semesters (4 or 8 months respectively) in Spain. The researchers measured productive vocabulary by means of a translation test (to measure vocabulary size) and Three Word Association Test (to measure lexical organization knowledge). The results showed the positive impact of the SA context on both size and degree of organizational knowledge. Contrary to Milton and Meara (1995), no difference in vocabulary development for intermediate and advanced students was found; while a longer SA experience was found to be more beneficial than the shorter one.

Collentine (2004) compared productive vocabulary development among 46 American students of Spanish in SA and AH contexts. The corpus consisting of two-minute segments from pretest and posttest interviews (OPI) was analyzed not by means of any test, but simply in terms of acquisition of new words and the rate of unique words within seven lexical categories: adjectives, adverbs, conjunctions, nouns, prepositions, pronouns, and verbs. The results indicated that AH students used slightly more new words while SA students used slightly more unique words; however the difference between the two contexts was not significant and therefore, Collentine (2004) stated that, “in response to the question of whether the SA context yields better overall [...] lexical abilities than the AH context, the answer is a qualified no” (p.244). Similar findings were presented by Freed et al. (2003) who also did not find any difference among two groups of students from SA and AH contexts assessed by six judges.

The means of measuring vocabulary in these two studies were doubted by Dewey (2008) who decided to use three different vocabulary tests: Vocabulary Matching Test (receptive knowledge), Vocabulary Knowledge Scale (depth of vocabulary knowledge) and The Situational Vocabulary Test (everyday vocabulary) to compare vocabulary development of 56 English-speaking learners of Japanese in three learning contexts: “study abroad (SA), intensive domestic immersion (IM) and academic year formal classroom (AY) learning” (p.127). Dewey (2008) found that the students who took part in the SA experience showed greater scores in all three tests when compared to learners from AY context, while the results from two tests (breadth and depth) for SA and IM students were similar, but in the third test SA showed significantly greater increase in the use of everyday language.

Jiménez-Jiménez (2010) studied the difference in the development of both productive and receptive vocabulary on 87 intermediate/advanced English-speaking learners of Spanish in the SA and AH contexts with a translation test and a specially-designed word association test (the participant had to cross out one word which was not associated with other two in terms of meaning (i.e., synonyms, antonyms, metonymies or collocations)) respectively. The results showed a positive impact of contact with a target-language community in the SA on both productive and receptive vocabulary, while no significant gains on either test were found for AH students.

Fitzpatrick (2012) used a word association test called Lex30 to track the changes in productive vocabulary in case study of a 21-year-old Chinese learner of English studying in the UK. The participant took the same Lex30 test six times during the eight months of the SA experience. However, the traditional Lex30 scoring procedure was changed to “the aspects of

word knowledge identified by Nation (2001) [Table 1]” (p.86) which consisted of a word’s form (number of responses given at each test time, percentage of items misspelled, occurrence of affixes), meaning (number of cue words responded to, misunderstood cue words) and use (percentage of native speaker-like responses). The goal of the study was to track the “micro-development” (p.92) of vocabulary from test to test in the SA context, but the development was rather “chaotic and elusive” (p.92) and it was “not possible to draw strong conclusions from this study” (p.92). Nevertheless, when focusing only on pretest and posttest results, the increase in number of responses given at each test time, occurrence of more affixes, number of cue words responded to, percentage of native speaker-like responses, number of collocational responses and decrease in number of misunderstood cue words, signaled the positive impact of the SA context on productive vocabulary development in general.

Serrano et al. (2012) investigated how depth of productive vocabulary changes after three months and nine months in the SA context. The data from 14 intermediate/advanced Spanish-speaking learners of English were collected by means of “The Picnic Story” (Heaton, 1966) and analyzed for the lexical richness with Guiraud’s Index of Lexical Richness (“word types divided by the square root of the word tokens (Types/Tokens)” (Serrano et al., 2012, p.145)). The results indicated the increase over the three SA months in lexical richness and even greater increase over nine months. Two years later, a similar study was conducted by Lara (2014) who using the same Guiraud’s Index measured lexical richness of 47 Catalan/Spanish-speaking learners of English after the three-month and six-month SA experience. Contrary to the result of Serrano et al. (2012), no significant changes were detected in lexical richness after three-month SA period, however the significant changes were found after six-month SA stay.

Dewey (2008) and Zaytseva et al. (2018) with a difference in ten years, both correctly noticed that there was no lack of studies on receptive vocabulary in the SA context, while more research should be done on productive vocabulary.

#### ***2.3.4. Studies on the Interaction Between Vocabulary and Fluency in the Study Abroad Context***

Greater vocabulary knowledge has usually been associated with greater reading skills (Qian, 2002; Van Gelderen et al., 2004), while the studies by De Jong et al. (2012), Uchihara and Clenton (2018) and Uchihara and Saito (2016) also highlighted the link between learners with greater vocabulary size and faster speaking skills (see section 2.2.4. Studies on the Relationships Between Vocabulary and Fluency). Despite the rise of the research on the

impact of the SA context which was found to be beneficial both for the development of fluency and vocabulary knowledge, there exist only few studies which partially concentrated on the interaction between vocabulary knowledge and fluency in this context that allows exploring the changes which a learner undergoes in a natural way.

For instance, the study by Leonard and Shea (2017) only partially covered the question of interaction between vocabulary knowledge and fluency. One of the main goals of their study was to find out whether the vocabulary scores predicted greater gains in fluency (and also in accuracy, lexical variety, syntactic complexity, and lexical complexity) in 39 English-speaking learners of Spanish who spent three months in a SA context. In order to measure fluency, 10 utterance fluency measures were used (e.g MLoR, AR, mid-clause pauses, end-of-clause pauses, etc.) while the “30-item untimed vocabulary test, adapted from the Diploma de Español como Lengua Extranjera (DELE)” (p.183) was used to measure vocabulary knowledge (without providing description of how and what exactly the test measured). The results indicated that the higher pre-SA vocabulary scores provided learners with “a slight advantage in accuracy, lexical variety, and complexity gains during SA, but not in fluency gains”. The posttest results indicated the development both in fluency and vocabulary knowledge; however, the researchers did not provide the analysis of their interaction (since it was not their goal).

Similarly, one of the goals of McManus, Mitchell and Tracy-Ventura’s (2021) study was to explore the relationships between fluency and lexis. Their study was based on 56 advanced English-speaking learners of French or Spanish who spent nine months abroad. The data was collected by means of picture-based narrative 1 time before, 2 times during and 1 after the SA period (and also 3 times in post-SA year). SR and MLoR were chosen to measure the changes in fluency while “lexis was operationalized as lexical diversity, computed using the VocD command in CLAN (MacWhinney, 2000), resulting in a score *D* (Malvern & Richards, 2002)” (p.15). The findings of the study showed that SA contributed to the increase in fluency and lexis after nine-month stay; when it comes to the interaction between them, the results “showed significant and long-lasting relationships between fluency and lexis” (p.25).

However, none of these studies has concentrated solely on the relationships between vocabulary knowledge and fluency (as i.e. Uchihara & Clenton, 2018; Uchihara & Saito, 2016) in the SA context which leaves room for further research in this direction.

### 3. The Study

The present study follows a pretest-posttest design that qualitatively examines the changes in productive vocabulary, fluency and their interaction over the SA period in the interviews of 5 French, intermediate-to-advanced learners of English who spent an academic year in English-speaking environment.

The study aims to answer the following Research Questions (RQ):

RQ.1. To what extent did oral fluency change over an academic year in the SA context?

RQ.2. To what extent did productive vocabulary change over an academic year in the SA context?

RQ.3. Do the changes in productive vocabulary appear to show parallel developmental trajectories with oral fluency over an academic year in the SA context?

### 3.1. Method Section

This section aims to describe in detail how this study was organized. Firstly, the information about the participants will be presented. Secondly, I will explain how the data were collected. Finally, the process and nuances of data coding and analysis will be described.

#### 3.1.1. Participants

The general information about the participants is presented in Table 2.

**Table 2**

*Background of the Participants.*

| Participant | Gender | Age<br>(Pre-sojourn) | Languages<br>(specialization) | L1                 | Pre-sojourn           | Country |
|-------------|--------|----------------------|-------------------------------|--------------------|-----------------------|---------|
| A           | F      | 19                   | English +<br>Italian          | French             | Lower<br>intermediate | Ireland |
| C           | F      | 18                   | English +<br>Chinese          | French             | Lower<br>intermediate | Ireland |
| M           | F      | 19                   | English +<br>Spanish          | French             | Advanced              | Ireland |
| N           | M      | 19                   | English +<br>Chinese          | French             | Advanced              | Ireland |
| Y           | M      | 19                   | English +<br>Arabic           | French+<br>Turkish | Advanced              | England |

The participants were five 18/19 year old students (A, C, M, N, Y (pseudonyms)) in their first or second year of a degree course in Applied Foreign Languages at Université Paul-Valéry, Montpellier, France. They were native speakers of French (Y was also a native speaker of Turkish, but the French was his dominant language) who studied English as a first language and Arabic, Chinese, Italian or Spanish as a second one. The participants' English proficiency, measured by means of the Oxford Quick Placement Test before SA experience, was considered either lower intermediate (A, C) or advanced (M, N, Y). Among the five participants, three participants were female (A, C, M), and two were male (N, Y). The participants had mainly studied English in a formal context and have never been to an English-speaking country for a long period of time. They spent 9 months in Ireland (A, C, M, N) or England (Y) being surrounded by native and non-native speakers of English in both formal (i.e., lectures and seminars in the host university, etc.) and informal (i.e., transport, job, supermarket, etc.) contexts, as a part of Erasmus+ programme.

### ***3.1.2. Data Collection***

This section has 2 main objectives. It describes how data for the fluency analysis were collected and then the description of vocabulary data collection is provided.

**3.1.2.1. Fluency.** The data for the fluency analysis were collected by means of an oral interview before the departure to English-speaking country (June 2018) and after the return to France (June 2019). The interviews were recorded on camera and conducted by Prof. Amanda Edmonds and/or Prof. Pascale Leclercq in order to compile the PROLinsa corpus (the data were also collected at three times during SA, but for the goals of my study only pretest and posttest interviews will be described). The format of a semi-interview was chosen because it provides more freedom for the participants to express themselves, not being subjected to one topic (as in i.e., picture-narration tasks or monologues) and thus allows to elicit the closest to natural use of participants' language in real life: "In an interview a rapport is established between the interviewer and the interviewee. Not only is physical distance between them annihilated, the social and cultural barrier is also removed; and a free mutual flow of ideas to and fro takes place" (Pandey & Pandey, 2015, p.60). Additionally, in order to get the participants' natural use of English, no corrective feedback was given during the interviews; however, the interviewers sometimes helped the interviewees with the forgotten or unknown words. Moreover, the interviewers established a set of questions which were designed with an aim to make the interviewee produce more speech. For example, the questions were open-ended, implying longer than yes/no answers with the possibilities of going well beyond the question and mainly concentrating on personal experience,



motivations, expectations and goals of the interviewee (e.g., What are your expectations regarding this stay abroad?, How do you feel about going abroad for a whole academic year?, How did it feel to come back to France?, Future plans?). The recorded interviews varied in length, for example, the shortest interview was 18 minutes 34 seconds (M, pre-SA interview) while the longest one was 53 minutes 47 seconds (Y, post-SA interview (see Table 3)). The collected interviews were further subjected to coding and analysis.

**3.1.2.2. Vocabulary.** Vocabulary data were elicited by means of productive word association task Lex30 before and after the SA period. Lex30 used in this study presented the participants with a list of 30 stimuli words and asked them to provide 4 responses. The 30 stimulus words (cues) were selected following three criteria: 1) cues were chosen from the first 1000 of most frequent words of English (following Meara & Fitzpatrick, 2000); 2) the chosen cues did not have one common association for all test takers which resulted in producing one same response; 3) chosen cues tended to elicit a high proportion of infrequent responses. The procedure of test taking consisted of a PowerPoint presentation with 30 stimulus words (each cue was shown for 30 seconds, while in general the test took no longer than 15 minutes) shown to the test takers who wrote down their association in a separate paper. Finally, all answers provided by 5 participants were typed into the Excel document without correcting spelling errors and were subjected for further coding and analysis. (See Appendix A, for the exemplification of a completed and not coded pre-SA test by Y).

### **3.1.3. Data Coding**

This section describes the process of fluency data coding and then vocabulary data coding is presented.

**3.1.3.1 Fluency.** The collected interviews were first manually transcribed in CLAN (program software used for audio and video transcriptions (MacWhinney & Wagner, 2010)) by a group of 9 transcribers. Each interview was at first transcribed by one transcriber and then checked and corrected by a second transcriber. The transcription was organized in the form of separate utterances (generally speaking, one verb corresponded to one utterance; one line corresponded to one utterance). Additionally, each utterance was linked to the corresponding segment of a video file (such organization allows more options for investigations on these interviews in future studies). CLAN transcription codes were used by transcribers to mark details of participants' speech, for example, all reformulations, contractions, errors, missing words, laughs and repetitions were marked as such in order to reflect orally produced speech as precisely as possible in a written transcription.

Following De Jong et al. (2015) and Huench and Tracy-Ventura (2017) the transcriptions were divided into ASU for the further analysis of pause location within/between ASU (see section 3.1.4.1. Fluency). Each transcription was copied to a separate Word file and manually divided into ASU according to the definition: “a single speaker’s utterance that consists of either an independent clause, or sub clausal unit, with any subordinate clause” (p.365) and following the rules developed by Foster et al. (2000). The use of ASU as a unit was chosen because it allows a more detailed division of the spoken data (in the following examples, the boundaries of ASU were marked by a double slash “//...//” and the length of pauses longer than 250ms were marked in parentheses “(...”). For example, contrary to other units which mainly focused on written data, the ASU may be not only an independent clause (\*INT1: Elementary? - \*M: // yeah it is elementary //) but also a sub-clausal unit which may be completed by means of recovery ((\*INT1: Elementary? - \*M: // yeah elementary //) or just one utterance (\*INT1: Elementary? - \*M: // yeah //). Additionally, the division into ASUs allows differentiating pause location. For instance there are two kind of pauses in the following example: “\*Y: // it wasn 't really (600ms) professional sorry . // (600ms) // &erh japanese people &erh yeah &uh we 've been[\*] there . //” . The first pause of 600ms occurred within ASU while the second pause occurred between 2 ASUs. However, the two mentioned ASUs are independent from each other, therefore the pause between them is considered as a pause between ASUs, while things are different for coordination clauses. For example, in coordinated clauses (coordinated by “and”, “but” or “or”) when the two coordinated verb phrases have the same subject, the two clauses usually belong to the same ASU. However, if a pause longer than 500ms separates the two verbs (either before or after the coordinator), then the clauses are considered as two separate ASUs. In the following example, “\*N: // i try to put &euh English subtitles (360ms) or try to put no subtitles at all. // (588ms) // and try to work on that. //”, the first pause of 360ms occurred within ASU while the second one of 588ms was considered as one which occurred between 2 ASUs because it was longer than 500ms, if the pause had been shorter than 500ms the two ASUs would be counted as one ASU. Once the division was done, it was subjected to the measurement of length and number of pauses within and between the ASU.

Phonetic coding was realized in PRAAT (program software used for speech analysis (Boersma & van Heuven, 2001)). Needless to mention that the interviews consisted not only of participants’ speech, but also of interviewers’ speech, therefore it was decided to use only participants’ speaking time (time when participant is speaking including pauses). Additionally, due to the fact that the length of participants’ speaking time varied from

participant to participant and from pre-SA and post-SA experience, it was decided to use eight minutes of that time from each interview. Eight minutes were chosen because it was the shortest speaking time of M from the whole pre-SA interview in comparison to other participants (and M in post-SA interview) who produced longer speaking time (see Table 3). The eight minutes were counted from the first open-ended question by the interviewer to which participants generally provided full answers (i.e., “Do you have any work experience?” in the pre-SA interviews and “When did you leave Cork/London/Dublin?” in the post-SA interviews). The sum of eight minutes from each interview was manually counted by listening to the interviews and adding the length of the participant’s speech.

**Table 3**

*Total Duration and Duration of Speaking Time of the Participants in Pretest and Posttest Interviews*

| <b>Participant</b> | <b>Duration of Pretest interview</b> | <b>Duration of speaking time analyzed in Pretest interview</b> | <b>Duration of Posttest interview</b> | <b>Duration of speaking time analyzed in Posttest interview</b> |
|--------------------|--------------------------------------|--|---------------------------------------|---|
| A                  | 27:00                                | 667s - 11:07   | 34:16                                 | 1412s - 23:32   |
| C                  | 26:49                                | 665s - 11:05   | 23:27                                 | 604s - 10:04  |
| M                  | 18:34                                | 485s - 8:05  | 20:51                                 | 686s - 11:26  |
| N                  | 23:55                                | 666s - 11:06   | 37:21                                 | 1447s - 24:47   |
| Y                  | 44:24                                | 1820s - 30:20  | 53:47                                 | 2433s - 40:33   |

In PRAAT, the eight minutes of speech time from each interview were extracted into a separate audio-file (i.e., M speaks for eight minutes in the entire interview thus the whole interview was an extract, while only the first eight minutes out of 11m07 seconds were extracted for A). Each of the ten audio-files was automatically annotated by PRAAT for silent and sounding intervals, labeled as “#” and “IPU” (inter-pausal unit) respectively. Based on the study by De Jong and Bosker (2013) the silence threshold was set at 250ms (see

section 2.1.2.3.2. Pause Threshold, for more details). However, since the automatic annotation in PRAAT cannot distinguish the voices of different speakers and separately annotate their speech, the annotations of the audio-files were done by manually correcting PRAAT's annotations. For instance, the speech (including pauses) of interviewers was labeled as "INT1" or "INT2" while the speech of a participant was subjected to more detailed annotation: for example, only silent pauses were labeled as "#" and the speech sequences without silent pause were labeled as "IPU", but also laughs ("laugh"), initial pauses ("Ip" (pause at the beginning of "IPU")) and non-English speech ("French" or "Spanish") were also annotated. These annotations were made in order to exclude laughs or languages other than English from the speaking time. Similarly, initial pauses were marked in order to be excluded from the general number of (within/between) pauses. The annotated extracts were subjected for further fluency analysis.

**3.1.3.2 Vocabulary.** Once the obtained answers from each participant were typed into the Excel document they were subjected to coding. First of all, since the aim of the Lex30 is to measure the knowledge of productive vocabulary and not grammatical accuracy, the spelling errors and errors of morphology in all words were corrected (i.e., "cummon" to "common", "retein" to "retain", "flue" to "flu", etc.). Secondly, all proper nouns were deleted because they cannot be assessed by any wordlist or database. Finally, based on Appendix B in Meara and Fitzpatrick (2000), the lemmatization of responses with certain inflectional suffixes (i.e., "-ing" as in "writing" was lemmatized to "write", or plural forms as in "games" to "game", etc.) and derivational affixes (i.e., "-ness" as in "illness" was lemmatized to "ill" or "-y" adjectives created from nouns as in "messy" to "mess" etc.) was done. See Appendix B, for the exemplification of a completed and coded pre-SA test by Y.

#### **3.1.4. Data Analysis**

This section provides the description of fluency data analysis first and then vocabulary analysis is presented.

**3.1.4.1. Fluency.** For the purposes of this study, it was decided to measure utterance fluency (Segalowitz, 2010) by both temporal and hesitation phenomena. For example, speed fluency was measured using three measures: AR and two "composite measures" (SR and MLoR) which were found to be the best reflectors of changes in fluency by most of the studies (see section 2.1.2.1. Speed Fluency). These speed measurements were calculated in the following way:

- 1) Speech Rate was expressed as the time necessary (in seconds) to produce a syllable (including pause time), and calculated as: total time (including silent pause time)

divided by total number of syllables in the extracted segments (8 minutes = 480 seconds).

2) Mean length of runs was calculated as: total number of syllables divided by total number of runs (IPUs).

3) Articulation rate was expressed as the time necessary (in seconds) to produce a syllable excluding pause time, and calculated as: total time (excluding silent pause time) divided by total number of syllables in the extracted segments (8 minutes = 480 seconds).

When it comes to breakdown fluency, four measurements were used, taking into consideration frequency, location, and duration of silent pauses (following De Jong (2016) and Huench & Tracy-Ventura (2017): number of silent pauses within ASU, number of silent pauses between ASU, mean length of silent pauses within ASU, mean length of silent pauses between ASU. Repair fluency was not measured since it has been argued that it rather reflects individual differences (see section 2.1.2.2. Repair Fluency). Based on Huensch and Tracy-Ventura (2017) these breakdown measures were calculated in the following way:

1) Number of silent pauses within/between ASU was calculated by counting the total number of silent pauses longer than 250ms at each of the two locations.

2) Mean length of silent pauses within/between ASU was calculated as: the total length of silent pauses longer than 250ms divided by their total number at each of the two locations.

All calculations were done using Excel. The obtained results from these measures are presented in 4. Results section.

**3.1.4.2. Vocabulary.** In comparison to the fluency analysis, the vocabulary analysis was more straightforward. Once coding was finished all responses were subjected to frequency analysis in the Corpus of Contemporary American English (*Corpus*, n.d.). The responses were manually filed in the search frequency system and then assessed due to their frequency. The words which were located within the first 1,000 of most frequent words (words from 1 to 1000) were given 0 points while the words which were located outside the first 1,000 of most frequent words were awarded 1 point. The maximum score that a participant could theoretically obtain was 120 points (30 cues x 4 responses). The final score is represented as the sum of the “infrequent” points and also the percentage of infrequent responses (following Fitzpatrick & Meara, 2004 and Fitzpatrick, 2007) in the following section.

## 4. Results

This section will be divided into three parts: Fluency in the SA context, productive vocabulary in the SA context and interaction of fluency and productive vocabulary in the SA context. Each part will present and describe the obtained results.

### 4.1. Fluency in the Study Abroad Context

Table 4 presents the summary for the three speed fluency measures in pretest and posttest interviews for the five participants. An improvement in SR and AR will be visible in a reduction of the amount of time needed to produce a syllable, while a positive development in MLoR will be visible in an increase of length of the mean run score (the degree of change in all three measures is presented in %).

**Table 4**

*Speed Fluency Results*

| Name:       | SR      |          |       | AR      |          |       | MLoR    |          |       |
|-------------|---------|----------|-------|---------|----------|-------|---------|----------|-------|
|             | Pretest | Posttest | %     | Pretest | Posttest | %     | Pretest | Posttest | %     |
| A           | 0,44    | 0,30     | -31,6 | 0,36    | 0,27     | -25,8 | 6,39    | 13,92    | 117,5 |
| C           | 0,48    | 0,40     | -17,6 | 0,36    | 0,33     | -7,4  | 5,33    | 7,02     | 31,6  |
| M           | 0,36    | 0,27     | -24,5 | 0,28    | 0,24     | -15   | 6,70    | 14,80    | 120,7 |
| N           | 0,43    | 0,32     | -24,9 | 0,29    | 0,27     | -8,1  | 5,59    | 8,63     | 54,4  |
| Y           | 0,34    | 0,25     | -26,4 | 0,25    | 0,22     | -14,3 | 6,53    | 15,23    | 133,2 |
| <b>Mean</b> | 0,41    | 0,31     |       | 0,31    | 0,26     |       | 6,11    | 11,92    |       |
| <b>SD</b>   | 0,06    | 0,05     |       | 0,04    | 0,04     |       | 0,60    | 3,81     |       |

Noteworthy changes were found for all 5 participants in all speed fluency measures. First of all, the time (including pausing) needed to produce a syllable was reduced in the posttest interviews of all participants, which means that SR showed positive improvement. Moreover, the results showed that the changes were similar for 3 participants: M (-24,5%), N (-24,9%) and Y (-26,4%) and to a lesser extent for A (-31,6%) and C (-17,6%). Secondly, a similar tendency of faster articulation (but excluding pauses) in syllable production was found in the posttest interviews of all participants. However, the development in AR was

more varied than in SR for each participant. For instance, the smallest percentages of changes were found for C (-7,4%) and N (-8,1%), greater changes were found in the posttest interviews of M (-15%) and Y (-14,3%), while A (-25,8%) showed the biggest change in AR. Finally, the biggest increase was found in the length of mean run for all participants, which indicated the increase in MLoR. However, in contrast to SR and AR, the post-test standard deviation for MLoR was bigger which means that the participants' increase in that measure was more dispersed. For example, although the percentages of change for C (31,6%) and N (54,4%) were bigger than any percentage in SR (except for equality with 31,6% for A) or AR, the percentages for other 3 participants, were immensely higher: A (117,5%), M (120,7%) and Y (133,2%). To sum up, the changes in all 3 measures of speed fluency were found in the posttest interviews of all 5 participants, which may suggest that fluency developed. However, in order to state that fluency developed one should also look at the breakdown fluency results.

Table 5 presents the summary of four breakdown fluency measures in pretest and posttest interviews of the participants. A decrease in the mean length of silent pauses within/between ASU and number of silent pauses within ASU will mark the development of fluency. When it comes to the number of silent pauses between ASU, it may be more difficult to treat the results, because the decrease in that measure may be treated as a positive indicator of fluency development only if the number of silent pauses within ASU also decreases, while the increase in that measure may also be treated in the same way only if number of silent pauses within ASU also decreases.

**Table 5***Breakdown Fluency Results*

| Name:       | M.L. of SP Within ASU |          |       | M.L. of SP Between ASU |          |       | Number of SP Within ASU |          |       | Number of SP Between ASU |          |      |
|-------------|-----------------------|----------|-------|------------------------|----------|-------|-------------------------|----------|-------|--------------------------|----------|------|
|             | Pretest               | Posttest | %     | Pretest                | Posttest | %     | Pretest                 | Posttest | %     | Pretest                  | Posttest | %    |
| A           | 667,49                | 563,27   | -15,6 | 789,06                 | 730,46   | -7,4  | 77                      | 43       | -44,1 | 15                       | 28       | 86,6 |
| C           | 867,31                | 538,71   | -37,8 | 1267,76                | 767,80   | -39,4 | 83                      | 74       | -10,8 | 21                       | 21       | 0    |
| M           | 630,36                | 549,46   | -12,8 | 855,73                 | 618,23   | -27,7 | 109                     | 45       | -58,7 | 19                       | 34       | 78,9 |
| N           | 783,07                | 572,14   | -26,9 | 1349,35                | 698,56   | -48,2 | 108                     | 62       | -42,5 | 31                       | 53       | 70,9 |
| Y           | 647,3                 | 535,06   | -17,3 | 737,97                 | 585,61   | -20,6 | 104                     | 29       | -72,1 | 44                       | 42       | -4,5 |
| <b>Mean</b> | 719,11                | 551,73   |       | 999,97                 | 680,13   |       | 96,2                    | 50,6     |       | 26                       | 35,6     |      |
| <b>SD</b>   | 102,16                | 15,82    |       | 286,22                 | 76,36    |       | 15,05                   | 17,55    |       | 11,66                    | 12,42    |      |

To begin with, the results indicate a decrease in mean length of silent pauses within ASU for all participants. Interestingly, the percentage of change was smaller for three participants (A (-15,6%), M (-12,8%) and Y (-17,3%)) who made shorter silent pauses in pretest interviews (667,49ms, 630,36ms and 647,3ms respectively) and higher for other two (N (-26,9%) and C (-37,8%)) who produced the longest silent pauses in pretest interviews (783,07ms and 867,31ms respectively). Such decrease led to the standardization of the mean length of silent pauses within ASU in posttest interviews, since the deviation from the mean decreased immensely (15,82) in comparison to the pretest interview (102,16). A similar tendency was found in the decrease of mean length of silent pauses between ASU for all participants in posttest interviews. For example, the percentage of change also varied for each participant: the smallest change was found in posttest interview of A (-7,4%) who produced the shortest silent pauses between ASU in her pretest interview (789,06ms), more dramatic changes of -20,6% and -27,7% were found for N and M respectively who made longer pauses (855,73ms and 737,97ms) at pre-test, while the highest percentages of change were found for C (-39,4%) and N (-48,2%) who made the longest silent pauses between ASU (1267,76ms and 1349,35ms respectively). In that way the deviation from the mean also immensely



decreased from pretest (286,22) to posttest interview (76,36), which means that the difference in mean length of silent pauses between ASU became less evident. To sum up, the decrease in mean length of silent pause both within and between ASU may be treated as the development of fluency. Additionally, an interesting tendency was found in the results of both measures indicating that the proportion of decrease over the SA context is higher for those who made longer silent pauses within/between ASU and lower for those who made shorter silent pauses within/between ASU.

In contrast to the changes in mean length of silent pauses, the changes in the number of silent pauses were more varied. To begin with, the percentages of change in the number of silent pauses within ASU were quite different for each participant: from the lowest -10,8% change for C, to higher -42,5%, -44,1% and -58,7% change for N, A and M respectively, and to the highest -72,1% change for Y. Contrary to the mean length of silent pauses within/between ASU, the changes in number of silent pauses within ASUs did not follow the same tendency of higher number – higher percentage. Nevertheless, all five participants decreased in the use of silent pauses within ASU which may be treated as an indicator of fluency development. More interesting results were found in the number of silent pauses between ASU. For example, the small decrease in that measure was found only for Y (-4,5%); along with the decrease in the number of silent pauses within ASU this result may be treated as an indicator of fluency development, since both measures decreased. At the same time a great increase was found for three participants (A (86,6%), M (78,9%) and N (70,9%)) which along with the decrease in number of silent pauses within ASU may also be treated as an indicator of fluency development (because when the participant improves, the number of silent pauses within ASU should decrease while the number of silent pauses between ASU may increase). Finally, no change was found for C (0%) which along with the decrease in the number of silent pauses between ASU may also be treated as an indicator of fluency development.

To sum up, changes that indicate fluency development over the SA period were found for all 5 participants in 6 utterance fluency measures (the change in number of silent pauses between ASU formally may also be treated as an indicator of fluency development, however it mainly depends on the changes in number of silent pauses between ASU, and may not be treated as a separate measure).

## 4.2. Productive Vocabulary in the Study Abroad Context

Table 6 presents the Lex30 results from the pretest and posttest interviews for the five participants. An increase in productive vocabulary will be visible in the increase of raw scores and/or % of infrequent responses.

**Table 6**

*Lex30 Results*

| Score       | Raw Score |          |       | % of infrequent responses |          |       |
|-------------|-----------|----------|-------|---------------------------|----------|-------|
|             | Pretest   | Posttest | %     | Pretest                   | Posttest | %     |
| A           | 20        | 41       | 105   | 33,3                      | 38,3     | 5     |
| C           | 33        | 39       | 18,1  | 36,2                      | 43,8     | 7,6   |
| M           | 58        | 68       | 17,2  | 49,57                     | 56,6     | 7,1   |
| N           | 63        | 54       | -14,2 | 60                        | 46,5     | -13,5 |
| Y           | 47        | 53       | 12,7  | 41,5                      | 44,5     | 3     |
| <b>Mean</b> | 44,2      | 52       |       | 51                        | 45,9     |       |
| <b>SD</b>   | 17,7      | 10,8     |       | 11,6                      | 6,6      |       |

To begin with, the changes in raw score were quite polarized. For example, N was the only participant who showed a slight decrease in raw score in post-SA test (-14,2%), while a slight increase was found for three participants (C (18,1%), M (17,2%) and Y (12,7%)) and an immense increase for A (105%). In conclusion, the productive vocabulary measured by the raw scores may be said to have developed for 4 out of 5 participants. The findings in the % of infrequent responses were of similar nature. N was also the only participant who showed a lower percentage of infrequent responses in post-SA test which decreased by -13,5% while the other four participants showed only slight increases (A (5%), C (7,6%), M (7,1%) and Y (3%)). In general the results measured by % of infrequent responses indicated a slight increase in the development of productive vocabulary for 4 out 5 participants.

To sum up, the changes that indicate the productive vocabulary development over the SA period were found for 4 out of 5 participants in both ways of scoring Lex30.

### 4.3. Interaction Between Productive Vocabulary and Fluency in the Study Abroad

#### Context

In this section, in order to study the interaction between the two constructs, the results will be presented in one table of combined productive vocabulary and fluency results (except for the number of silent pauses between ASU, because it is not possible to treat it as a separate measure) for each participant. The changes of the same degree for most measures of productive vocabulary and fluency may signal the parallel development between the two constructs. Table 7 presents fluency and productive vocabulary results obtained from A.

**Table 7**

*Utterance Fluency and Productive Vocabulary Results for A*

| A                            |         |          |       |
|------------------------------|---------|----------|-------|
| Measure/ Time:               | Pretest | Posttest | %     |
| SR                           | 0,44    | 0,30     | -31,6 |
| AR                           | 0,36    | 0,27     | -25,8 |
| MLoR                         | 6,39    | 13,92    | 117,5 |
| M.L. of SP within<br>ASU     | 667,49  | 563,27   | -15,6 |
| M.L. of SP between<br>ASU    | 789,06  | 730,46   | -7,4  |
| Number of SP within<br>ASU   | 77      | 43       | -44,1 |
| Raw Score                    | 20      | 41       | 105   |
| % of infrequent<br>responses | 33,3    | 38,3     | 5     |

When analyzing the results obtained from A, on one hand one may refer to the greater fluency changes in SR (-31,6%), AR (-25,8%), MLoR (117,5%) and number of silent pauses

within ASU (-44,1%) and greater productive vocabulary changes in raw scores (105%) and argue that there may be an interaction between the two since they show parallel developmental trajectories. However, when adding the other measures, for example the slight changes in % of infrequent responses (5%) and mean length of silent pauses between ASU (-7,4) which contradict the other measures, the developmental trajectories become less similar. To conclude, both productive vocabulary and fluency results showed changes of different degrees and thus they did not show parallel developmental changes in data of A.

The results of fluency and productive vocabulary obtained from C are presented in Table 8.

**Table 8**

*Utterance Fluency and Productive Vocabulary Results for C*

| C                            |         |          |       |
|------------------------------|---------|----------|-------|
| Measure/ Time:               | Pretest | Posttest | %     |
| SR                           | 0,48    | 0,40     | -17,6 |
| AR                           | 0,36    | 0,33     | -7,4  |
| MLoR                         | 5,33    | 7,02     | 31,6  |
| M.L. of SP within<br>ASU     | 867,31  | 538,71   | -37,8 |
| M.L. of SP between<br>ASU    | 1267,76 | 767,80   | -39,4 |
| Number of SP within<br>ASU   | 83      | 74       | -10,8 |
| Raw Score                    | 33      | 39       | 18,1  |
| % of infrequent<br>responses | 36,2    | 43,8     | 7,6   |

The provided results are also difficult to compare. For instance, the changes in two productive vocabulary measures (raw score ((18,1%) and % of infrequent responses (7,6%)) and three fluency measures (SR (-17,6%), AR (-7,4%) and number of silent pauses within ASU (-10,8%)) may be considered rather low in comparison to moderate changes in the other 3 measures of fluency (MLoR (31,6%), mean length of silent pauses within (-37,8%) and between (-39,4%) ASU). Therefore, due to the fact that differences of the same degree were found for two productive vocabulary measures and only for the half of fluency measure (which means not for majority), it is hard to connect them to the interaction of the mentioned constructs.

Less ambiguity was found in the results obtained from M presented in Table 9.

**Table 9**

*Utterance Fluency and Productive Vocabulary Results for M*

| <b>M</b>                     |                |                 |          |
|------------------------------|----------------|-----------------|----------|
| <b>Measure/ Time:</b>        | <b>Pretest</b> | <b>Posttest</b> | <b>%</b> |
| SR                           | 0,36           | 0,27            | -24,5    |
| AR                           | 0,28           | 0,24            | -15      |
| MLoR                         | 6,70           | 14,80           | 120,7    |
| M.L. of SP within<br>ASU     | 630,36         | 549,46          | -12,8    |
| M.L. of SP between<br>ASU    | 855,73         | 618,23          | -27,7    |
| Number of SP within<br>ASU   | 108            | 62              | -42,5    |
| Raw Score                    | 58             | 68              | 17,2     |
| % of infrequent<br>responses | 49,57          | 56,6            | 7,1      |

The similar slight changes in raw score (17,2) and % of infrequent responses (7,1) for productive vocabulary and slight changes in AR (-15) and mean length of silent pauses (-12,8) for fluency are not sufficient to state that fluency and productive vocabulary show the parallel developmental trajectories, because the other four fluency measures showed either moderate (SR (-24,5), mean length of silent pauses between ASU (-27,7) and number of silent pauses within ASU (-42,5) or immense (MLoR (120,7)) changes. To conclude, no interaction between productive vocabulary and fluency was found for this participant.

More interesting results, presented in Table 10, were found for N.

**Table 10**

*Utterance Fluency and Productive Vocabulary Results for N*

| N                            |         |          |       |
|------------------------------|---------|----------|-------|
| Measure/ Time:               | Pretest | Posttest | %     |
| SR                           | 0,43    | 0,32     | -24,9 |
| AR                           | 0,29    | 0,27     | -8,1  |
| MLoR                         | 5,59    | 8,63     | 54,4  |
| M.L. of SP within<br>ASU     | 783,07  | 572,14   | -26,9 |
| M.L. of SP between<br>ASU    | 1349,35 | 698,56   | -48,2 |
| Number of SP within<br>ASU   | 108     | 62       | -42,5 |
| Raw Score                    | 63      | 54       | -14,2 |
| % of infrequent<br>responses | 60      | 46,5     | -13,5 |

The results collected from N are the most straight-forward to discuss. Since the changes in both raw score and % of infrequent responses indicate the decline of productive vocabulary while all 6 measures of fluency indicate the opposite (development of fluency).

Therefore, no discussion about parallel trajectories and thus about the interaction may be required a priori.

The results obtained from the last but not least participant Y are presented in Table 11.

**Table 11**

*Utterance Fluency and Productive Vocabulary Results for Y*

| Y                            |         |          |       |
|------------------------------|---------|----------|-------|
| Measure/ Time:               | Pretest | Posttest | %     |
| SR                           | 0,34    | 0,25     | -26,4 |
| AR                           | 0,25    | 0,22     | -14,3 |
| MLoR                         | 6,53    | 15,23    | 133,2 |
| M.L. of SP within<br>ASU     | 647,3   | 535,06   | -17,3 |
| M.L. of SP between<br>ASU    | 737,97  | 585,61   | -20,6 |
| Number of SP within<br>ASU   | 104     | 29       | -72,1 |
| Raw Score                    | 47      | 53       | 12,7  |
| % of infrequent<br>responses | 41,5    | 44,5     | 3     |

The results obtained from Y are similar to the results from M, that is to say that the slight changes in both raw score (12,7%) and % of infrequent responses (3%) and the slight changes in AR (-14,3%) and mean length of silent pauses within ASU (-17,3%) are not enough to indicate parallel developmental trajectories between productive vocabulary and fluency, because the other four fluency measures showed either moderate (SR (-26,4%) and mean length of silent pauses between ASU -20,6%)) or high changes (MLoR (133,2%) and number of silent pauses within ASU (-72,1%)).

To conclude, based on the results obtained from all five participants, the changes in productive vocabulary did not show parallel developmental trajectories with oral fluency over an academic year in the SA context for any of the participants.

## 5. Discussion

This paper examined the changes in productive vocabulary, fluency and their interaction over an academic year in the SA context. The findings of the study will be discussed in the following paragraphs.

The first research question focused on changes in oral fluency over an academic year in the SA context. Fluency results showed great development in all measures for all participants, which allows us to conclude and answer the Research Question №1 in the following way: the comparison of pretest and posttest interviews in the 7 utterance fluency measurements showed an increase of fluency over a period spent in the SA context.

The findings of this study are in line with the previous research (Huensch & Tracy-Ventura, 2017; Llanes & Muñoz, 2009; Leonard & Shea, 2017; Mora & Valls-Ferrer, 2012; Towell, et al., 1996) showing the development of fluency over the SA experience, despite a number of differences. For instance, the present study differed from the previous ones in the number of participants (5 in comparison to 30, 39, 24 in the studies by Mora & Valls-Ferrer (2012), Leonard & Shea (2017) and Huensch & Tracy-Ventura (2017) respectively) and the size of a corpus (8 minutes of speech from each participant in comparison to 1,5 minute of speech from each participant in Leonard and Shea (2017) or 9,5 hours of talk in Huensch and Tracy-Ventura (2017)). The way of data collection may have also influenced the results. For instance, the participants in both Towell et al. (1996) and Huensch and Tracy-Ventura completed a story retelling task, while the participants in the present study were engaged in a semi-guided interview. When it comes to the fluency measurements, the development of fluency measured by three speed fluency measures (SR, AR and MLoR) was in line with all of the above-mentioned studies, while the results from breakdown fluency measured with an emphasis on location of pauses (mean length of silent pauses within/between ASU and number of silent pauses within/between ASU) also indicated the development of fluency as in the studies by Leonard and Shea (2017) and Huensch and Tracy-Ventura (2017), and confirmed the reliability and need for pause location measures. The three breakdown fluency measures (mean length of silent pauses within/between ASU and number of silent pauses within ASU) reflect not only the general changes in fluency but allow understanding the nature of pauses. Based on Skehan, Foster, and Shum, (2016), one may argue that the



decrease in mean length of silent pauses within ASU and/or number of silent pauses within ASU in this study may be treated as decrease in length or number of pauses in Levelt's (1989) formulation and/or articulation stages which means that the participants stopped less within ASU to search for a forgotten word, grammatical form or pronunciation (micro-planning). Similarly, the decrease in mean length of silent pauses between ASU in this study may reflect the decrease in the length of pauses in Levelt's (1989) conceptualization stage which means that the participants stopped less between ASU to search for new ideas (macro-planning). Needless to say that the findings from the last breakdown fluency measure (the number of silent pauses between ASU) may also be treated as reflection the changes in conceptualization stage, however due to the fact that this measure fully depends on the mean length of silent pauses between ASU, more research should be done in order to establish this measure as a separate one.

The second research question focused on the changes in productive vocabulary over an academic year in the SA context. The productive vocabulary results indicated slight development for 4 (out of 5) participants in two measures, therefore one may conclude and answer Research Question №2 in the following way: the comparison of pretest and posttest interviews in both scoring measures of Lex30 showed an increase of productive vocabulary over a period spent in the SA context.

The study by Fitzpatrick (2012) will be used in this discussion as the most similar study to the present one, since it measured productive vocabulary with the Lex30 test. However, the study by Fitzpatrick (2012) was a case study of only one participant and used a less usual scoring procedure for Lex30 (i.e., number of responses given at each test time, occurrence of affixes, number of cue words responded to, percentage of native speaker-like responses, number of collocational responses etc.) while the present study was based on 5 participants and raw scores and percentage of infrequent responses for Lex30. Additionally, the main goal of the Fitzpatrick study was the measurement of changes in productive vocabulary at various points during the SA period while one of the main goals of the present study was the measurement of changes in productive vocabulary over the SA. Nevertheless, the results from both studies indicated development of productive vocabulary over the SA period. Similarly, the findings of this study are in line with other previous studies which measured productive vocabulary with different tests translation tests (Ife et al., 2000; Jiménez-Jiménez, 2010) or Guiraud's Index of Lexical Richness test (Lara, 2014; Serrano et al., 2012) and also showed the development of productive vocabulary over the SA context.

The final research question focused on the interaction between productive vocabulary and oral fluency over an academic year in the SA context. The present study based on the comparison of productive vocabulary and fluency results obtained from five participants found only slight changes in productive vocabulary while oral fluency showed greater or immense changes over an academic year in the SA context, therefore one may conclude and answer Research Question №3 in the following way: the comparison of the changes in productive vocabulary and fluency in posttest interviews did not show parallel developmental trajectories over an academic year in the SA context.

These findings are contrary to the findings of the only similar study by McManus et al. (2021) who found “significant and long-lasting relationships between fluency and lexis” (p.25) and also to four other studies which were based not in the SA context (De Jong et al., 2012; Koizumi and In’nami, 2013; Uchihara & Saito, 2016). Therefore, one may argue that the results may be influenced by one or several differences between the studies. For example, one of the differences may be the SA context itself, since the three above-mentioned studies found interaction of different degrees between vocabulary and fluency in other contexts. The number of participants may be another difference that could have influenced the results, since the data from previous studies was based on a much greater number of participants (i.e, 56 participants in McManus et al. (2021), 181 in De Jong et al. (2012), 224 in Koizumi & In’nami (2013) and 39 participants in Uchihara & Saito (2016). The method of the vocabulary measurement may also be the reason of contradicting findings, for example, De Jong et al., (2012), Koizumi and In’nami (2013) and McManus et al. (2021) used other productive vocabulary tests: PVLТ, L1-L2 translations and VocD command in CLAN respectively; while although Uchihara and Saito (2016) also used Lex30 tests, they measured fluency only with one measure (speech rate) in comparison to the six measures in the current study. Additionally, the methods of fluency data collection may be another major reason of differences, since the current study was the only one to use the interviews which allow for more natural conversations while the previous studies were mainly based on more artificial monologue tasks: De Jong et al. (2012) used role-play (monologues) speaking tasks, Koizumi and In’nami (2013) used monologues and picture-description tasks, Uchihara and Saito (2016) used a series of picture description tasks and McManus et al. (2021) used picture-based narrative. To conclude, the differences in findings of the current study and previous studies indicate that more research using different methods needs to be done in order to find out more profound findings on the interaction between productive vocabulary and fluency in the SA context.

## **6. Limitations of the Study and Implications for Future Research**

The reported results from this study should be considered in the light of some limitations. First of all, although the corpus of the study (eight minutes from each interview) was quite large, the limitation consisted in the difference of quality in eight minutes for each participant. For example the eight minutes were chosen because it was the shortest speaking time produced by M in Time 1, however although she produced less speech she was more talkative towards the end, while for example N was also more talkative at the end of the Time 1 interview, but since only first eight minutes (of eleven produced by N) were chosen, the last three minutes were not included. Therefore, smaller (or bigger in other corpora) but more balanced corpus may be used in future studies. Secondly, the rules for ASU division were based on Foster et al. (2000) and further developed for the purposes of this study, while there is a need for further development and establishment of more detailed general rules which can be used in future studies. Thirdly, although it was found that the difference between written and oral versions of Lex30 was rather not significant (Clenton, 2010), it would be more logical to use the oral version of the test when researching on the interaction of oral productive vocabulary with other oral skills. Additionally, other productive vocabulary tests may also be used in order to check the interaction with fluency, in other ways, similarly other fluency measures may be used in future research. Finally, the inferential statistics may be applied to show more precise and valid results instead of the descriptive statistics in future studies.

## **7. Conclusion**

The present study has sought to fill gaps and extend the body of research on the relationships between productive vocabulary and fluency in the SA context. This paper investigated how fluency and productive vocabulary changed over the SA period and then compared their results to find out the possible interaction between the two constructs in the SA context.

This paper provided a lot of theory related to vocabulary, fluency and the SA context, and described the methodology of the study, before presenting and discussing the results. The results indicated fluency development over the SA period for all participants in all measures. Productive vocabulary analysis showed less unanimous results, since the development over the SA period was slight in both measures and only for 4 participants. Finally, the comparison of slight changes in productive vocabulary results and greater changes in fluency

results did not show parallel developmental trajectories and therefore the interaction between the two constructs was not found.

Nevertheless, there is still a need for further research which means that more studies using various measurements, data collection methods, participants of different proficiency, etc., should be conducted in order to contribute to the research in this direction and confirm or deny the findings of this study.

## References

- Aitchison, J. (1989). *Words in the mind*. Basil Blackwell Publishers.
- Anderson, R. C., & Freebody, P. (1981). Vocabulary knowledge. In J. T. Guthrie (Ed.), *Comprehension and teaching: Research reviews* (pp. 77–117). International Reading Association.
- Anderson, R. C., & Freebody, P. (1983). Reading comprehension and the assessment and acquisition of word knowledge. In B. Huston (Ed.), *Advances in reading/language research*. JAI Press.
- Arvidsson, K. (2019). « *C'est ça, en fait.* » *Développer l'idiomaticité dans une L2 pendant un séjour linguistique: Trois études sur le rôle des différences individuelles* (Doctoral dissertation). Stockholm University.
- Baba, K. (2002). Test review: Lex30. *Language Testing Update*, 32, 68–71.
- 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.  
<https://doi.org/10.1111/flan.12110>
- Beglar, D. (2009). A Rasch-based validation of the Vocabulary Size Test. *Language Testing*, 27(1), 101–118. <https://doi.org/10.1177/0265532209340194>
- Bialystok, E., & Sharwood-Smith, M. S. (1985). Interlanguage is not a state of mind: An evaluation of the construct for second-language acquisition. *Applied Linguistics*, 6(2), 101–117. <https://doi.org/10.1093/applin/6.2.101>
- Boers, F., & Lindstromberg, S. (2012). Experimental and intervention studies on formulaic sequences in a second language. *Annual Review of Applied Linguistics*, 32, 83–110.  
<https://doi.org/10.1017/s0267190512000050>
- Boers, F., Eyckmans, J., Kappel, J., Stengers, H., & Demecheleer, M. (2006). Formulaic sequences and perceived oral proficiency: putting a Lexical Approach to the test. 37 *Language Teaching Research*, 10(3), 245–261.
- Boersma, P., & van Heuven, V. (2001). Speak and unSpeak with PRAAT. *Glott International*, 5(9-10), 341–347. <https://doi.org/10.1191/1362168806lr195oa>
- 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. <https://doi.org/10.1177/0265532212455394>

- Brecht, R. D., Davidson, D. E., & Ginsberg, R. B. (1995). Predicting and measuring language gains in study abroad settings. In B. F. Freed (Ed.), *Second language acquisition in a study abroad context* (pp. 37–66). Benjamins.
- Brown, G. (1994). Modes of understanding. In G. Brown, K. Malmkjær, A. Pollitt, & J. Williams (Eds.), *Language and understanding* (pp. 10–20). Oxford University Press.
- Brown, J. D. (2003). Promoting fluency in EFL classrooms. *Proceedings of the 2013 JALT Pan-SIG Conference*, 1–12.
- Cambridge English Dictionary. (n.d.). Vocabulary. In Dictionary.Cambridge.org. Retrieved March 22, 2021, from <https://dictionary.cambridge.org/dictionary/english/vocabulary>  
<https://thesestest.ncl.ac.uk/jspui/bitstream/10443.1/2265/1/Cordier%2013.pdf>
- Carroll, J. B. (1967). Foreign language proficiency levels attained by language majors near graduation from college. *Foreign Language Annals*, 1(2), 131–151.  
<https://doi.org/10.1111/j.1944-9720.1967.tb00127.x>
- Chafe, W. L. (1980). The deployment of consciousness in the production of narrative. In W. L. Chafe (Ed.), *The Pear stories : cognitive, cultural, and linguistic aspects of narrative production* (pp. 9–50). Ablex.
- Chafe, W. L. (1985). Some reasons for hesitating. In D. Tannen & M. Saville-Troike (Eds.), *Perspectives on Silence* (pp. 21–30). Ablex.
- Chambers, F. (1997). What do we mean by fluency? *System*, 25(4), 535–544.  
[https://doi.org/10.1016/s0346-251x\(97\)00046-8](https://doi.org/10.1016/s0346-251x(97)00046-8)
- Channell, J. (1988). Psycholinguistic considerations in the study of L2 vocabulary acquisition. In R. Carter & M. McCarthy (Eds.), *Vocabulary and Language Teaching* (pp. 83–96). Longman.
- Chen, J. (2019). What are discourse markers? *Advances in Social Science, Education and Humanities Research*, 268. <https://doi.org/10.2991/sohe-19.2019.1>
- Clenton, J. (2010). *Investigating the construct of productive vocabulary knowledge with Lex30* (Doctoral dissertation). Swansea University.
- Collentine, J. (2004). The effects of learning contexts on morphosyntactic and lexical development. *Studies in Second Language Acquisition*, 26(02), 227–248.  
<https://doi.org/10.1017/s0272263104262040>
- Collentine, J. (2009). Study abroad research: Findings, implications, and future directions. In M. H. Long & C. J. Doughty (Eds.), *The handbook of language teaching* (pp. 218–233). Blackwell.

- Cordier, C. (2013). *The presence, nature and role of formulaic sequences in English advanced learners of French: a longitudinal study* (Unpublished doctoral dissertation). Retrieved from <https://thesestest.ncl.ac.uk/jspui/bitstream/10443.1/2265/1/Cordier%2013.pdf>
- Corpus of Contemporary American English (COCA). (n.d.). <https://www.english-corpora.org/coca/>.
- Cronbach, L. J. (1942). An analysis of techniques for diagnostic vocabulary testing. *The Journal of Educational Research*, 36(3), 206–217. <https://doi.org/10.1080/00220671.1942.10881160>
- Crystal, D., & Davy, D. (1975). *Advanced conversational English*. Longman.
- Cubillos, J. H., Chieffo, L., & Fan, C. (2008). The impact of short-term study abroad programs on L2 listening comprehension skills. *Foreign Language Annals*, 41(1), 157–186. <https://doi.org/10.1111/j.1944-9720.2008.tb03284.x>
- Cucchiaroni, 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.
- D'Anna, C. A., Zechmeister, E. B., & Hall, J. W. (1991). Toward a meaningful definition of vocabulary size. *Journal of Reading Behavior*, 23(1), 109–122. <https://doi.org/10.1080/10862969109547729>
- De Jong, N. H. (2016). Predicting pauses in L1 and L2 speech: the effects of utterance boundaries and word frequency. *International Review of Applied Linguistics in Language Teaching*, 54(2), 113–132. <https://doi.org/10.1515/iral-2016-9993>
- De Jong, N. H., & Bosker, H. R. (2013). Choosing a threshold for silent pauses to measure second language fluency. In R. Eklund (Ed.), *Proceedings of the 6th workshop on Disfluency in Spontaneous Speech (DiSS)* (pp. 17–20). <http://hdl.handle.net/11858/00-001M-0000-0015-0FB8-8>
- 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. <https://doi.org/10.1017/S0142716413000210>
- 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. <https://doi.org/10.1017/s0272263111000489>

- DeKeyser, R. M. (1997). Beyond explicit rule learning: Automatizing second language morpho-syntax. *Studies in Second Language Acquisition*, 19(2), 195–221.  
<https://doi.org/10.1017/s0272263197002040>
- Dewey, D. P. (2004). A comparison of reading development by learners of Japanese in intensive and domestic Immersion and study abroad contexts. *Studies in Second Language Acquisition*, 26(02), 303–327. <https://doi.org/10.1017/s0272263104262076>
- Dewey, D. P. (2008). Japanese vocabulary acquisition by learners in three contexts. *Frontiers: The Interdisciplinary Journal of Study Abroad*, 15(1), 127–148.  
<https://doi.org/10.36366/frontiers.v15i1.223>
- Díaz-Campos, M. (2004). Context of learning in the acquisition of Spanish second language phonology. *Studies in Second Language Acquisition*, 26(2), 249–273.  
<https://doi.org/10.1017/s0272263104262052>
- Ejzenberg, R. (2000). The juggling act of oral fluency: A psycho-sociolinguistic metaphor. In H. Riggenbach (Ed.), *Perspectives on fluency* (pp. 287–313). University of Michigan Press.
- Ellis, R., & Barkhuizen, G. P. (2005). *Analysing learner language*. Oxford University Press.
- Encyclopædia Britannica. (n.d.). *The 18th century*. Encyclopædia Britannica.  
<https://www.britannica.com/topic/art-market/The-18th-century#ref1052123>.
- Erasmus+ annual report 2018 - Publications Office of the EU. Publications Office of the EU.  
<https://op.europa.eu/en/publication-detail/-/publication/7985705e-41b7-11ea-9099-01aa75ed71a1/language-en>.
- Evans, M., & Fisher, L. (2005). Measuring gains in pupils' foreign language competence as a result of participation in a school exchange visit: The case of Y9 pupils at three comprehensive schools in the UK. *Language Teaching Research*, 9(2), 173–192.  
<https://doi.org/10.1191/1362168805lr162oa>
- Fan, M. (2000). How big is the gap and how to narrow it? An investigation into the active and passive vocabulary knowledge of L2 learners. *RELC Journal*, 31(2), 105–119.  
<https://doi.org/10.1177/003368820003100205>
- Fillmore, C. (1979). On fluency. In C. Fillmore, D. Kempler, & W. Wang (Eds.), *Individual differences in language ability and language behavior* (pp. 85–101). Academic Press.
- Fitzpatrick, T. (2007). Productive vocabulary tests and the search for concurrent validity. In H. Daller, J. Milton, & J. Treffers-Daller (Eds.), *Modelling and assessing vocabulary knowledge* (pp. 116–133). Cambridge University Press.



- Fitzpatrick, T. (2012). Tracking the changes: vocabulary acquisition in the study abroad context. *The Language Learning Journal*, 40(1), 81–98.  
<https://doi.org/10.1080/09571736.2012.658227>
- Fitzpatrick, T., & Clenton, J. (2017). Making sense of learner performance on tests of productive vocabulary knowledge. *TESOL Quarterly*, 51(4), 844–867.  
<https://doi.org/10.1002/tesq.356>
- Fitzpatrick, T., & Meara, P. (2004). Exploring the validity of a test of productive vocabulary. *Vigo international journal of applied linguistics*, 1, 55–74.
- Foster, P. (2009). Lexical diversity and native-like selection: The bonus of studying abroad. In B. Richards (Ed.), *Vocabulary studies in first and second language acquisition: The interface between theory and application* (pp. 91–106). Palgrave Macmillan.
- Foster, P. (2020). Oral fluency in a second language: A research agenda for the next ten years. *Language Teaching*, 53(4), 446–461.  
<https://doi.org/10.1017/s026144482000018x>
- Foster, P., & Skehan, P. (1999). The influence of source of planning and focus of planning on task-based performance. *Language Teaching Research*, 3(3), 215–247.  
<https://doi.org/10.1177/136216889900300303>
- Foster, P., & Tavakoli, P. (2009). Native speakers and task performance: Comparing effects on complexity, fluency, and lexical diversity. *Language Learning*, 59(4), 866–896.  
<https://doi.org/10.1111/j.1467-9922.2009.00528.x>
- Foster, P., Tonkyn, A., & Wigglesworth, G. (2000). Measuring spoken language: a unit for all reasons. *Applied Linguistics* Foster, 21(3), 354–375.  
<https://doi.org/10.1093/applin/21.3.354>
- Freed, B. F. (1995a). Language learning and study abroad. In B. F. Freed (Ed.), *Second language acquisition in a study abroad context* (pp. 3–34). Benjamins.
- Freed, B. F. (1995b). *Second language acquisition in a study abroad context*. Benjamins.
- Freed, B. F. (1995c). What makes us think that students who study abroad become fluent? In B. F. Freed (Ed.), *Second language acquisition in a study abroad context* (pp. 123–148). Benjamins.
- Freed, B. F. (2000). Is fluency like beauty in the eyes (and ears) of the beholder? In H. Riggensbach (Ed.), *Perspectives on fluency* (pp. 243–265). University of Michigan Press.
- Freed, B. F., Segalowitz, N., & Dewey, D. P. (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(02). <https://doi.org/10.1017/s0272263104262064>
- Freed, B. F., So, S., & Lazar, N. A. (2003). Language learning abroad: How do gains in written fluency compare with gains in oral fluency in French as a second language? *ADFL Bulletin*, 34(3), 34–40. <https://doi.org/10.1632/adfl.34.3.34>
- Fulcher, G., & Davidson, F. (2007). *Language testing and assessment: An advanced resource book*. Routledge.
- García-Amaya, L. (2018). Utterance fluency in the study abroad context: An overview of research methodologies. In C. Sanz & A. Morales-Front (Eds.), *The Routledge Handbook of Study Abroad Research and Practice* (pp. 181–192). Routledge.
- 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. <https://doi.org/10.1177/0265532210364407>
- Goldman-Eisler, F. (1968). *Psycholinguistics: experiments in spontaneous speech*. Academic Press.
- González, R., & Piriz, A. (2016). Measuring the productive vocabulary of secondary school CLIL students: Is Lex30 a valid test for low-level school learners? *Vigo International Journal of Applied Linguistics*, 13, 31–53.
- Goulden, R., Nation, P., & Read, J. (1990). How large can a receptive vocabulary be? *Applied Linguistics*, 11(4), 341–363. <https://doi.org/10.1093/applin/11.4.341>
- Harley, B. (1995). The lexicon in language research. In B. Harley (Ed.), *Lexical issues in language learning* (pp. 1–28). Benjamins.
- Hazlitt, H. (1993). *The wisdom of Henry Hazlitt*. (H. F. Sennholz, Ed.). Foundation for Economic Education.
- He, C. (2019). A review on diagnostic vocabulary tests. *International Journal of Higher Education*, 8(3), 126. <https://doi.org/10.5430/ijhe.v8n3p126>
- Heaton, J. B. (1966). *Composition through pictures*. Longman.
- Henriksen, B. (1999). Three dimensions of vocabulary development. *Studies in Second Language Acquisition*, 21(2), 303–317. <https://doi.org/10.1017/s0272263199002089>
- Hilton, H. (2008). Connaissances, procédures et production orale en L2. *Acquisition Et Interaction En Langue Étrangère*, 27, 63–89. <https://doi.org/10.4000/aile.4113>

- Hilton, H. (2014). Oral fluency and spoken proficiency: Considerations for research and testing. In P. Leclercq, A. Edmonds & H. Hilton (Ed.), *Measuring L2 proficiency* (pp. 27-53). Bristol, Blue Ridge Summit: Multilingual Matters.  
<https://doi.org/10.21832/9781783092291-005>
- Hirsh, D., & Nation, I. S. P. (1992). What vocabulary size is needed to read unsimplified texts for pleasure? *Reading in a Foreign Language*, 689–696.  
<https://doi.org/10.26686/wgtn.12560417>
- Hu, M., & Nation, I. S. P. (2000). Vocabulary density and reading comprehension. *Reading in a Foreign Language*, 13(1), 403–430. <https://doi.org/10.26686/wgtn.12560354.v1>
- Huensch, A., & Tracy-Ventura, N. (2017). L2 utterance fluency development before, during, and after residence abroad: A multidimensional investigation. *The Modern Language Journal*, 101(2), 275–293. <https://doi.org/10.1111/modl.12395>
- Hunt, K. W. (1970). Syntactic maturity in schoolchildren and adults. *Monographs of the Society for Research in Child Development*, 35(1). <https://doi.org/10.2307/1165818>
- Ife, A., Vives Boix, G., & Meara, P. (2000). The impact of study abroad on the vocabulary development of different proficiency groups. *Spanish Applied Linguistics*, 4, 55–84.
- Ishii, T., & Schmitt, N. (2009). Developing an integrated diagnostic test of vocabulary size and depth. *RELC Journal*, 40(1), 5–22. <https://doi.org/10.1177/0033688208101452>
- Jiménez-Jiménez, A. F. (2010). A comparative study on second language vocabulary development: Study abroad vs. classroom settings. *Frontiers: The Interdisciplinary Journal of Study Abroad*, 19(1), 105–124.  
<https://doi.org/10.36366/frontiers.v19i1.276>
- 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.  
<https://doi.org/10.1111/lang.12084>
- 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). <https://doi.org/10.4304/jltr.4.5.900-913>
- Kormos, J. (1999). Monitoring and self-repair in L2. *Language Learning*, 49(2), 303–342.  
<https://doi.org/10.1111/0023-8333.00090>
- Kormos, J. (2006). *Speech production and second language acquisition*. Lawrence Erlbaum Associates.

- Kormos, J., & Dénes, M. (2004). Exploring measures and perceptions of fluency in the speech of second language learners. *System*, 32(2), 145–164.  
<https://doi.org/10.1016/j.system.2004.01.001>
- Krashen, S. D., & Seliger, H. W. (1976). The role of formal and informal environments in second language learning: A pilot study. *Linguistics*, 14(172), 15–20.  
<https://doi.org/10.1515/ling.1976.14.172.15>
- Kroll, B. (1977). Combining ideas in written and spoken English: a look at subordination and coordination. In E. Ochs & T. L. Bennett (Eds.), *Discourse across time and space. Southern California occasional papers in linguistics*. University of Southern California.
- Lapkin, S., Hart, D., & Swain, M. (1995). A Canadian interprovincial exchange: Evaluating the linguistic impact of a three-month stay in Quebec. In B. F. Freed (Ed.), *Second language acquisition in a study abroad context* (pp. 67–94). Benjamins.
- Lara, A. R. (2014). *Complexity, accuracy and fluency development through study abroad programmes varying in duration* (dissertation). Universitat Pompeu Fabra. Retrieved from [www.tdx.cat/handle/10803/284229](http://www.tdx.cat/handle/10803/284229)
- Laufer, B. (1990). Ease and difficulty in vocabulary learning: Some teaching implications. *Foreign Language Annals*, 23(2), 147–155. <https://doi.org/10.1111/j.1944-9720.1990.tb00355.x>
- Laufer, B. (1992). How much lexis is necessary for reading comprehension? In P. Arnaud & H. Béjoint (Eds.), *Vocabulary and Applied Linguistics* (pp. 126–132). Palgrave Macmillan.
- Laufer, B. (1993). Appropriation du vocabulaire: mots faciles, mots difficiles, mots impossibles. *Acquisition Et Interaction En Langue Étrangère*, (3), 97–113.  
<https://doi.org/10.4000/aile.4895>
- Laufer, B. (1998). The development of passive and active vocabulary in a second language: Same or different? *Applied Linguistics*, 19(2), 255–271.  
<https://doi.org/10.1093/applin/19.2.255>
- Laufer, B. (2005). Focus on form in second language vocabulary learning. *EUROSLA Yearbook*, 5, 223–250. <https://doi.org/10.1075/eurosla.5.11lau>
- Laufer, B., & Nation, P. (1995). Vocabulary size and use: Lexical richness in L2 written production. *Applied Linguistics*, 16(3), 307–322.  
<https://doi.org/10.1093/applin/16.3.307>

- Laufer, B., & Nation, P. (1999). A vocabulary-size test of controlled productive ability. *Language Testing*, 16(1), 33–51. <https://doi.org/10.1177/026553229901600103>
- Laufer, B., & Paribakht, T. S. (1998). The relationship between passive and active vocabularies: Effects of language learning context. *Language Learning*, 48(3), 365–391. <https://doi.org/10.1111/0023-8333.00046>
- Laufer, B., Elder, C., Hill, K., & Congdon, P. (2004). Size and strength: do we need both to measure vocabulary knowledge? *Language Testing*, 21(2), 202–226. <https://doi.org/10.1191/0265532204lt277oa>
- Lennon, P. (1984). Retelling a story in English. In H. W. Dechert & M. Raupach (Eds.), *Second language productions* (pp. 50–68). Narr Verlag.
- Lennon, P. (1990). Investigating fluency in EFL: A quantitative approach\*. *Language Learning*, 40(3), 387–417. <https://doi.org/10.1111/j.1467-1770.1990.tb00669.x>
- Lennon, P. (2000). The lexical element in spoken second language fluency. In H. Riggenbach (Ed.), *Perspectives on fluency* (pp. 43–60). University of Michigan Press.
- Leonard, K. R., & Shea, C. E. (2017). L2 speaking development during study abroad: Fluency, accuracy, complexity, and underlying cognitive factors. *The Modern Language Journal*, 101(1), 179–193. <https://doi.org/10.1111/modl.12382>
- Levelt, W. (1989). *Speaking: From intention to articulation*. MIT Press.
- Llanes, À. (2011). The many faces of study abroad: an update on the research on L2 gains emerged during a study abroad experience. *International Journal of Multilingualism*, 8(3), 189–215. <https://doi.org/10.1080/14790718.2010.550297>
- Llanes, À., & Muñoz, C. (2009). A short stay abroad: Does it make a difference? *System*, 37(3), 353–365. <https://doi.org/10.1016/j.system.2009.03.001>
- MacMillan dictionary. (n.d.). Vocabulary. In *Macmillandictionary.com* Retrieved March 22, 2021, from <https://www.macmillandictionary.com/dictionary/british/vocabulary>
- MacWhinney, B., & Wagner, J. (2010). *Transcribing, searching and data sharing: The CLAN software and the TalkBank data repository*. NCBI. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4257135/>.
- Mcguire, M., & Larson-Hall, J. (2017). Teaching formulaic sequences in the classroom: Effects on spoken fluency. *TESL Canada Journal/Revue TESL Du Canada*, 34(3), 1–25. <https://doi.org/10.18806/tesl.v34i3.1266>
- McLean, S., & Kramer, B. (2015). The creation of a New Vocabulary Levels Test. *Shiken*, 19(2), 1–11.

- McManus, K., Mitchell, R., & Tracy-Ventura, N. (2021). A longitudinal study of advanced learners' linguistic development before, during, and after study abroad. *Applied Linguistics*, 42(1), 136–163. <https://doi.org/10.1093/applin/amaa003>
- Meara, P. (1980). Vocabulary acquisition: A neglected aspect of language learning. *Language Teaching*, 13(3-4), 221–246. <https://doi.org/10.1017/s0261444800008879>
- Meara, P. (1987). *Vocabulary in a second language* (Vol. 2). CILT.
- Meara, P. (1990). Some notes on the Eurocentres vocabulary tests. In J. Tommola (Ed.), *Foreign Language Comprehension and Production* (pp. 103–113). AFinLa Yearbook.
- Meara, P. (1992). *Efl Vocabulary Tests*. Lognostics.
- Meara, P. (1996). The dimensions of lexical competence. In G. Brown, K. Malmkjaer, & J. Williams (Eds.), *Performance and Competence in Second Language Acquisition* (pp. 35–53). Cambridge University Press.
- Meara, P., & Fitzpatrick, T. (2000). Lex30: An improved method of assessing productive vocabulary in an L2. *System*, 28(1), 19–30. [https://doi.org/10.1016/s0346-251x\(99\)00058-5](https://doi.org/10.1016/s0346-251x(99)00058-5)
- Meara, P., & Jones, G. (1987). Tests of vocabulary size in English as a foreign language. *Polyglot*, 8(1), 1–40.
- Meara, P., & Jones, G. (1988). Vocabulary size as a placement indicator. In P. Grunwell (Ed.), *Applied linguistics in society* (pp. 80–87). CILT.
- Meara, P., & Olmos, J. C. (2010). Words as species: An alternative approach to estimating productive vocabulary size. *Reading in a Foreign Language*, 22(1), 222–236. <https://doi.org/10125/66651>
- Melka-Teichroew, F. J. (1982). Receptive versus productive vocabulary: A survey. *Interlanguage Studies Bulletin*, 6, 5–33.
- Melka-Teichroew, F. J. (1997). Receptive vs. productive aspects of vocabulary. In N. Schmitt & M. McCarthy (Eds.), *Vocabulary, description, acquisition and pedagogy* (pp. 84–102). Cambridge University Press.
- Miller, G. A. (1996). *The science of words*. Scientific American Library.
- Milton, J., & Meara, P. (1995). How periods abroad affect vocabulary growth in a foreign language. *ITL - International Journal of Applied Linguistics*, 107, 17–34. <https://doi.org/10.1075/itl.107-108.02mil>

- Mora, J. C., & Valls-Ferrer, M. (2012). Oral fluency, accuracy, and complexity in formal instruction and study abroad learning contexts. *TESOL Quarterly*, 610–641. <https://doi.org/10.1002/tesq.34>
- Nation, I. S. P. (1983). Teaching and testing vocabulary. *Guidelines*, 5(1), 12–25.
- Nation, I. S. P. (1990). *Teaching and learning vocabulary*. Newbury House.
- 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. <https://doi.org/10.3138/cmlr.63.1.59>
- Nation, I. S. P., & Beglar, D. (2007). A vocabulary size test. *The Language Teacher*, 31(7), 9–13. <https://doi.org/10.26686/wgtn.12552197>
- OECD charts a slowing of international mobility growth*. ICEF Monitor - Market intelligence for international student recruitment. (2017, December 4). <https://monitor.icef.com/2017/09/oecd-charts-slowing-international-mobility-growth/>.
- Opper, S., Teichler, U., & Carlson, J. S. (Eds.). (1990). *Impacts of study abroad programmes on students and graduates*. Jessica Kingsley Publishers.
- 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(4), 557-581.
- Pandey, P., & Pandey, M. (2015). *Research methodology: Tools and techniques*. Bridge Center.
- Pawley, A., & Syder, F. H. (1983). Two puzzles for linguistic theory: Nativelike selection and nativelike fluency. In J. C. Richards & R. W. Schmidt (Eds.), *Language and communication* (pp. 191–226). Longman.
- Pawley, A., & Syder, F. H. (2000). The one-clause-at-a-time hypothesis. In H. Riggenbach (Ed.), *Perspectives on Fluency* (pp. 163–199). University of Michigan Press.
- Pica, T., Holliday, L., Lewis, N., & Morgenthaler, L. (1989). Comprehensible output as an outcome of linguistic demands on the learner. *Studies in Second Language Acquisition*, 11(1), 63–90. <https://doi.org/10.1017/s027226310000783x>
- 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. <https://doi.org/10.1177/0265532215579530>

- Qian, D. (2002). Investigating the relationship between vocabulary knowledge and academic reading performance: An assessment perspective. *Language Learning*, 52(3), 513–536. <https://doi.org/10.1111/1467-9922.00193>
- Quirk, R., Greenbaum, S., Leech, G., & Svartvik, J. (1985). *A comprehensive grammar of the english language*. Longman.
- Raupach, M. (1984). Formulae in second language production. In H. W. Dechert, D. Möhle, & M. Raupach (Eds.), *Second language productions* (pp. 114–137). Narr Verlag.
- Read, J. (1988). Measuring the vocabulary knowledge of second language learners. *RELC Journal*, 19(2), 12–25. <https://doi.org/10.1177/003368828801900202>
- Read, J. (1993). The development of a new measure of L2 vocabulary knowledge. *Language Testing*, 10(3), 355–371. <https://doi.org/10.1177/026553229301000308>
- Read, J. (2000). *Assessing Vocabulary*. Cambridge University Press.
- Read, J. (2004). Research in teaching vocabulary. *Annual Review of Applied Linguistics*, 24. <https://doi.org/10.1017/s0267190504000078>
- Rehbein, J. (1987). On fluency in second language speech. In H. W. Dechert & M. Raupach (Eds.), *Psycholinguistic models of production* (pp. 97–105). Ablex.
- Riazantseva, A. (2001). Second language proficiency and pausing: A study of Russian speakers of English. *Studies in Second Language Acquisition*, 23(4), 497–526. <https://doi.org/10.1017/s027226310100403x>
- Richards, J. (1976). The role of vocabulary teaching. *TESOL Quarterly*, 10(1), 77–89. <https://doi.org/10.2307/3585941>
- Richards, J., Platt, J., & Weber, H. (1985). Fluency. *Longman Dictionary Of Applied Linguistics*. Longman.
- Riggenbach, H. (1991). Toward an understanding of fluency: A microanalysis of nonnative speaker conversations. *Discourse Processes*, 14(4), 423–441. <https://doi.org/10.1080/01638539109544795>
- Ringbom, H. (1987). *The role of the first language in foreign language learning*. Multilingual Matters.
- Sasaki, M. (2004). A multiple-data analysis of the 3.5-year development of EFL student writers. *Language Learning*, 54(3), 525–582. <https://doi.org/10.1111/j.0023-8333.2004.00264.x>
- Schmitt, N. (1998). Tracking the incremental acquisition of second language vocabulary: A longitudinal study. *Language Learning*, 48(2), 281–317. <https://doi.org/10.1111/1467-9922.00042>



- Schmitt, N. (2008). Instructed second language vocabulary learning. *Language Teaching Research*, 12(3), 329–363. <https://doi.org/10.1177/1362168808089921>
- Schmitt, N. (2010). *Researching vocabulary: A vocabulary research manual*. Palgrave Macmillan.
- Schmitt, N. (2014). Size and depth of vocabulary knowledge: What the research shows. *Language Learning*, 64(4), 913–951. <https://doi.org/10.1111/lang.12077>
- Schmitt, N., & Meara, P. (1997). Researching vocabulary through a word knowledge framework: Word associations and verbal suffixes. *Studies in Second Language Acquisition*, 19(1), 17–36. <https://doi.org/10.1017/s0272263197001022>
- Schmitt, N., Nation, P., & Kremmel, B. (2019). Moving the field of vocabulary assessment forward: The need for more rigorous test development and validation. *Language Teaching*, 53(1), 109–120. <https://doi.org/10.1017/s0261444819000326>
- Schmitt, N., Schmitt, D., & Clapham, C. (2001). Developing and exploring the behaviour of two new versions of the Vocabulary Levels Test. *Language Testing*, 18(1), 55–88. <https://doi.org/10.1177/026553220101800103>
- Segalowitz, N. (2000). Automaticity and attentional skill in fluent performance. In H. Riggensbach (Ed.), *Perspectives on fluency* (pp. 200–219). University of Michigan Press.
- Segalowitz, N. (2003). Automaticity and second language. In C. Doughty & M. Long (Eds.), *The handbook of second language acquisition* (pp. 382–408). Blackwell Publishers.
- Segalowitz, N. (2010). *Cognitive bases of second language fluency*. Routledge.
- Segalowitz, N. (2016). Second language fluency and its underlying cognitive and social determinants. *International Review of Applied Linguistics in Language Teaching*, 54(2). <https://doi.org/10.1515/iral-2016-9991>
- 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. <https://doi.org/10.1017/s0272263104262027>
- Segalowitz, N., Freed, B., Collentine, J., Lafford, B., Lazar, N., & Díaz-Campos, M. (2004). A comparison of Spanish second language acquisition in two different learning contexts: Study abroad and the domestic classroom. *Frontiers: The Interdisciplinary Journal of Study Abroad*, 10(1), 1–18. <https://doi.org/10.36366/frontiers.v10i1.130>
- Serrano, R., Llanes, À., & Tragant, E. (2011). Analyzing the effect of context of second language learning: Domestic intensive and semi-intensive courses vs. study abroad in Europe. *System*, 39(2), 133–143. <https://doi.org/10.1016/j.system.2011.05.002>

- Serrano, R., Tragant, E., & Llanes, À. (2012). A longitudinal analysis of the effects of one year abroad. *Canadian Modern Language Review*, 68(2), 138–163.  
<https://doi.org/10.3138/cmlr.68.2.138>
- Skehan, P. (2002). Theorising and updating aptitude. In P. Robinson (Ed.), *Individual differences and instructed language learning* (pp. 69–93). John Benjamins.
- Skehan, P. (2003). Task-based instruction. *Language Teaching*, 36(1), 1–14.  
<https://doi.org/10.1017/s026144480200188x>
- Skehan, P. (2009). Modelling second language performance: Integrating complexity, accuracy, fluency, and lexis. *Applied Linguistics*, 30(4), 510–532.  
<https://doi.org/10.1093/applin/amp047>
- Skehan, P., & Foster, P. (1999). The influence of task structure and processing conditions on narrative retellings. *Language Learning*, 49(1), 93–120. <https://doi.org/10.1111/1467-9922.00071>
- Skehan, P., & Foster, P. (2008). Complexity, accuracy, fluency and lexis in task-based performance: a meta-analysis of the Ealing research. In S. Van Daele, A. Housen, F. Kuiken, M. Pierrard, & I. Vedder (Eds.), *Complexity, accuracy, and fluency in second language use, learning, and teaching*. University of Brussels Press.
- Skehan, P., Foster, P., & Shum, S. (2016). Ladders and snakes in second language fluency. *International Review of Applied Linguistics in Language Teaching*, 54(2), 97–111.  
<https://doi.org/10.1515/iral-2016-9992>
- 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), 143–167.  
<https://doi.org/10.1017/s0272263119000421>
- Takala, S. (1984). *Evaluation of students' knowledge of English vocabulary in the Finnish comprehensive school* (dissertation). Institute of Educational Research.
- Tavakoli, P. (2011). Pausing patterns: Differences between L2 learners and native speakers. *ELT Journal*, 65(1), 71–79. <https://doi.org/10.1093/elt/ccq020>
- Tavakoli, P., & Skehan, P. (2005). Strategic planning, task structure, and performance testing. In R. Ellis (Ed.), *Planning and task performance in a second language* (pp. 239–276). John Benjamins.
- Tavakoli, P., & Wright, C. (2020). Measuring Fluency. In *Second Language Speech Fluency: From Research to Practice* (pp. 43–63). Cambridge University Press.

- Tavakoli, P., Nakatsuhara, F., & Hunter, A. (2020). Aspects of fluency across assessed levels of speaking proficiency. *The Modern Language Journal*, 104(1), 169–191.  
<https://doi.org/10.1111/modl.12620>
- Tolstoy, L. (1878). *Анна Каренина* [Anna Karenina]. ЭКСМО.
- Towell, R. (2002). Relative degrees of fluency: A comparative case study of advanced learners of French. *International Review of Applied Linguistics in Language Teaching*, 40(2), 117–150. <https://doi.org/10.1515/iral.2002.005>
- Towell, R., & Dewaele, J.M. (2005). The role of psycholinguistic factors in the development of fluency amongst advanced learners of French. *Focus on French as a Foreign Language*, 210–239. <https://doi.org/10.21832/9781853597688-011>
- Towell, R., Hawkins, R., & Bazergui, N. (1996). The development of fluency in advanced learners of French. *Applied Linguistics*, 17(1), 84–119.  
<https://doi.org/10.1093/applin/17.1.84>
- Tseng, W.-T., & Schmitt, N. (2008). Toward a model of motivated vocabulary learning: A structural equation modeling approach. *Language Learning*, 58(2), 357–400.  
<https://doi.org/10.1111/j.1467-9922.2008.00444.x>
- Uchihara, T., & Clenton, J. (2018). Investigating the role of vocabulary size in second language speaking ability. *Language Teaching Research*, 24(4), 540–556.  
<https://doi.org/10.1177/1362168818799371>
- Uchihara, T., & Saito, K. (2016). Exploring the relationship between productive vocabulary knowledge and second language oral ability. *The Language Learning Journal*, 47(1), 64–75. <https://doi.org/10.1080/09571736.2016.1191527>
- Valls-Ferrer, M. (2011). *The development of oral fluency and rhythm during a study abroad period* (Doctoral dissertation). Retrieved from <http://hdl.handle.net/10803/52064>
- Van Gelderen, A., Schoonen, R., de Glopper, K., Hulstijn, J., Simis, A., Snellings, P., & Stevenson, M. (2004). Linguistic knowledge, processing speed, and metacognitive knowledge in first- and second-language reading comprehension: A componential analysis. *Journal of Educational Psychology*, 96(1), 19–30.  
<https://doi.org/10.1037/0022-0663.96.1.19>
- Webb, S., Sasao, Y., & Ballance, O. (2017). The updated Vocabulary Levels Test: Developing and validating two new forms of the VLT. *ITL - International Journal of Applied Linguistics*, 168(1), 33–69. <https://doi.org/10.1075/itl.168.1.02web>

- Wiese, R. (1984). Language production in foreign and native languages: Same or different? In H. W. Dechert, D. Möhle, & M. Raupach (Eds.), *Second language productions* (pp. 11–25). Narr Verlag.
- Wilkins, D. A. (1972). *Linguistics in language teaching*. Arnold.
- Wood, D. (2006). Uses and functions of formulaic sequences in second language speech: An exploration of the foundations of fluency. *Canadian Modern Language Review*, 63(1), 13–33. doi: 10.3138/cmlr.63.1.13
- Wood, D. (2009). Effects of focused instruction of formulaic sequences on fluent expression in second language narratives: A case study. *Canadian Journal of Applied Linguistics*, 12(1), 39–57.
- Zaytseva, V., Pérez-Vidal, C., & Miralpeix, M. (2018). Vocabulary acquisition during study abroad: A comprehensive review of the research. In C. Sanz & A. Morales-Front (Eds.), *The Routledge Handbook of study abroad research and practice* (pp. 210–224). Routledge.
- Zellner, B. (1994). Pauses and the temporal structure of speech. In E. Keller (Ed.), *Fundamentals of speech synthesis and speech recognition* (pp. 41–62). John Wiley.

### Appendix A

Exemplification of a Completed and Not Coded Lex30 Test by Y in pre-SA Experience

| Cue                   | Responses    |              |              |             |
|-----------------------|--------------|--------------|--------------|-------------|
| <b>1. attack</b>      | bomb         | terrorist    | nightmare    | awful       |
| <b>2. board</b>       | class        | pupil        | hotel        | writing     |
| <b>3. close</b>       | open         | door         | mouth        | shop        |
| <b>4. cloth</b>       | wear         | pants        | textile      | protection  |
| <b>5. dig</b>         | soil         | ground       | agriculture  | earth       |
| <b>6. dirty</b>       | clean        | rubbish      | insalubrious | messy       |
| <b>7. disease</b>     | illness      | sick         | sore throat  | medecine    |
| <b>8. experience</b>  | novelty      | personal     | travel       | awesome     |
| <b>9. fruit</b>       | passionfruit | banana       | vegetable    | pomegranate |
| <b>10. furniture</b>  | rug          | desk         | room         | cupboard    |
| <b>11. habit</b>      | custom       | conservatory | traditional  | time        |
| <b>12. hold</b>       | paper        | secret       | pencil       | tight       |
| <b>13. hope</b>       | magic        | persevere    | stars        | life        |
| <b>14. kick</b>       | mean         | rugby        | football     | ball        |
| <b>15. map</b>        | geography    | travel       | freedom      | Turkey      |
| <b>16. obey</b>       | order        | submission   | respectful   | good        |
| <b>17. pot</b>        | melting      | cultures     | difference   | –           |
| <b>18. potato</b>     | tomato       | vegetable    | yellow       | heavy       |
| <b>19. real</b>       | fake         | realistic    | hard         | suffer      |
| <b>20. rest</b>       | vacation     | good         | breathe      | lay         |
| <b>21. rice</b>       | india        | China        | Japan        | –           |
| <b>22. science</b>    | respect      | awesome      | philosophy   | brain       |
| <b>23. seat</b>       | chair        | down         | movie        | –           |
| <b>24. spell</b>      | word         | sorcerer     | witch        | potion      |
| <b>25. substance</b>  | drug         | toxic        | unknown      | colour      |
| <b>26. stupid</b>     | intelligent  | poor         | meannes      | inferiority |
| <b>27. television</b> | stupid       | games        | politics     | waste       |
| <b>28. tooth</b>      | toothpaste   | fragile      | tongue       | white       |
| <b>29. trade</b>      | commerce     | economics    | exchange     | world       |

**30. window**      door                      room                      house                      wind

---

Note: “-” was used to mark the absence of the response.

## Appendix B

Exemplification of a Completed and Coded Lex30 Test by Y in pre-SA Experience

| Cue                   | Responses    |              |              |             |
|-----------------------|--------------|--------------|--------------|-------------|
| <b>1. attack</b>      | bomb         | terrorist    | nightmare    | awful       |
| <b>2. board</b>       | class        | pupil        | hotel        | WRITE       |
| <b>3. close</b>       | open         | door         | mouth        | shop        |
| <b>4. cloth</b>       | wear         | pants        | textile      | protection  |
| <b>5. dig</b>         | soil         | ground       | agriculture  | earth       |
| <b>6. dirty</b>       | clean        | rubbish      | insalubrious | MESS        |
| <b>7. disease</b>     | ILL          | sick         | sore throat  | medecine    |
| <b>8. experience</b>  | novelty      | personal     | travel       | awesome     |
| <b>9. fruit</b>       | passionfruit | banana       | vegetable    | pomegranate |
| <b>10. furniture</b>  | rug          | desk         | room         | cupboard    |
| <b>11. habit</b>      | custom       | conservatory | traditional  | time        |
| <b>12. hold</b>       | paper        | secret       | pencil       | tight       |
| <b>13. hope</b>       | magic        | persevere    | STAR         | life        |
| <b>14. kick</b>       | mean         | rugby        | football     | ball        |
| <b>15. map</b>        | geography    | travel       | freedom      | –(P/N)      |
| <b>16. obey</b>       | order        | submission   | respectful   | good        |
| <b>17. pot</b>        | MELT         | CULTURE      | difference   | –           |
| <b>18. potato</b>     | tomato       | vegetable    | yellow       | heavy       |
| <b>19. real</b>       | fake         | realistic    | hard         | suffer      |
| <b>20. rest</b>       | vacation     | good         | breathe      | lay         |
| <b>21. rice</b>       | –(P/N)       | –(P/N)       | –(P/N)       | –           |
| <b>22. science</b>    | respect      | awesome      | philosophy   | brain       |
| <b>23. seat</b>       | chair        | down         | movie        | –           |
| <b>24. spell</b>      | word         | sorcerer     | witch        | potion      |
| <b>25. substance</b>  | drug         | toxic        | KNOW         | colour      |
| <b>26. stupid</b>     | intelligent  | poor         | MEAN         | inferiority |
| <b>27. television</b> | stupid       | GAME         | politics     | waste       |
| <b>28. tooth</b>      | toothpaste   | fragile      | tongue       | white       |
| <b>29. trade</b>      | commerce     | economics    | exchange     | world       |

---

|                   |      |      |       |      |
|-------------------|------|------|-------|------|
| <b>30. window</b> | door | room | house | wind |
|-------------------|------|------|-------|------|

---

Note: 1) Lemmatized words were marked by capitalization. 2) Proper nouns were marked “–(P/N)” and excluded from calculations. 3) “–” was used to mark the absence of the response.