

Vocabulary Acquisition in Study Abroad and Formal Instruction: An Investigation on Oral and Written Lexical Development

Victoria Zaytseva

TESI DOCTORAL UPF / 2016

DIRECTORES DE LA TESI

Dra. Carmen Pérez Vidal (Departament de Traducció i Ciències del
Llenguatge, Universitat Pompeu Fabra)

Dra. Imma Miralpeix Pujol (Departament de Filologia Anglesa i
Alemanya, Universitat de Barcelona)

DEPARTAMENT DE TRADUCCIÓ I CIÈNCIES DEL
LLENGUATGE



This dissertation is dedicated to the memory of my grandmother,
Raísa Tkachova (1927-2010).

*Милая моя бабушка, прости меня за то,
что не была рядом с тобой в последнюю минуту.*

Acknowledgements

First and foremost, I would like to express my deepest gratitude to my two supervisors, Dr. Carmen Pérez Vidal and Dr. Imma Miralpeix Pujol, for all of their support and dedication throughout the whole process. Carmen, thank you for encouraging me to pursue doctoral studies in Barcelona and making me part of the SALA research team. I really appreciate your generosity and insight in this project. Imma, I am extremely grateful for your involvement with this study from day one, your expertise, and immediate feedback. The enthusiasm with which you embraced it, your empathy, and advice have meant a world to me.

Without question, I am also greatly indebted to all the members of the SALA project who collaborated in data collection and coding tasks. My sincerest thanks go to Dr. Margalida Valls-Ferrer, Dr. Elisa Barquin, Dr. Pilar Avello, and Dr. Rebecca Lara for their help and advice and whose respective dissertations have been a great source of inspiration for my own writing. Many thanks also to Dr. Joan Carles Mora for allowing me to attend his course on statistics, which contributed greatly to this dissertation. In a similar vein, I would like to acknowledge all the student participants who have taken part in the SALA project over the years.

I would also like to thank the Department of Translation and Language Sciences at the UPF for the financial support I received to spend a quarter studying at Georgetown University and for funding the final stage of this thesis. I am also very much indebted to the ALLENCAM group for the many opportunities I have been given to present my work at international conferences. Likewise, I am particularly grateful for the three-year doctoral scholarship provided by AGAUR (Generalitat de Catalunya), which has enabled me to undertake this project in the first place.

Special thanks are due to Dr. Cristina Sanz for her warm welcome, attention and help during my visit in Georgetown University and to Dr. Lourdes Ortega and Dr. John Norris for kindly allowing me to sit in on their lectures. I would also like to thank Prof. Doreen Metzner, who has been the best host I could possibly imagine and really made me feel at home.

Next, I would like to thank my former and current officemates: Dr. Isabel Tejada, Dr. Estela García Alcaraz, Dr. Itziar Arechederra, Mònica Tarrés, Andrea Bíró, Jennifer Ament, Tamara Vorobyeva and Robert Bailey. You made my time at UPF a truly enjoyable experience. Further thanks to my teaching colleagues, Sonia López and Dr. Laia Mayol, for being exceptionally understanding with my heavy workload and a hectic schedule over the last few months.

On a more personal note, I would like to thank my friend and colleague, Dr. Anna Denissenko, who has shared with me the joys and sorrows of a PhD life and always encouraged me to carry on in this endeavour. Thank you for having come into my life.

Finally, my most heartfelt thanks go to my parents for their faith in me and unconditional love: *Дорогие мама и папа! Спасибо вам за любовь и поддержку; за понимание и веру в меня в самые трудные часы моей жизни; за ваши искренние молитвы. Всем своим достижениям и заслугам я обязана только вам. Мне не хватает слов, чтобы выразить как сильно вы мне дороги и как я вас люблю.* A special mention to my local family: Muchas gracias a mi familia española por todo el apoyo, comprensión y paciencia brindados durante un tiempo tan delicado para mí.

And to Robert, you have lived this with me and have always believed that I could do it. It is impossible to put into words how much your presence has meant to me. There is much that I owe to you.

Victoria Zaytseva

Barcelona, April 2016

Abstract

The present study investigates the impact of two different consecutive learning contexts, formal instruction (FI) at home and a 3-month stay abroad (SA), on second language (L2) vocabulary acquisition in oral and written production. Data were obtained from a group of 30 Catalan/Spanish advanced learners of English before and after each learning period by means of an oral interview and a written composition. These samples were analyzed in terms of quantitative lexical proficiency measures in the domains of fluency, density, diversity, sophistication and accuracy, and through qualitative native-like selections. Baseline data from 29 native speakers of English, elicited through the same tasks, were also used for comparison purposes. Results reveal that SA is particularly beneficial for written productive vocabulary, and less so for oral, and that progress occurs especially in lexical fluency and diversity. FI, in contrast, shows a modest effect on the improvement of oral productive vocabulary and affects namely lexical sophistication. Furthermore, initial level of vocabulary knowledge is found to be a significant predictor of gains.

Resum

Aquest estudi investiga l'impacte de dos contextos d'aprenentatge consecutius diferents, el de la instrucció formal al país d'origen i l'estada de 3 mesos a l'estranger, en l'adquisició de vocabulari de la segona llengua (L2) a la producció oral i escrita. Les dades s'han obtingut d'un grup de 30 estudiants bilingües català/castellà aprenents avançats d'anglès, abans i després de cada període d'aprenentatge mitjançant una entrevista oral i una redacció escrita. Aquestes mostres s'han analitzat amb mesures quantitatives de proficiència lèxica en les àrees de fluïdesa, densitat, diversitat, sofisticació i correcció, i també a través d'una anàlisi qualitativa de la idiomaticitat en la L2. Igualment, s'han recollit dades de 29 nadius d'anglès a efectes comparatius. Els resultats revelen que l'estada a l'estranger té un efecte de millora en el desenvolupament del vocabulari productiu escrit, però no tant en el vocabulari productiu oral, i afecta sobretot la fluïdesa i diversitat lèxiques. El període d'instrucció formal, al contrari, té un efecte moderat en la millora de vocabulari productiu oral i és més notable en la sofisticació lèxica. També es troba que el nivell inicial de competència lèxica explica en gran part els guanys obtinguts.

Table of contents

Acknowledgements.....	v
Abstract.....	vii
Table of contents.....	ix
List of tables.....	xiii
List of figures.....	xv
List of abbreviations.....	xvii
1. Introduction.....	1
2. Literature review.....	7
2.1. Study Abroad.....	8
2.1.1. Study Abroad in the European context.....	9
2.1.2. SA from its inception: a new context for analysis.....	12
2.1.3. The impact of SA on linguistic abilities.....	22
2.1.4. The impact of SA on vocabulary acquisition.....	28
2.1.4.1. Receptive vocabulary.....	29
2.1.4.2. Productive vocabulary.....	33
2.1.4.3. Formulaic language.....	38
2.1.5. SA, initial level, and vocabulary.....	40
2.2. Vocabulary acquisition.....	42
2.2.1. Vocabulary in SLA research.....	43
2.2.2. Vocabulary knowledge.....	47
2.2.3. Measuring vocabulary.....	53
2.2.3.1. Units of measurement.....	54
2.2.3.2. Measuring vocabulary in free writing and speech...	57
2.2.3.2.1. Lexical fluency.....	59
2.2.3.2.2. Lexical density.....	61
2.2.3.2.3. Lexical diversity.....	63
2.2.3.2.4. Lexical sophistication.....	67
2.2.3.2.5. Lexical accuracy.....	71
2.2.3.3. Measuring native-like selection.....	73
2.3. The relationship between writing and speech.....	74
2.4. Summary.....	77
3. The study.....	81
3.1. The SALA project.....	82
3.2. Objectives and research questions.....	84
3.3. Method.....	87
3.3.1. Design.....	87
3.3.2. Participants.....	88

3.3.3. Learning contexts.....	91
3.3.3.1. Formal Instruction (FI).....	92
3.3.3.2. Study Abroad (SA).....	93
3.3.4. Instruments and procedures.....	95
3.3.4.1. Written task.....	96
3.3.4.2. Oral task.....	97
3.3.4.3. General proficiency tests.....	99
3.3.5. Data analysis.....	101
3.3.5.1. Transcription.....	102
3.3.5.2. Methodological decisions.....	103
3.3.5.3. Computational tools.....	107
3.3.5.3.1. VocabProfile.....	107
3.3.5.3.2. D_Tools.....	109
3.3.5.3.3. CLAN.....	110
3.3.5.3.4. CLAWS part-of-speech tagger.....	110
3.3.6. Measures.....	113
3.3.6.1. Lexical fluency.....	113
3.3.6.2. Lexical density.....	116
3.3.6.3. Lexical diversity.....	117
3.3.6.4. Lexical sophistication.....	118
3.3.6.5. Lexical accuracy.....	120
3.3.6.6. Native-like selections (NLSs).....	122
4. Results.....	125
4.1. RQ1: Vocabulary development in oral and written production	126
4.1.1. Longitudinal and context-specific changes in productive vocabulary.....	128
4.1.2. Quantitative comparisons with native speakers.....	137
4.1.3. Qualitative comparisons with native speakers.....	142
4.2. RQ2: Vocabulary gains across contexts and modalities.....	152
4.2.1. Results for vocabulary gains across contexts.....	152
4.2.2. Results for vocabulary gains across modalities.....	156
4.3. RQ3: The role of initial level.....	161
4.3.1. The role of initial vocabulary knowledge.....	162
4.3.2. The role of initial general proficiency.....	170
4.4. Summary.....	175
5. Discussion.....	179
5.1. Vocabulary development in oral and written production.....	179
5.1.1. The effects of FI and SA on productive vocabulary.....	180
5.1.2. Comparisons with native speakers.....	194
5.2. Vocabulary gains across contexts and modalities.....	199

5.2.1. On the learning context.....	200
5.2.2. On the production modality.....	202
5.3. The role of initial level.....	203
5.3.1. Initial vocabulary knowledge.....	204
5.3.2. Initial general proficiency.....	205
5.4. Summary.....	206
6. Conclusions and future research.....	209
References.....	217
Appendices.....	245

List of tables

Table 1	Components of word knowledge (Nation, 2001).....	48
Table 2	Participants' demographics.....	89
Table 3	Learners' mean perceived improvement in EFL proficiency after SA.....	95
Table 4	Lexical fluency measures.....	116
Table 5	Lexical density measures.....	117
Table 6	Lexical diversity measures.....	118
Table 7	Lexical sophistication measures.....	119
Table 8	Lexical accuracy measures.....	122
Table 9	Descriptive statistics for written productive vocabulary.....	129
Table 10	Descriptive statistics for oral productive vocabulary.....	131
Table 11	RM-ANOVAs: Longitudinal changes with context-specific (FI or SA) differences in written and oral productive vocabulary.....	132
Table 12	Independent-samples <i>t</i> -tests comparing written productive vocabulary of learners and NSs.....	139
Table 13	Independent-samples <i>t</i> -tests comparing oral productive vocabulary of learners and NSs.....	140
Table 14	Frequency of use of impersonal forms in written production.....	143
Table 15	Excerpts from learners' and NSs' written compositions illustrating the use of impersonal forms.....	143
Table 16	Frequency of correct and incorrect uses of <i>hold on to</i> in written production.....	144
Table 17	Frequency of use of adverbs modifying adjectives in written production.....	144
Table 18	Examples of adverbs modifying adjectives in learners' and NSs' written compositions.....	145
Table 19	Frequency of topic-related verbs in written production I..	146
Table 20	Frequency of topic-related verbs in written production II.	146
Table 21	Frequency of use of <i>another</i> and <i>new</i> as examples of L1 transfer in written production.....	147
Table 22	Frequency of use of deceptive cognates in oral production.....	148
Table 23	Frequency of use of emphasis in oral production.....	149
Table 24	Frequency of use of delexicalised and specific verb combinations in oral production.....	151
Table 25	Paired-samples <i>t</i> -tests comparing FI vs. SA mean gains in written and oral production.....	153

Table 26	Paired-samples <i>t</i> -tests comparing written vs. oral modality gains in each learning context.....	156
Table 27	Lexical proficiency mean scores of high/low initial level groups in written production.....	163
Table 28	Lexical proficiency mean scores of high/low initial level groups in oral production.....	163
Table 29	Changes in lexical proficiency scores of high/low initial level learners across contexts.....	164
Table 30	Independent-samples <i>t</i> -tests comparing context-specific gains for high/low initial level groups.....	168
Table 31	Correlations between grammar-cloze scores and lexical proficiency measures in written production.....	171
Table 32	Correlations between grammar-cloze scores and lexical proficiency measures in oral production.....	171
Table 33	RM-ANOVAs: Longitudinal changes in grammar-cloze scores.....	174

List of figures

Figure 1	Three main contexts of acquisition (Pérez-Vidal & Juan-Garau, 2011).....	15
Figure 2	Macro and micro-level features of specific learning contexts (Pérez-Vidal, 2014b).....	18
Figure 3	Diagram of SALA data collection.....	88
Figure 4	Learners' contact with English media during SA.....	94
Figure 5	Boxplots of significant context-specific changes by group in written productive vocabulary.....	134
Figure 6	Boxplots of significant context-specific changes by group in oral productive vocabulary.....	136
Figure 7	Circle charts denoting the proportion of different adverbs used in oral production.....	150
Figure 8	Scatterplots of individual FI gains in written vs. oral production modalities.....	158
Figure 9	Scatterplots of individual SA gains in written vs. oral production modalities.....	159
Figure 10	Scatterplots showing significant associations between written and oral vocabulary gains.....	160
Figure 11	Line graphs comparing written lexical development over time of high/low initial level learners.....	165
Figure 12	Line graphs comparing oral lexical development over time of high/low initial level learners.....	166
Figure 13	Line graphs depicting significant context-sensitive changes in written and oral productive vocabulary.....	181

List of abbreviations

ACTFL	<i>American Council on the Teaching of Foreign Languages</i>
Adj	Adjectives
Adv	Adverbs
AH	At Home (learning context)
ALLENCAM	<i>Grup d'Adquisició de Llengües a la Catalunya Multilingüe</i>
AWL	Academic Word List
BNC	British National Corpus
CAF	Complexity, Accuracy, and Fluency
CEFR	Common European Framework of Reference
CHILDES	Child Language Data Exchange System
CLAN	Computerized Language ANalysis (software)
CLIL	Content and Language Integrated Learning
Conj	Conjunctions
CW	Content Words
D	D value (lexical diversity measure)
E/w	Errors per word
EFL	English as a Foreign Language
ESL	English as a Second Language
EVST	Eurocentres Vocabulary Size Test
FI	Formal Instruction
FW	Function Words
GI	Guiraud's Index
GLI	Greco-Latin Index
ICC	Intraclass Correlation Coefficient
IDs	Individual Differences
IM	Domestic immersion (learning context)
L1	First/native language
L2	Second/non-native language
LexE/w	Lexical errors per word
LFP	Lexical Frequency Profile
K1	first 1,000 most frequent English words
K2	second 1,000 most frequent English words
MjSp/w	Major spelling errors per word
MLAT	Modern Language Aptitude Test
MnSp/w	Minor spelling errors per word
NNS	Non-Native Speaker
NS	Native Speaker
N	Nouns
<i>n</i>	Number

OffL	Off-list words
Prep	Prepositions
POS	Part of Speech
Pron	Pronouns
SA	Study Abroad
SAC	Study Abroad Conditions (questionnaire)
SALA	Study Abroad and Language Acquisition (project)
sd	Standard Deviation
SLA	Second Language Acquisition
T1	First data collection time
T2	Second data collection time
T3	Third data collection time
TL	Target Language
TTR	Type/Token ratio
POS	Part of Speech
V	Verbs
VLT	Vocabulary Levels Test
w	Word

Chapter 1

Introduction

The number of university students choosing to study overseas has rapidly expanded over the last few decades. Whether for a few weeks, a semester, or a full academic year, thousands of students embark on study abroad (SA) programmes, placing a great deal of faith in its efficacy and believing that they will return home being proficient target language (TL) speakers. The widespread assumption that SA is the ideal scenario for foreign language learning does not only persist in the general population. Language experts and education authorities also continue to promote academic mobility and actively encourage their students to spend a sojourn abroad.

In the European context, the popularity of SA exchanges has grown in tandem with the initiative of the governing bodies to devise new educational linguistic policies, aimed at fostering multilingualism and foreign language (FL) learning. According to the latest statistics released by the European Commission¹, the 2012-2013 school year set a record by sending nearly 270,000 students overseas through

¹ The report is retrieved from http://ec.europa.eu/education/library/statistics/ay-12-13/facts-figures_en.pdf (European Commission, 2014).

ERASMUS, the most popular SA programme within the European framework and a fundamental engine to cultivate mobility among the youth up until now. This figure represents a substantial increase over the 3,244 students who participated in the exchange for the first time in 1987. Spain, where the current study has been developed, remains one of the top senders and receivers of ERASMUS students, reflecting the growing importance of internationalisation within the European borders.

Against such backdrop, it has become essential to demonstrate the actual benefits of SA on language learning empirically so as to support the European Commission's strategy towards multilingualism and justify its investment in human capital. Yet, it was not until the mid 1990s that scholars took up the challenge to approach SA as an area of investigation in its own right and set out to unravel the real effectiveness of SA on a full range of language skills. After several decades of theoretical and empirical inquiry into the effects of SA on second language acquisition (SLA), the emerging picture is still incomplete (e.g., Collentine, 2009; Collentine & Freed, 2004; DeKeyser, 2007; Freed, 1995a; Freed, Segalowitz, & Dewey, 2004; Llanes, & Muñoz, 2009; Pérez-Vidal, 2014a; Sanz, 2014). Although the authors do coincide that many factors may dictate the relative success of this experience (e.g., the programme design, the length of stay, personality traits, or pre-programme proficiency level), SA does not seem to exert the same impact on all the different linguistic skills across the board. For example, while there is substantial evidence of the positive outcomes as a result of SA on speaking skills (i.e., temporal fluency and pragmatics), research is relatively scarce in others

Introduction

domains, such as writing or phonology, and where it exists, results are somewhat contradictory. Much less is known about the impact of SA on both written and spoken production modalities and no study to date, as far as we are aware, has examined such combination from the perspective of vocabulary.

The purpose of this study is to fill this void, by investigating L2 vocabulary development in written and oral production modes as a result of two learning contexts, a 6-month period of formal instruction (FI) at home in contrast to a 3-month SA period, experienced one after the other. Following a repeated-measures design, we assess the written and oral vocabulary in a robust sample of 30 Catalan/Spanish speaking undergraduate learners of English and formulate a series of research questions based on our review of the literature on study abroad, vocabulary acquisition, and the relationship between oral and written production modalities.

The dissertation that follows is made of six different chapters. Chapter 1 – the present chapter – offers a brief introduction to the topic and sets the stage for the empirical study that follows, presenting the rationale behind the work at hand. Here we also provide an overall organisation of the dissertation together with a brief description of each chapter.

Chapter 2 presents the theoretical background to the study based on a thorough review of the literature surrounding SA and vocabulary acquisition. The first part of the chapter begins with an introduction to SA research, its treatment in the literature as a context of SLA, and

Chapter 1

its impact on the different dimensions of language abilities. In Chapter 2, we review the main findings of SA research, particularly in the domain of vocabulary, referring to its receptive and productive knowledge as well as the mastery of the formulaic language kind. The second part is dedicated to the nature of vocabulary acquisition and presents selected comprehensive frameworks or approaches to the study of vocabulary in SLA research that help to contextualise our study. In this section, we also consider ways in which vocabulary has usually been assessed in free writing and speech and elaborate on the quantitative measures associated with the different domains of lexical proficiency (fluency, density, diversity, sophistication, and accuracy) and qualitative native-like idiomaticity. Also contained in Chapter 2 is an overview of the relationship between writing and speech, as contrasting the two production modalities is of particular interest in the current work.

Chapter 3 is devoted to the empirical study and addresses our objectives and research questions. In this chapter, we situate our study within the institutional setting in which it was conducted and provide a description of the large-scale research project (SALA) that made the elaboration of this dissertation possible. Also detailed in Chapter 3 is the method, which covers the design of the study and provides information on the participants, the learning contexts, data collection instruments and procedures. The process of data analysis is described in Chapter 3 as well, as is the assortment of computational tools used along with the selection of quantitative and qualitative lexical proficiency measures.

Introduction

Chapter 4 reports the results of statistical analyses and is organised around the research questions that motivated the study. Data are analysed longitudinally (to measure development over time) and cross-sectionally (to compare learners' performance to that of NSs at different testing times). Repeated-measures analyses are also used to compare students' lexical proficiency when learning in two different contexts: at-home FI and SA. To meet the main objectives of this study, all writing and speech samples in the corpus are analysed both quantitatively and qualitatively.

The main findings are then summarised and discussed in detail, and in relation to our theoretical background and previous research, in Chapter 5. Finally, Chapter 6 aims to draw overall conclusions from our study. It also acknowledges some limitations that the study may present and identifies possible directions for future research.

Chapter 2

Literature review

This chapter introduces the theoretical background to the present study and is divided into three main sections. In Section 2.1., we offer an introduction to SA research, its trajectory and impact on SLA and place a special emphasis on European mobility, as it is the context in which the empirical study was developed. We also summarise the impact of SA on different linguistic abilities with particular attention given to vocabulary and the ways in which it has been analysed in SA research. Finally, we conclude Section 2.1. with the discussion of the role of initial level proficiency on SA outcomes, and its relation to vocabulary. In Section 2.2., we turn our attention to L2 vocabulary acquisition. We first describe the complex nature of vocabulary knowledge and review some of the underlying concepts related to lexical competence. We then specifically focus on free productive vocabulary and review the different approaches used in SLA to analyse it in writing and speech, as well as the theoretical rationales behind these approaches. We consider measures in the domains of lexical fluency, density, diversity, sophistication, and accuracy, as well as measures of qualitative analysis (native-like selections) which are often argued to be the main components that describe L2 lexical

proficiency (e.g., Bulté, Housen, Pierrard, & Van Daele, 2008; Foster, 2009; Read, 2000; Wolfe-Quintero, Inagaki, & Kim, 1998) and are an integral part of this dissertation. Finally, the chapter concludes with a reflection on the connection between writing and speech (Section 2.3.) and the different implications that each mode of language production may have for SLA and vocabulary development.

2.1. Study Abroad

The popular belief that the best way to learn a foreign language is to spend a period of time in the country where it is spoken natively has been around for so long that it is now almost an article of faith. The romanticised image of SA as a “real soul-melding, transcendent experience that comes with wrapping yourself up in another language²” and soaking it up like a sponge persists to this day, encouraging learners to participate in SA programmes on one hand and requiring scientific support for SA benefits on the other. In order to back up this sort of anecdotal observation and address the gap in our knowledge of the true efficacy of SA, researchers in the field of SLA have sought to investigate SA as a learning context and its implications for language learning. The last several decades of empirical research in the field have led to a massive amount of data, generally finding the context beneficial, but also illustrating the extremely complex nature of the SA learning environment.

² Original source (Librande, 1998) unavailable, but as cited in Kinginger, 2009: 114.

2.1.1. Study Abroad in the European context

In outlining a European perspective on SA and language learning, Coleman (1998) emphasises that in order to understand the SA phenomenon one needs to take into account the historical, geographical and political context of Europe. Coleman goes back to the aftermath of World War II, when the weakened nations were facing an urgent need to secure peace and stability on the continent. After a long history of conflict between the nations, European policymakers became deeply committed to guaranteeing peaceful coexistence and mutual understanding of their citizens. As a response to this commitment, two principal institutions, the European Union (EU) and Council of Europe (COE), were founded not only to promote security and construct a united Europe but also to preserve its cultural and linguistic diversity.

Against such backdrop, mobility became a key feature of subsequent EU policy, enabling the free movement of citizens between EU member states and developing educational initiatives to enhance the learning of foreign language through interuniversity exchange programmes. One of the most well-known and successful programmes of this kind, as we have already noted in the introduction, is the European Community Action Scheme for the Mobility of University Students, popularly known as the ERASMUS programme. Launched in 1987, the programme facilitates the transfer of academic credits between partnering institutions as well as grant support for students who undertake the exchange, typically for a period of 3 to 12 months. While the EU has usually devised ways to promote academic

mobility among European citizens, the COE has been in charge of preserving Europe's cultural and linguistic heritage in all its diversity. Its major contribution to language learning to date has been the drafting of the Common European Framework of Reference for Languages (CEFR), which has since been used as a key tool for standardizing language proficiency tests and a guide for preparing curricula for language instruction³. Within this context of promoting multilingualism, positioned at the heart of European education programmes, knowledge of additional foreign languages other than one's own mother tongue was the new goal. In fact, it was in 1995 that the 'formula 1 + 2' – implying the mastery of L1 plus knowledge of two other languages – was suggested for the first time laying the foundation for subsequent educational and linguistic policies (Pérez-Vidal, 2015).

In such a scenario, as Pérez-Vidal (2015: 3) notes, the ERASMUS scheme established itself as “one of the most successful programmes for cross-border academic experiences at the HE [Higher Education] level.” While in most European universities SA tends to be optional, spending a period abroad has long been a compulsory requirement of language related studies in the UK and is closely integrated into their degree structure (Coleman, 1998, 2005). The mandatory nature of SA (as a part of their home-based degree) is also shared by the institution where the present study has been conducted and is unique in the Spanish university context (Beattie, 2014).

³ Source: http://www.coe.int/t/dg4/linguistic/cadre1_en.asp.

The European model of student mobility and demographics related to L2 learning differs in important ways from the SA in North America. One substantial difference is, for instance, the design of the programmes. While the North American SA model generally envisages “the short-term transfer of cohesive groups of American students to a different geographical base, where they may benefit from formal (classroom) and informal (naturalistic) language learning, but without necessarily abandoning an American educational framework and academic/administrative support,” in the European SA experience, “the emphasis has historically been on individuals or at most small groups living independently for a relatively long time in a totally L2land context, relying on local social, academic and institutional support” (Coleman, 1997: 1). Kinginger (2009: 16) speaks of a ‘core dilemma’ that American language learners face abroad, as they are encouraged to maintain strong ties with their compatriots while also trying to blend in with the locals.

American students also differ greatly from their continental European counterparts in the pre-departure proficiency level. Whereas the former typically embark upon their SA with relatively modest L2 proficiency, the sociolinguistic reality for Europeans is rather different: they normally have advanced skills in the TL prior to arrival (generally 10 years of classroom instruction in English) and previous experience of travelling abroad (Coleman, 1998; Sanz, 2014). Thus, it is important to keep these differences in mind, as the earliest contributions to SA research come from North American databases, whereas studies

focusing on continental European students are relatively recent⁴. We now turn our focus to the trajectory of SA research, its treatment in the literature as a learning context and the different context-internal and external parameters that describe it.

2.1.2. SA from its inception: a new context for analysis

In his brief overview of the research on SA, Collentine (2009) defines two different periods. The first period can be traced to the efforts of researchers examining factors that best predict foreign language ability. It dates back to the early 1960s with the emergence of macro studies assessing benefits of SA on overall L2 abilities largely through broad measurement instruments (Brecht, Davidson, & Ginsberg, 1990; Carroll, 1967; Meara, 1994; Willis, Doble, Sankarayya, & Smithers, 1977, among others). The second period is marked by the year 1995 with Barbara Freed's landmark publication *Second language acquisition in a study abroad context*, a milestone that represents a turning point in SA research. Although optimistic about the linguistic benefits of SA, research in the early 1960s was characterised by several shortcomings, with most studies heavily relying on non-validated instruments or self-reports, involving one testing time, or lacking comparative data (Carroll, 1967; Meara, 1994; Willis et al. 1977). In this sense, Freed's (1995a) volume, consisting of an edited collection of papers, was the first attempt to synthesise SA research, giving rise to systematic investigation and setting the course for future studies in the area. By

⁴ More recently, there has been a growing percentage of SA students across the globe originating from Asian countries (Collentine, 2009; Kinginger, 2009), which represents an important pocket of research in the field of SA.

displaying a number of limitations in scope and design, Freed and her colleagues succeeded in framing SA within the SLA research agenda, suggesting to regard SA as a means of studying the impact of 'learning context' on acquisition. They also called into question the superiority of the SA context over other learning environments, especially in the area of morphosyntax.

In the years following Freed's publication, several noteworthy compilations have appeared expanding in the diversity of populations, languages and approaches under scrutiny. Some of them have concentrated on American learners going abroad; with the special issue of *Studies in Second Language Acquisition (SSLA)* in 2004 comparing the outcomes of SA with formal classroom instruction (FI) and domestic immersion (IM); Paige, Cohen, Kappler, Chi, & Lassegard (2002) assessing language and culture learning strategies during SA, DuFon and Churchill (2006) investigating the acquisition of pragmatic and sociolinguistic competence; DeKeyser (2007) addressing the role of practice from a cognitive approach; or Kinginger (2009; 2013) summarizing SA knowledge base and concentrating on a sociocultural perspective. Others set out to investigate SA in the European context, with Coleman (1997, 1998) offering an extensive historical review of SA in Europe and its impact on intercultural competence; Regan, Howard and Lemée (2009) examining the acquisition of sociolinguistic competence in Irish students placed in francophone countries, Jackson (2008) raising the issue of identity and intercultural communication, or Perez-Vidal's (2014) up-to-date volume dealing with EFL acquisition in subsequent FI and SA contexts by Spanish-Catalan bilingual students. More recently, with the emergence of information and

communication technologies, SA research has taken a new direction examining the role of social networks on foreign language development, as SA in the age of Facebook is not what it used to be years ago (Kinginger, 2013; Mitchell, 2015). Before presenting a brief account of these studies and summarizing its findings (see Section 2.1.3.), we first define the SA as a learning context and describe the different features that characterise it.

SA as a learning context

It is generally accepted among SLA researchers that one of the crucial variables influencing language acquisition is the context of learning, an idea that has become of ever-greater interest in the field of SA research (Collentine & Freed, 2004; Collentine, 2009; Freed, 1995a). As early as the 1970s Dell Hymes noted that “the key to understanding language in context is to start not with language but with context,⁵” suggesting that in order to understand the complexity of language acquisition in a specific learning environment it is important to appreciate its context, in the first place, and then “systematically relate the two.”

A relatively recent context-sensitive account of language acquisition is that of Batstone (2002), who uses the terms *learning context* and *communicative context* to distinguish formal from naturalistic acquisition settings. Batstone’s dichotomy roughly echoes Krashen’s (1976) separation between learning and acquisition or Ellis’s (1994) later

⁵ Original source unavailable, but as cited in Collentine and Freed, 2004: 153.

learning contexts and refers to intensive programmes where learners spend all of their time studying the L2 without leaving the AH L1 setting. One variation of immersion in the European strategy towards multilingualism is the Content and Language Integrated Learning (CLIL) approach. Although somewhat different from IM, CLIL follows the same premise as IM and also takes place in the home country (Muñoz, 2007; Ruiz de Zarobe, 2007). Finally, SA is an example of rather a naturalistic context and is a priori characterised by a massive exposure to authentic input and potentially unlimited opportunities for TL practice and interaction in a variety of real-life situations. Juan-Garau (2012: 226) considers the SA learning environment as “an opportunity to enhance both quantity and quality of input and interaction causing acquisition to take place.” Although the differences between one context and another are not always clear-cut, Collentine (2009), following Batstone’s distinction, notes that FI tends to favour learning contexts, IM combines both types of contexts, and SA presumably provides more opportunities for practising the L2 in communicative contexts, yet classroom-based language learning remains an integral part of most exchange programmes. For this reason, Collentine and Freed (2004) regard SA as both in-class and out-of-class language learning and define it as “a hybrid communicative-learning context” (p. 156). In our study, we define SA as a “form of international mobility which takes place within a student’s programme of study in higher education” (Carbonell, 2011, as cited in Coleman, 2013: 22) on a markedly temporary basis.

Characterisation of SA

In her critical reading of SA research, Kinginger (2009) remarks that an initial challenge to researchers taking an interest in the qualities of the SA experience is “the immense variety of programs, host communities, and students involved” (p. 115). There is a complex mix of variables playing an active role on L2 acquisition, which has been the subject of investigation in several studies of SA. Individual characteristics and personality traits can potentially influence linguistic outcomes just as the specific SA programme design or the amount and type of contact with the TL. Given the highly complex nature of SA experience, it is no wonder that any generalisation about SA benefits has to be made with great caution (Huebner, 1998). As Freed and colleagues (2004b) put it, “... it is not *the learning context per se* that promotes various types of learning but rather, [...] the nature of the interactions, the quality of the experiences, and the efforts made to use the L2 that render one context superior to another with respect to language gain” (p. 298, emphasis added). Ultimately, it is the learner himself who will have the final say on his predisposition to enjoy each and every opportunity afforded in the SA environment.

In order to frame the mix of variables at play in the SA learning context, Pérez-Vidal (2014b) identifies the following parameters: macro-level features, micro-level features and the architecture of SA programmes – alternatively, exchange programme design (see Figure 2).

Figure 2. Macro and micro-level features of specific learning contexts from Pérez Vidal (2014b), p. 22.



Following the authors, the macro-level features of SA are those elements which revolve around the context and can be defined in terms of input, output, interaction and meaning-oriented cognitive processes. In contrast to FI, in which input is limited and restricted to the classroom, as we have already mentioned, learners on a residence abroad programme are by and large immersed in the TL and culture with potentially limitless amounts of input and output, as well as ample opportunities for practice and interaction. According to classic theories in SLA such as the *Input Hypothesis* by Krashen (1985), the *Output Hypothesis* by Swain (1985, 1993), and the *Interaction Hypothesis* by Long (1996) SA is an optimal context for L2 acquisition. One needs a great deal of meaningful input, with resulting implications for intake, in order to acquire an L2 (Krashen), which is especially boosted when

being immersed in the L2 environment. Successful attainment in L2 also relies on output, an equally essential component for the improvement of language proficiency, with SA being the best learning context that “pushes the learner to produce” unlike any other, Sanz (2014: 2). Another learning condition that SA can provide, which is a considerable advantage over the FI context, is opportunities for interaction. From the interactionist perspective (Long), L2 acquisition takes place by means of specific cognitive mechanisms (e.g., noticing, awareness or attention), which are activated during ‘negotiation of meaning’ while practising language in conversation (Gass & Mackey, 2006, 2007). To put it another way, the interaction with expert and NSs of the TL will enable learners to take on different social roles in a myriad of everyday situations (e.g., going shopping, doing the laundry, ordering lunch, or getting tickets) and engage them in a series of implicit meaning-oriented processes, conducive to potential gains in the TL competence. Under these circumstances, language learning will take place incidentally, that is, in the absence of consciousness or ‘awareness at the point of learning’ (Ellis, 1994; Hulstijn, 2006; Schmidt, 1994: 2). In the FI context, in contrast, there is an explicit instructional treatment of language, requiring learners’ attentional focus on the form of language rather than on meaning. As regards the quantity and quality of the type of practice available while abroad, DeKeyser (2007) posits that SA is ideally suited for the automatization stage, best explained from the perspective of skill acquisition theory (Anderson, 1993), which traditionally distinguishes three stages: (1) declarative knowledge, (2) proceduralisation, and (3) automatization.

Focusing on the contrast between FI and SA, Pérez-Vidal (2014b) summarises DeKeyser's approach as follows:

[...] declarative knowledge requires repeated careful, slow and deliberate production of instances of rules, something which can take place in the classroom. However, large amounts of practice are not required for declarative knowledge to become procedural knowledge; feedback, in contrast, is required. The ideal context for this to happen is still the *classroom*. In contrast, the process from such early proceduralization to automatization [...] requires a great many instances of practice, something which can easily take place while *abroad*, because, as explained, the sociolinguistic context makes it possible, provided learners manage to avail themselves of the opportunities for such a practice which the SA context offers. (p. 24)

Despite these contextual characteristics positioning SA at the top of the L2 learning setting hierarchy, learners do not benefit from the SA experience uniformly. Gains in language competence as a result of SA may also be conditioned by individual variables, which leads us to consider the second parameter of the framework, the SA micro-level features.

The micro-level features of SA, as we have just noted, refer to learner individual differences (IDs). These may include learners' affective factors (attitude, motivation and beliefs), their language aptitude and learning strategies, as well as intercultural awareness. Individual variation is known to play a key role in the relative success of L2 attainment and to largely account for the differential SA outcomes among learners (e.g., Grey, Cox, Serafini, & Sanz, 2015). Pérez-Vidal (2014b) point to attitude and motivation as the variables that are best linked to the ability for establishing quality contact with locals, as “those learners who engage in more interactional encounters should

be in a better position to learn” (p. 25). Another key individual variable that has attracted significant attention in the SLA community and proved to be a robust predictor of success in mastering an L2, even in SA contexts, is language aptitude (Anderson, 2014; see also Dörnyei & Skehan, 2003; Robinson, 2002, for further reviews). In sum, it seems that IDs are ‘a big issue’ in SA, presumably even more so than in FI, because of the difficulty to control how learners spend their time, who they talk to, how often, and about what (DeKeyser, 2014: 320).

Finally, the programme architecture or exchange programme design makes up the last parameter at play in the SA learning context. The success of SA can be equally predicted by factors such as the length of the programme, accommodation, and living conditions while abroad, or employment opportunities. Similarly, issues specific to the academic programme from the pre-departure planning right through to SA completion must also be taken into consideration. Thus, the degree of pre-SA preparation, for instance, may vary from one institution to another, ranging from a brief meeting with academic personnel to having to attend a full preparatory module for credit. Also, students may (or may not) be required to have a specific level of language proficiency on enrolment and undertake the exchange at different points in the academic curriculum (either preceding or following an FI period). Last but not least, the type of academic assignments and student workload during the period abroad can be subject to variation, as well as the follow-up and de-briefing conditions.

Having revised the different factors at work in learning a language over an SA period, we may conclude that, together, the macro- and micro-level features and the programme design make up a complex web of variables that play into the efficacy of SA in the ultimate L2 attainment to a great degree.

2.1.3. The impact of SA on linguistic abilities

Over the last past decades, the research stockpile on SA has expanded in scope and scale, providing new insights into the nature of the SA experience and SLA: from the very early attempts to analyse multiple skills through large-scale projects (Brecht et al., 1995; Carroll, 1967; Coleman, 1996; Meara, 1994; Willis et al., 1977, among others) to subsequent studies zooming in on more specific areas by either using more finely tuned instruments or adopting rather narrow measures (DeKeyser, 2014: 319). In the following, we will bring together the most important empirical findings in SA research, summarising the impact of SA on linguistic competence from the early multiple-skill reports to most recent and rigorous contributions (especially after Freed's publication) focusing on specific skills. We dedicate subsection 2.1.4. to the review of SA literature dealing with vocabulary acquisition, as it is of particular interest in the empirical study to follow.

One early widely-cited macro study to offer unqualified support for SA research is that of Carroll (1967). Although not directly concerned with SA as a learning context and based exclusively on the results of test scores, the study was an ambitious one, aimed at investigating

global FL learning from a broad perspective. In his study, Carroll assessed 2,872 US seniors majoring in five foreign languages (Spanish, French, German, Russian, and Italian) on the four language skills – reading, writing, listening, and speaking – in order to determine their average attainment of FL proficiency upon graduation together with factors associated with variation in performance. He named time spent abroad as “one of the most potent variables” predicting L2 proficiency noting that “even a tour abroad, or a summer school course abroad, is useful, apparently, in improving the student’s skill” (p. 137). In addition to Carroll’s solid claim about the effectiveness of SA, his secondary finding was that SA especially favoured weaker learners.

Another noteworthy contribution to SA research came from a collaborative project sponsored by the American Council of Teachers of Russian (ACTR) and the National Foreign Language Center (NFLC). The project, consisting of individual and joint publications (Brecht et al., 1990; Ginsberg, 1992), was among the first to statistically address the development of FL ability in an SA context. More specifically, it focused on predictors of language gain during a four-month SA period involving 658 learners of Russian in the former Soviet Union. Following a pre- and post-test design with a multiple-skill approach, researchers assessed learners’ speaking proficiency via the Oral Proficiency Interview⁶ (OPI), as well as their listening and

⁶ Designed by the American Council on the Teaching of Foreign Languages (ACTFL), the OPI is a “standardised procedure for the global assessment of functional speaking ability. It is a face-to-face or telephonic interview between a certified ACTFL tester and an examinee that determines how well a person speaks a language.” Source: <http://www.actfl.org/professional-development/certified-proficiency-testing-program/testing-proficiency>.

reading abilities. They found that SA was an effective mode for FL learning, and that knowledge of other languages, gender, pre-programme proficiency level and language aptitude, as measured by the Modern Language Aptitude test (MLAT) proved to be reliable predictors of linguistic gains.

During the same period in the 1990s, parallel projects were emerging in Europe to also document language learning in the SA context. In the British context, Meara (1994) considered improvement in four skills by 586 British university students on a national survey, the Nuffield Modern Languages Enquiry. Although based on learners' subjective perceptions of how they had improved rather than empirical assessments of language development, the study confirmed to a large extent previous findings (Willis et al., 1977) attributing growth in the area of spoken language to the 'year abroad.' Moreover, according to these self-reports, the majority of respondents indicated that their speaking and listening skills had improved more than their reading or writing abilities. As part of his European Languages Proficiency Survey, Coleman (1996) also looked at the effects of SA, drawing on data from a large pool of subjects ($n=18,825$) in a cross-sectional study, built around a C-test. He found that SA enhanced general language proficiency of British students (SA students scored substantially higher than those who had yet to study abroad) and also reported a significant association between length of stay and scores on the C-Test.

In the years following the aforementioned large-scale projects evaluating general language proficiency, SA scholars have refined and

multiplied the measures of assessment focusing on more specific skills and providing new findings. Most of these studies have looked into the development of speaking skills, pointing to oral fluency as the biggest beneficiary in comparison with other linguistic areas. For example, major gains were observed for fluency when it was measured through temporal aspects of L2 speech production (speech rate, hesitation, pauses) or based on judges' evaluations of perceived fluency (Freed, 1995b; Towell, Hawkins, & Bazergui, 1996; Segalowitz & Freed, 2004). Less progress was found for fluency when the construct was operationalised as quantity of speech, as measured by total number of words produced (Freed, 1995b; Freed, So, & Lazar, 2003; Freed, Segalowitz, & Dewey, 2004; we will return to the review of lexical fluency in Section 2.1.4.2.). For instance, of the seven components of fluency⁷, only rate of speech (words per minute) proved significant in Freed's (1995b) study comparing AH and SA groups, indicating that students who had spent a semester abroad spoke at a significantly faster rate than did those who stayed at home. Similar findings were reported in Towell et al. (1996) who found significant improvement in fluency (as measured by speaking rate) in advanced learners of French after a six-month period abroad. Oral fluency was also found to benefit in SA over AH contexts in a study by Segalowitz and Freed (2004), involving American undergraduates studying Spanish in Spain. According to the authors, SA participants improved significantly from the pre-test to the post-test in terms of oral fluency (measured by speech rate and hesitation phenomena) and speaking turn based on OPI.

⁷ The other six components measured amount of speech, filled and unfilled pauses, length of fluent speech runs, repairs, and speech dysfluencies.

A handful of more recent European-based studies, examining EFL development in advanced Catalan-Spanish learners undertaking an ERASMUS exchange, provided similar evidence for SA as a potential oral fluency booster (Lara, 2014; Pérez-Vidal & Juan-Garau, 2011; Valls-Ferrer, 2011; Valls-Ferrer & Mora, 2014). For example, Valls-Ferrer, using oral interview samples, found that a 3-month SA period was highly beneficial for learners' fluency and rhythm development, reporting significant improvement in three dimensions (utterance fluency, perceived fluency, and rhythm). In a follow-up study using the same population, Valls-Ferrer & Mora (2014) confirmed that SA had a significant positive effect on both speed fluency measures and breakdown fluency measures. They also found that lower initial fluency levels and a larger amount of contact with the TL were associated with greater gains during SA.

Unlike temporal aspects of oral fluency, studies focusing on phonological development under SA conditions provide inconclusive findings, most of them reporting little advantage for the SA context (Avello, 2013; Diaz-Campos, 2004; Mora, 2008), although with some exceptions (Nagle, Morales-Front, Moorman, & Sanz, *in press*).

As for listening comprehension skills, there is a paucity of studies focusing explicitly on this area, and those that do have reported contrasting results (Beattie, 2008; Cubillos, Chieffo, & Fan, 2008; Tanaka & Ellis, 2003). Likewise, very little attention has been given to reading-related ability, yet results in this skill (although based on self-assessment) have pointed to improvement as a result of SA (Dewey, 2004; Hayden, 1998).

Surprisingly, writing has also been relatively under-represented in the SA literature and where it exists, results are somewhat contradictory. While Sasaki (2007, 2009) found moderate differential gains in written production between the SA and AH groups, Freed and colleagues (2003) failed to show a clear superiority of SA in written fluency. More recently, Barquin (2012) examined writing development in Catalan-Spanish learners of English, using qualitative and quantitative measures. She found that learners made considerably more progress after the SA experience than after the FI context at their home university (we will reconsider Freed et al. and Barquin's studies in Section 2.1.4.2, as some of the measures used in their studies targeted various aspects of lexical proficiency).

As regards grammatical competence, results are controversial and do not consistently show greater gains after spending a period abroad. Although some studies document a beneficial impact of SA on grammatical accuracy (Howard, 2005, 2006; Isabelli & Nishida, 2005), others find no substantial differences between FI and SA contexts (DeKeyser, 1991, 2010; Isabelli-García, 2010) or even report superior gains for the AH FI setting, especially for discrete grammatical forms (Collentine, 2004), with rather limited SA effects on morphosyntactic system restructuring (Howard, 2001). Despite these discrepancies, Churchill and DuFon (2006: 9) suggest to regard these findings not as conflicting results, but rather as “further evidence that grammatical development patterns are tremendously complex, vary considerably depending on the linguistic feature highlighted in the research (and quite probably the method used to collect the data) and are undoubtedly confounded by learner initial proficiency.”

Finally, sociolinguistic and pragmatic dimensions of language learning abroad have also been investigated, providing evidence of positive outcomes as a result of SA on learners' communication strategies, particularly in informal interaction. Examples of recent studies are Dewaele (2004) examining informal language use, Grieve (2011)⁸ tackling the acquisition of pragmatic markers for vagueness, or Barron (2006) and Kinginger (2008) dealing with the use of L2 address system.

Here, we have attempted to summarise the main findings of SA research in the different domains of linguistic competence. We reserve the review of SA effects on vocabulary development for the following subsection, as it is the focus of the empirical study to follow in the coming chapters and therefore merits special attention.

2.1.4. The impact of SA on vocabulary acquisition

Vocabulary knowledge has often been reported as one of the areas that noticeably improves in the SA context. However, the very pioneering studies on the effects of periods abroad on vocabulary either involved small sample sizes (DeKeyser, 1991; Lennon, 1990) or were simply based on impressionistic observations of such improvements (Davie, 1996). Davie comments on “marked improvement as a result of the year abroad in vocabulary” (p. 75), as inferred from learners' perceptions that their vocabulary had improved, rather than from empirical measures analysing the

⁸ Original source unavailable, but as cited in Kinginger, 2013.

improvements. Later studies on vocabulary acquisition in the SA setting have primarily focused on the domain of receptive vocabulary knowledge, i.e., understanding the meaning of words (Dewey, 2008; Ife, Vives Boix, & Meara, 2000; Jiménez Jiménez, 2010; Milton & Meara, 1995), or analysed free written and oral productive vocabulary, i.e., retrieving words to be used in appropriate situations without external support, generally from the perspective of complexity, accuracy and fluency (CAF) (Barquin, 2012; Lara, 2014; Llanes & Muñoz, 2009; Pérez-Vidal, Juan-Garau, Mora, & Valls-Ferrer, 2012). In recent years, with the growing recognition among applied linguists of the relevance of multi-word units in lexical knowledge, some studies have examined vocabulary from the perspective of formulaic language (Foster, 2009; Foster, Bolibaug, & Kotula, 2014; Siyanova & Schmitt, 2008). In the following subsections, we will review the most prominent SA studies on receptive and productive vocabulary knowledge as well as formulaic language.

2.1.4.1. Receptive vocabulary

One of the first oft-cited and groundbreaking studies on vocabulary growth during SA is that of Milton and Meara (1995). In their study on EFL receptive vocabulary, the authors assessed 53 European exchange students upon entry to a British university and six months later. Students' vocabulary knowledge in English was measured through the Eurocentres Vocabulary Size Test (EVST), a computerised Yes/No test which uses a graded sample of words at different frequency levels and makes an estimate of learners' receptive *vocabulary size* (also called *vocabulary breadth*). Results showed that on

Chapter 2

average participants acquired new vocabulary nearly five times faster during SA than at home, implying a growth rate of over 2500 words per year – an estimate of L1 vocabulary growth in English-speaking adolescents (Nagy & Herman, 1987, as cited in Milton & Meara, 1995). The authors also found that not all students benefited equally from the SA experience and reported “a clearly marked tendency for students with small initial vocabularies to make a great deal of progress” as opposed to students with the largest starting vocabularies, who, on the contrary, produced the smallest changes (p. 25).

Five years later, receptive vocabulary development was re-examined in a study by Ige and colleagues (2000) conducted among 36 intermediate and advanced British students spending 1 or 2 semesters (4-8 months) away in Spain. The instruments used in their study were a translation test, aimed at measuring their vocabulary size, and the Three Word Association Test (A3VT), permitting assessment of subjects’ lexical organisation knowledge. According to the authors, A3VT was used to overcome the ceiling-effect problem associated with the conventional vocabulary size tests and to capture more demanding aspects of vocabulary knowledge at advanced levels of proficiency. Although the study, with a pre-test/post-test design, was constrained by the lack of a control group, as acknowledged by the authors, the researchers also looked into the amount of progress made by learners differing in initial proficiency levels and length of stay during SA. Results revealed considerable lexical gains during the SA period for both proficiency groups in both types of test in both types of tests and for both proficiency groups, contrary to Milton and Meara’s (1995) findings, which had suggested greater improvement for low-level students. Data

Literature review

also pointed to a trend toward larger vocabulary gains as a result of a longer stay, that is, participants who went on a two-semester SA improved their lexical knowledge 3 times more than those who went away for only one semester.

In order to corroborate the results from Ife et al.'s study and provide more reliable findings by using a more rigorous design, Jiménez Jiménez (2010) examined the development of lexical knowledge in both FI and SA contexts. He collected data from 51 North American students enrolled at different courses of Spanish at their home university, and 30 American graduates having undertaken an SA programme by using a translation test and a word association task. The results of the study indicated that classroom instruction did not foster vocabulary development either in size or in depth of lexical knowledge, unlike the SA context, where learners experienced gains at both levels. The author concludes that “deeper level of vocabulary knowledge is more likely to be acquired in study abroad contexts since classroom instruction does not seem to offer the appropriate elements to trigger its development” (p. 122).

A study by Dewey (2008) also provided evidence for significant benefits of SA context over AH classroom instruction. The researcher set out to investigate vocabulary acquisition in 56 intermediate-level students learning of Japanese in three contexts, FI AH ($n=22$), SA ($n=20$), and intensive domestic immersion (IM) ($n=14$), for 9-13 weeks. He administered three tests designed to capture the breadth and depth of learners' vocabulary knowledge (i.e., how many words they know and how much they know about them) before and after

Chapter 2

each treatment and found that SA participants received significantly higher scores on all three vocabulary tests than their AH counterparts, with the IM participants showing a fairly similar performance to the SA group (except for one test where they performed significantly worse).

Finally, in a more recent study, Fitzpatrick (2012) set out to track developmental changes in vocabulary knowledge of a Chinese undergraduate studying English in the UK. Using a word association task, Lex30, at six time-points over an eight-month period, Fitzpatrick was able to elicit information on different aspects of the participant's lexical knowledge (i.e., associations and collocations). In her findings, the learner's vocabulary developed in a non-linear way: while data revealed a gradual growth in some aspects of vocabulary knowledge (collocations, NS-like associations), there were uneven striking inconsistencies in others (connections, form-meaning). Although the author explicitly cautioned against extrapolating from a single case study to make generalisations about vocabulary learning, she advocated for future research to focus on the "micro-development" of the lexicon, where the acquisition process may not necessarily be linear, but rather "chaotic and elusive" (p. 92).

As we have seen then, some of the attempts to measure the amount of words that students learned receptively (either at home or abroad), were complemented in later studies by other indications on how well learners knew the words (e.g., by including word association tests in vocabulary assessment).

2.1.4.2. Productive vocabulary

Due to the multiple foci of most SA research, as Dewey (2008) points out, only limited vocabulary-related data have been usually collected. In fact, a number of studies in the SA literature have looked into L2 productive vocabulary knowledge in a more general sense, typically relying on a few lexical complexity indices in terms of the CAF framework based on L2 output in response to writing or speaking tasks.

Written production

One of the few studies that examined several aspects of lexical richness in writing is that of Barquin (2012). Following the CAF approach, Barquin considered the development of fluency and lexical complexity of 30 Spanish/Catalan learners of English before and after a 6-month FI period and a 3-month SA. She found that learners improved considerably in fluency (as measured by the number of words and sentences that learners produced in their essays), producing significantly longer essays after SA than after FI. Results also showed that learners improved significantly in the Guiraud's Index (GI) of lexical diversity, showing a greater variation in their word choice as a result of SA. No changes, however, were found in any measure of lexical sophistication evaluating learners' use of rare words (neither through Advanced Guiraud 1000, nor via noun and verb hyponymy). Barquin's findings were in line with the results of a companion study by Pérez-Vidal and colleagues (2012) on oral and written development,

Chapter 2

drawing from the same population (29 Spanish/Catalan learners of English following sequential periods of FI and SA). Pérez-Vidal et al. reported significant improvement in written lexical diversity, as measured by GI, also as a result of the SA period.

Together Barquin and Perez-Vidal's results are comparable to those reported earlier in Freed et al. (2003). The authors investigated the acquisition of written and spoken fluency in American L2 learners of French: 15 of whom went on SA to France for a semester, and 15 of whom remained on campus AH. Although Freed's team was not especially concerned with vocabulary acquisition, their study targeted two dimensions that somehow reflected participants' lexical proficiency: the length of compositions and lexical density (the proportion of content words). Freed and colleagues found that the post-test compositions written by the SA group were much longer (though not statistically significant) and slightly denser in lexical use than their pre-test compositions, whereas the AH group did not show this change. More recently, Serrano and others (2012) examined changes in oral and written production in 14 Spanish-speaking learners of English spending an academic year in the UK from a CAF perspective. The authors reported a significant improvement in the GI of lexical diversity, yet this improvement occurred only after the second semester abroad (8 months approximately).

Finally, the main findings of Laufer and Paribakht (1998), on the contrary, did not seem to indicate an advantage for L2 learning abroad over classroom instruction AH in terms of a more sophisticated vocabulary. The authors, investigating the effects of language learning

contexts on receptive and productive vocabularies in adult learners of English in Israel ($n=79$) and in Canada ($n=103$), found that advanced learners following FI in Israel had significantly higher lexical sophistication than their counterparts studying in Canada, as measured through Lexical Frequency Profile (LFP) in their free written expression⁹.

Oral production

Early research on SA and word knowledge is to be found at the interface between lexical and grammatical competence. Ryan and Lafford (1992), and Lafford and Ryan (1995) studied the developmental stages of the Spanish verbs *ser* and *estar* and the prepositions *por* and *para* in learners acquiring Spanish in Granada and compared them to those in learners following classroom instruction. The authors found that the input received in the SA environment accounted to a great extent for the variation in the order of acquisition of these features across groups. Also, in his study of learners' copula choice, as determined through performance on grammar tests and communicative tasks, DeKeyser (1990) showed that those learners studying abroad were slightly more accurate than their peers receiving classroom instruction at home.

Although not focused solely on the development of L2 vocabulary knowledge, some years later Collentine's (2004) study illuminated

⁹ Laufer and Paribakht (1998) also found that intermediate and advanced-level learners studying in Canada outperformed the Israeli learners in controlled active vocabulary size, as measured through the Vocabulary Levels Test (VLT), a form-recognition test designed to measure passive vocabulary knowledge.

interesting results concerning the acquisition of lexico-grammatical competence. In his comparative analysis of oral discourse, captured through OPI (conducted among 46 participants prior to and following a semester-long treatment in either FI at the home university or SA exchange), Collentine looked at *semantic density*¹⁰, following Biber's (1988) definition to code features associated with informational richness. He found that the FI group outperformed the SA group in the ability to generate unique word types (adjectives and nouns), but SA participants' speech was more semantically dense (although by virtue of speaking more fluently and thus producing more words in the given time frame). Collentine's findings parallel those of Freed and colleagues (2003), following an analogous population and using the same OPI to elicit L2 speech. According to Freed's results, by the end of the semester, SA students spoke significantly more and faster than their AH peers.

At the level of lexical diversity in free oral productive vocabulary, results are somewhat inconclusive. A recent study by Lara (2014) examining the impact of SA programmes varying in length (3-month stay vs. 6-month stay) on CAF oral development by bilingual advanced learners of English did not detect significant changes in lexical variety of student speech productions (as elicited by means of a role-play task and measured through GI). However, after SA, GI scores of the 6-month stay group approached native-like patterns to the point at which they were no longer statistically distinguishable

¹⁰ The features that Biber (1988) identified to be indicative of semantically dense (i.e., informationally rich) discourse included a great proportion of nouns, adjectives, and prepositions, as well as multisyllabic words (as measured by number of letter or phones), and a high TTR (Collentine, 2004).

from NS scores in GI. Lara's results confirmed previous findings reported by Pérez-Vidal and Juan-Garau (2011) for the same role-play task and analogous participants undertaking a 3-month SA period, where the authors did not find any improvement for lexical diversity (via GI). These results were also in tune with yet another study by Pérez-Vidal et al. (2012) analysing the differential effects of FI and SA on oral and written EFL development. In their analysis of oral performance involving 29 Spanish/Catalan learners of English through sequential FI and SA learning experiences, gains in lexical diversity, as measured by GI, did not reach significance in any of the contexts under scrutiny.

While together these results point to a lack of significant progress in lexical diversity for free speech, there are two studies that have reported the opposite. A study done by Serrano, Tragant and Llanes (2012), for example, investigated changes in writing and speech in an analogous pool of Spanish learners of English studying in the UK over a course of a year. They found that after the first semester (3 months) alone students were already making significant progress in terms of oral lexical diversity, as measured through GI. Similar findings were reported in Foster (2009), although differing in methodology and participants' demographic. By combining quantitative and qualitative approaches to gauge the development of productive vocabulary knowledge, Foster used a more innovative measure of lexical diversity, D. She contrasted two comparable groups of intermediate EFL learners in two learning contexts: the SA group ($n=40$) from different L1 backgrounds having spent at least one year in the UK, and the AH group ($n=60$) involving L1 speakers of Farsi in Tehran. She also

included baseline data from 40 London-based NSs of English and used cartoon picture prompts to elicit learners' and NSs' speech productions. Foster found a significant effect for group and showed that not only were the SA participants' productions significantly more diverse than those of the AH group, their vocabulary was also no less diverse than that of the NSs. The qualitative analysis also showed evidence that the SA learners were to some degree framing their sentences in more native-like ways (we further discuss these issues in Subsection 1.2.1.3., as it deals with the analysis of formulaic language).

One final study that sheds some light on the domain of oral lexical accuracy in an SA context is that of Llanes and Muñoz (2009). Although originally aimed at evaluating the effects of short stays abroad (3-4 weeks) on the development of EFL listening comprehension, fluency and accuracy in a general sense, the study looked exhaustively at the number of lexical errors. Twenty-four Catalan/Spanish students provided the L2 English data pre- and post-SA through oral narratives. Results showed that lexical errors decreased significantly after the SA experience with low-level participants "showing comparatively greater gains in using L2 words (and hence in acquiring vocabulary) and in producing more accurate and fluent speech" (p. 361).

2.1.4.3. Formulaic language

Investigation of lexical knowledge has also been approached in SA-related studies from the perspective of formulaic language use and idiomaticity, albeit initially in an indirect way. Thus, a set of early

studies investigating fluency or morphosyntax in SA environment attributed gains to an expanded use of formulaic sequences rather than anything else. For instance, a series of studies conducted by Möhle and Raupach in the early 1980s (as cited in Foster, 2009) showed that SA learners, compared to their AH peers, improved in the area of fluency thanks to their ability to sound more natural in the L2, while Towell et al., (1996), and Regan (1998) reported significant gains in fluency due to the learners making greater use of formulaic sequences. Marriott (1995), and Siegal (1995) also found a greater number of formulaic expressions in SA learners, suggesting that these, rather than syntactic knowledge, accounted for growth in learners' overall morphosyntactic complexity. Similarly, Regan (1995) did not find improvement in morphosyntactic development for SA learners of French, though she did find that learners had acquired an ability to delete the negative particle *ne*, which made them sound more informal and native-like. Broadly, as Foster (2009: 93) points out, "these studies are pointing to lexical organisation, especially of the formulaic language kind, as the main area of benefit for SA learners."

In the same study, centred strictly on vocabulary acquisition, Foster (2009) set out to explore productive vocabulary quantitatively (assessing learners' lexical diversity scores through D, as we explained earlier) and qualitatively, describing learners' lexical choices and comparing them to native speech, a linguistic capacity called *native-like selections*.¹¹ Foster was able to show that SA learners were framing their

¹¹ The term *native-like selection* was first coined by Pawley and Syder in 1983 and refers to "the ability of the native speaker routinely to convey his meaning by an expression that is not only grammatical but also nativelike; [...] how he selects a sentence that is natural and idiomatic from among the range of grammatically

speech in more native-like ways than their AH counterparts. They used more narrowly defined lexical choices instead of broad general vocabulary, more colloquial language, and target-like collocate phrasing. Five years later, Foster (2015) retook a similar analysis, although this time tapping L2 receptive knowledge of native-like selection, and found that an early age of exposure to the TL in an immersion situation, in particular, served as a guarantee of nativelikeness.

Foster's findings were in tune with those reported by Siyanova and Schmitt (2008) examining native and non-native collocational knowledge. According to the results, "extended stays in an L2-speaking environment lead to a more native-like idiomaticity" (p. 447), as participants who had spent a year in an English-speaking country showed significantly better intuitions of collocation than those who had never been abroad. The authors concluded that whether it is shorter or longer than a year, a prolonged stay in the L2 country (with presumably a great deal of exposure to natural L2) can help learners become more native-like in their perception of collocations than learners without any L2 natural exposure.

2.1.5. SA, initial level, and vocabulary

As we have discussed in Section 2.1.2, one consistent observation in SA-related literature is that individual differences may play a vital role in the relative success of the L2 learning experience. Learners differ in

correct paraphrases, many of which are non-nativelike or highly marked usages" (p. 191).

their cognitive abilities, personality traits and level of motivation and such individual variation may determine the quality and quantity of learners' contact with locals and thus strongly influence the final result of SA on L2 progress.

According to Collentine (2009), there is a wide consensus among researchers that learners must have reached a particular *threshold level* of proficiency to fully benefit from the SA learning context. Although it is unclear what exactly that level is, once all learners have 'crossed' the threshold, the higher level ones will presumably gain less, in line with the normal learning curve.

Brecht et al.'s (1990), also reproduced in Freed in 1995, was one of the early studies to provide evidence that pre-programme proficiency level influences SA outcomes and should be adequately controlled for in research design. Finding gains on both reading and listening proficiency to be negatively related to pre-programme level, that is, "the higher the initial level, the less the gain (p. 46)," Brecht attributed this phenomenon to the nature of the learning curve, being initial level the most robust predictor of differential SA outcomes in the study. As of that time, much of the available body of research on the role of initial level in SA context has also documented greater gains for learners with lower initial proficiency than for the more advanced learners (e.g., Juan-Garau, 2014; Lapkin et al., 1995; Llanes and Muñoz, 2009; Marriott, 1995; Milton and Meara, 1995; Mora, 2014; Regan, 1995; Towell, 2002).

At the level of vocabulary, however, results underlying the level of proficiency prior to SA are less clear-cut. The first pioneering study by Milton and Meara (1995) reported a strong negative correlation between students' entry scores and vocabulary growth rates, showing that learners with poor vocabulary at the outset of exchange were the ones who benefited the most from the SA experience. The trend for learners with lower proficiency to make greater gains after SA was later confirmed in another study by Llanes and Muñoz (2009). The authors named participants' proficiency level as the key variable in their lexical accuracy progress.

However, the opposite conclusion, where major gains in vocabulary knowledge were not associated with lower initial proficiency, was drawn by Ife et al. (2000). In their analysis of the impact of SA on the vocabulary development of different proficiency groups, Ife found that both the more proficient as well as the less proficient learners improved in much the same way. The authors also encouraged further research in this area, relating the persistent lack of linguistic progress among more proficient learners with the inability of testing instruments to capture this progress, and thus giving the impression that it is always the less advanced learners who make the greatest gains (p. 4).

2.2. Vocabulary acquisition

While Section 2.1. has presented the main concepts and issues surrounding SA research, Section 2.2. aims to address the complexity of the L2 vocabulary acquisition process, with a special focus on

productive vocabulary and its assessment in free writing and speech. We dedicate the first subsection to the theoretical underpinnings of vocabulary knowledge in language acquisition by discussing its different components and presenting some of the approaches from which L2 vocabulary acquisition has been studied. We then go on to describe the several dimensions of lexical proficiency which can be used to evaluate vocabulary in free oral and written production. We discuss the different existing quantitative measures of vocabulary within the domain of fluency, density, diversity, sophistication and accuracy, as well as measures of qualitative analysis, such as native-like selection, which are relevant for the empirical study contained in this dissertation.

2.2.1. Vocabulary in SLA research

The importance of vocabulary as an area for investigation is by now well established in the field of SLA research. After years of relatively little attention, we are long past the time when vocabulary acquisition was regarded as ‘a neglected aspect of language learning’ (Meara, 1980). However, despite a widespread consensus among learners, teachers and the lay public that learning vocabulary is a crucial component for L2 mastery, an idea that has often been illustrated with learners carrying around dictionaries instead of grammar books, the role of vocabulary has been underestimated up until the last few decades.

In his attempt to summarise the historical trends on L2 vocabulary teaching, Zimmerman (1997) notes that SLA researchers, under the

influence of the Chomskyan school of linguistics, have particularly prioritised phonology and syntax as “more serious candidates for theorizing” (Richards, 1976: 77). By the same token, this general oversight of vocabulary studies in favour of studies on syntactic development was acknowledged by Levenston in his well-known critique in 1979. Vocabulary was neither the focus of attention in communicative language research, yet began to assume a greater role with regards to the primacy of grammar in L2 learning. David Wilkins (1974: 20), one of the chief collaborators of the communicative syllabi adopted by the Council of Europe, summarised his view of the relevance of vocabulary in language pedagogy and research, suggesting that “knowledge of a language demands mastery of its vocabulary as much as of its grammar,” which he considered possible through enough exposure to the language.

Despite these early claims in support of vocabulary, it was not until the early 1990s when researchers began to draw attention to the need for more systematic work on vocabulary acquisition, convincingly reasserting its status in the applied linguistics arena. Key to this shift are works by Sinclair (1991), emphasising the use of large lexical phrases or chunks in communication, or Nattinger and DeCarrico (1992), and Lewis (1993), challenging the traditional grammar-vocabulary dichotomy through the *lexical approach* in language learning. In Lewis’s terms, language consisted of “grammaticalised lexis, not lexicalised grammar,” a view that involved designing syllabi based on vocabulary rather than grammar in the EFL classroom. In more recent years, a number of good collections have appeared looking at L2 vocabulary from various perspectives: some of the most influential

ones being on vocabulary learning and assessment (e.g., Milton, 2009; Nation, 2001; Read, 2000; Richards et al., 2009; Schmitt, 2010), lexical processing (e.g., Fitzpatrick & Barfield, 2009), mental lexicon (e.g., Singleton, 1999) or word associations (e.g., Meara, 2009). In a parallel way, a series of articles have been produced dealing with varying aspects of vocabulary acquisition and providing empirical evidence of how important lexical issues are in L2 learning.

From this ongoing body of research, it becomes clear that vocabulary knowledge makes a substantial contribution to numerous domains of L2 proficiency (Zareva, Schwanenflugel, & Nikolova, 2005). Thus, research shows that learners with big vocabularies typically outperform those with smaller ones in a variety of language skills (Meara, 1996c) and reports high correlations between vocabulary (mostly measures of vocabulary size) and receptive and productive language abilities. For instance, a strong relationship has been established between vocabulary size and reading comprehension (Laufer, 1992; Laufer & Goldstein, 2004; Nation, 2001), or vocabulary and listening, being the lack of lexical knowledge a major obstacle for successful comprehension even for advanced language learners (Kelly, 1991). In fact, Kelly's study showed that more than 60% of errors that impeded understanding were lexical in nature. As regards to productive skills, lexical knowledge may be a potential indicator of learners' writing quality. Alderson (2005), in his analysis of the relationship between vocabulary and other linguistic skills as part of the DIALANG¹² project, reported a particularly strong association between vocabulary knowledge and writing, noting that "language ability is to quite a large extent a function of vocabulary size" (p. 88).

Chapter 2

Similarly, vocabulary knowledge has been shown to play a vital role in L2 communicative competence and to predict a considerable proportion of variance in speaking ability (Koizumi, 2005; 2013).

From the perspective of L1 acquisition, the idea that without vocabulary no language learning can take place is emphasised in Clark's book (1993), where the author tracks L1 lexical development from the very early stages of language acquisition. In her foreword to the book, Clark argues for the centrality of the lexicon in language, which, in her words, offers "a unique window on the process of acquisition for language as a whole" (p. 1). The fact that even native speakers develop their vocabulary throughout their whole life (unlikely any other linguistic domain) makes research in this field particularly challenging.

Thus far, we have elaborated on the trajectory of research on vocabulary, highlighting the revival of interest in the field, particularly since the mid-1990s. In subsections that follow, we will synthesise what is known about vocabulary knowledge as a construct and will expound upon the different interrelated components that comprise it. Our aim is to make sense of its complex and multifaceted nature, and to gain a better understanding of the theoretical implications involved.

¹² DIALANG was developed as a language assessment tool, examining different linguistic abilities – reading, writing, listening, grammar and vocabulary – in 14 European languages (Alderson, 2005).

2.2.2. Vocabulary knowledge

A common observation among vocabulary researchers is that there is still no generally accepted theory of vocabulary acquisition and is to date a quest for the ‘Holy Grail of vocabulary studies’ (Schmitt, 2010). However, we have several indications of what vocabulary knowledge consists of and explanations on how it is believed to be learned. As the present study examines learners’ productive vocabulary when performing two tasks after experiencing different learning contexts, we will also briefly comment on why dimensions of vocabulary knowledge play such a prominent role in models of language production.

One early widely-referenced contribution to vocabulary research was the work of Richards (1976) and Nation (1990), who considered that lexical knowledge implied knowing a series of characteristics associated with the word beyond the mere notion of its semantic meaning. In 1976, Jack Richards published an article in *TESOL Quarterly*, where he outlined several assumptions about vocabulary knowledge. In his assumptions, Richards considered that knowing a word involved knowledge of its spelling, pronunciation, syntactic behaviour, collocations, associations, frequency of occurrence, and register. Inspired by Richards’s conceptualisation, Nation (1990) refined his approach by incorporating additional components of vocabulary knowledge and emphasised the distinction between receptive and productive knowledge. A decade later, Nation (2001) presented an expanded version of his word-knowledge components approach, which continues to be “the best specification of the range

of word knowledge aspects” even until the present day, Schmitt (2010: 16) and a point of reference for research on aspects of lexical knowledge (see Milton & Fitzpatrick, 2014, for an updated overview). According to Nation (2001), the components of both receptive and productive knowledge entail three dimensions: (1) *form*, involving knowledge of spoken and written aspects of the word, as well as word parts, (2) *meaning*, consisting of knowledge of form-meaning links, semantic relations and associations, and (3) *use*, including knowledge of the grammatical functions of the word, its collocations, and register and frequency constraints, as presented in Table 1.

Table 1. Components of word knowledge from Nation (2001), p. 27.

		Receptive knowledge	Productive knowledge
Form	spoken	What does the word sound like?	How is the word pronounced?
	written	What does the word look like?	How is the word written/spelled?
	word parts	What parts are recognizable in this word?	What word parts are needed to express this meaning?
Meaning	form/meaning	What meaning does this word form signal?	What word form can be used to express this meaning?
	concepts and referents	What is included in the concept?	What items can the concept refer to?
	associations	What other words does this make us think of?	What other words could we use instead of this one?
Use	grammatical functions	In what patterns does the word occur?	In what patterns must we use this word?
	collocations	What words or types of words occur with this one?	What words or types of words must we use with this one?
	constraints on use (register, frequency)	Where, when, and how often would we expect to meet this word?	Where, when, and how often can we use this word?

In Table 1, as we have already commented upon, Nation made explicit one of the classical dichotomies in the vocabulary research literature; that is, *receptive* and *productive* vocabulary knowledge (sometimes referred to as *passive* and *active*).¹³ In most cases, receptive knowledge

¹³ While some authors use the terms *active/productive vocabulary* or *production*, and *receptive/passive vocabulary* or *reception* interchangeably (Melka Teichroew, 1997), others use them for different concepts (e.g., Laufer & Goldstein, 2004).

implies recognising the word when it is heard or seen, whereas productive knowledge extends receptive knowledge by adding the ability to speak or write the word in the appropriate context (Nation, 1990). With regard to the latter, some authors also distinguish two types of productive knowledge: *controlled*, which involves “producing words when prompted by a task” (usually in cloze and translation tests), and *free*, dealing with the “use of words at one’s free will, without any specific prompts for particular words,” as is the case of free writing or speech (Laufer, 1998: 257).

A commonly held idea in vocabulary studies is that receptive knowledge usually precedes productive knowledge (i.e., some receptive knowledge of words is acquired before becoming part of the productive repertoire), although the relationship between the two is not as straightforward as it might seem at first glance. Some authors suggest that the receptive-productive relationship is better understood in terms of a continuum, where words move along a path from receptive to productive status as more is learned about them (Melka Teichroew, 1997). Others, in contrast, argue that the distinction between knowing a word receptively or productively is rather clear-cut, since active vocabulary can be freely accessed and, at the same time, give access to other lexical items, whereas passive vocabulary cannot be accessed without the help of a stimulus (e.g., encountering a word when reading, see Meara, 1990). Part of the problem in describing the receptive-productive vocabulary knowledge lies in the difficulty of measurement, as the results largely depend on the test instruments used (Laufer & Goldstein, 2004).

Another widespread distinction in vocabulary-related studies, put forth by Anderson and Freebody in 1981, is that of *size* (or *breadth*) and *depth* of vocabulary knowledge. Following their definition, vocabulary size refers to the quantity of words that a person knows (how many words are known?), while vocabulary depth specifies the quality of such knowledge (how well are these words known?). Just as with the receptive-productive distinction, the relationship between size and depth of vocabulary knowledge is unclear and highly dependent on how each dimension is conceptualised and measured (see Schmitt, 2014, for a review).

In the years following Richards's publication, researchers have been actively engaged in defining vocabulary knowledge in terms of more comprehensive frameworks that integrate the different dimensions of the construct. Although simple binary divisions like receptive-productive or breadth-depth do not really do justice to the complexity of vocabulary knowledge (Milton, 2009), they do help to achieve greater specification of the construct and to assess more accurately the different word knowledge aspects involved. Some of these frameworks have been based on at least two dimensions of vocabulary, such as breadth and depth, as in Read (1988), Wesche and Paribakht (1996), and Qian (1999), others have incorporated the receptive-productive component among others (Henriksen, 1999). Meara (1996a) describes lexical competence in terms of three dimensions: namely, size, organisation and access (an approach also adopted by other researchers such as Daller et al., 2007). He argues that while depth is a property of individual words, the three dimensions proposed are properties of the whole lexicon, they are all

measurable and therefore could be more useful to describe learners' lexicons as well as vocabulary growth.

These dimensions, especially size and access, are essential in Levelt's (1989) theoretical model of speech production – the well-known *blueprint for the speaker* – in which vocabulary knowledge rather than grammatical knowledge is also the engine driver. In his model, the number of lemmas in the lexicon, the information stored about them and how quickly they can be accessed is paramount in formulating a message, and once a lexical item is retrieved and selected for use, its collocations are also systematically activated.

In an attempt to account for the acquisition of L2 vocabulary knowledge and the mental processes involved in vocabulary learning, Vanniarajan (1997) proposes an interactive model of vocabulary acquisition, where the acquisition of phonology, morphosyntax, pragmatics, semantics and world knowledge converge in the acquisition of lexis. In Vanniarajan's terms, the depth of word knowledge very much depends on learners' previous encounters and experiences with the word and the different affective values built upon these experiences. In sum, this model parallels the idea that vocabulary acquisition is incremental in nature and constitutes a “gradual process of one meeting with a word adding to or strengthening the small amount of knowledge gained from previous meetings” (Nation, 2001: 155).

Although L2 vocabulary acquisition is slow and laborious, large vocabularies can be acquired given enough input in the TL (Meara,

1988), with exposure to a massive amount of natural texts being the key to vocabulary acquisition. Following this line of reasoning, Kirsner (1994) highlights the importance of unconscious, implicit learning and the role of practice in lexical acquisition in that the more learners experience lexical units or phrases, the better these units are stored and retrieved from memory. What emerges from these arguments is that input, output, practice, and interaction accelerate language learning and this is not an exception for L2 vocabulary acquisition.

As discussed in a recent volume by Dóczy and Kormos (2016), L2 writing and speaking tasks are among the best tools to help learners identify gaps in their knowledge, gain feedback and as a result expand their existing lexical repertoires. In their meta-analysis of twelve studies, for instance, Huang et al. (2012) showed that composition writing was one of the most beneficial output tasks for vocabulary acquisition, lending weight to the *Involvement Load Hypothesis*, e.g., the greater the involvement with a word, the more efficient the learning. In addition to the output tasks and active engagement with the L2 vocabulary, interaction, and opportunities for negotiating meaning may similarly promote vocabulary acquisition by reinforcing form-meaning links and automatizing lexical retrieval processes. As shown in a recent study by Newton (2013), negotiating the meaning of unfamiliar words can greatly contribute to vocabulary development and improve students' knowledge of the target words. All in all, although all the aforementioned factors may speed up vocabulary learning to some extent, if lacking appropriate feedback, explicit vocabulary teaching focus, and motivation, successful attainment of L2 vocabulary may still not be guaranteed (Dóczy & Kormos, 2016).

In this section, we have offered insights on the multidimensional nature of vocabulary knowledge and the factors influencing L2 vocabulary acquisition. We have seen that despite the difficulty to identify what best contributes to its expansion, vocabulary is undoubtedly an essential component of communicative language ability. Whether one framework is more comprehensive than the other, or which model of representation is more adequate as a psycholinguistic approach to vocabulary processing¹⁴ is still a matter for debate and goes beyond the scope of the present dissertation. One thing that seems obvious in this discussion is that vocabulary knowledge is multifaceted and “any single measure of it will give only a very minimal impression of the overall lexical knowledge constellation” (Schmitt, 2010). For this reason, vocabulary assessment should employ multiple measures for the different word knowledge aspects (Laufer, Elder, Hill, & Congdon, 2004; Read, 2000; Schmitt, 2010; Webb, 2005), as will be discussed further in the next subsection, especially dedicated to the overview of the measurement of free productive vocabulary.

2.2.3. Measuring vocabulary

Measuring vocabulary is not less troublesome than attempting to define what is involved in knowing a word or what vocabulary knowledge is. The results of such examinations are by and large conditioned by the kind of vocabulary being measured and the measurement instrument used. In the following subsections, we will

¹⁴ E.g., Revised Hierarchical Model (Kroll & Stewart, 1994), Distributed Representation Model (Van Hell, 1998), etc.

briefly review how free productive vocabulary has been assessed in the literature, which is particularly relevant for the aims of our study, and will describe various measures of productive vocabulary in the domain of fluency, density, diversity, sophistication and accuracy. Before delving into measurement of each domain, we will first define the units of measurement (how lexical items are generally counted) and then will discuss each domain separately, expounding upon the interactions in written and oral production.

2.2.3.1. Units of measurement

Every manual on vocabulary research or assessment (Milton, 2009; Read, 2000; Schmitt, 2010, to name but a few) is very likely to contain a section devoted exclusively to the overview of the term *word* and the different units of ‘counting vocabulary’. What is meant by a word or how we count words is not a trivial matter, as it is generally recognised that the different ways of counting lexical items may lead to vastly different results. Accordingly, a persistent issue in lexical studies is that “figures are reported, but without a clear indication of how they were derived” (Schmitt, 2010: 188). Therefore, some basic points have to be spelled out from the start.

The term *word* is used very loosely in SLA research, so much so that it may not be understood in exactly the same way across studies, requiring more specific definitions. Thus, Carter (2012), for instance, considers the term *word* inaccurate and proposes a more neutral label, which, to some extent, helps to overcome ambiguity and confusion. In his view, the variable phonological, orthographic, grammatical and

semantic properties of words are best captured by the use of the term *lexical item*. Similarly, Sinclair (2004: 281) embraces the idea that the lexical item is better described maximally, not minimally, comprising “one or more words that together make up a unit of meaning.” On the other hand, when providing an estimate of productive vocabulary size or calculating the number of words written in a composition, other units of measurement have come into use, the most cited of which are *tokens*, *types*, *lemmas*, and *word families*. *Tokens* refer to the total number of running words in the text, regardless of how often the same word forms are repeated. This unit of counting is mostly used for quantifying the length of texts. *Types*, in turn, are the total number of different words. Here the word forms that are repeated are counted only once. For example, the sentence *The dog chases the cat* consists of 5 tokens but 4 types, as the article *the* is repeated. The relation between the number of types and tokens constitutes the basis of lexical diversity measures, as we will discuss in a later section. Another way to count vocabulary is through *lemmas*. A lemma is defined as the root¹⁵ form of a word (e.g., *speak*) plus its inflected forms (e.g., *speaks*, *spoke* and *speaking*). Thus, the four forms of the word *speak* are counted as one lemma. Using lemmas is more appropriate in psycholinguistic research (Schmitt, 2010) or in studies involving highly inflected languages (e.g., French in Treffers-Daller, 2013). One last widely-used unit of measurement is that of a *word family*, which consists of a base form of a word, its inflected forms and semantically related derivatives. Following this definition, the items *speak*, *speaks*, *spoke*, *speaking* and *speech* would make up a word family. Although it is not easy to determine what items should (and should not) be included in a

¹⁵ The root (or base) form is defined as the simplest form of a word.

word family (Nation, 2001), this unit of measurement is the best at capturing all of the word forms linked to a concept.

Deciding the best unit of counting vocabulary depends mostly on the technical resources available, the research questions or conceptual considerations of lexical storage (Schmitt, 2010). Recent trends in vocabulary research point to the inclusion of even higher units consisting of more than one word, variously known as *formulaic sequences* or *chunks*¹⁶. The phenomenon of formulaic language is becoming increasingly popular among vocabulary researchers (Nattinger & DeCarrico, 1992; Schmitt & Carter, 2004; Wray, 2002), yet its definition is rather broad and inclusive. Encompassing a wide range of subcategories (prefabricated phrases, idioms, phrasal verbs, collocations, set or fixed phrases and semi-fixed phrases, as in Wray 2002), characterised by having an idiomatically determined meaning, formulaic language is one of the major hurdles in learning English and a “hallmark of the highest stages of language mastery” (Schmitt, 2010: 145). As Schmitt (2010) points out, it takes a long time to master formulaic language in native-like ways, and its lack or misuse is a primary reason why L2 discourse tends to sound unnatural and odd.

To sum up, specifying the unit of measurement is of paramount importance, as the results of any research will vary depending on what is meant by ‘word’. Careful consideration as to the type of vocabulary being measured (receptive versus productive) or participants’ profile

¹⁶ Wray (2002) encountered over 50 terms (e.g., *chunks*, *collocations*, *formulaic speech*, *formulas*, *multi-word units*, *prefabricated routines*, etc.) to refer to the phenomenon of formulaic language.

(early L1 vocabulary acquisition in children versus L2 vocabulary learning in advanced bilingual learners) is key to adequate results interpretation. When it comes to the analysis of free written and oral production, some authors may adopt specific units of analysis (e.g., Treffers-Daller (2013) lemmatises her French data because of the highly inflectional forms, while Laufer (1991) uses the lexeme as her unit of analysis writing in Israeli students of English), although most studies usually count the number of types and tokens. We also follow this approach in the present study and use tokens and types as the basis of our word counts. Having considered the units of vocabulary measurement and reported our unit of counting, we can now shift our attention to the assessment of free productive vocabulary in writing and speech, which is of particular interest for our study.

2.2.3.2. Measuring vocabulary in free writing and speech

In his timely book on vocabulary assessment, Read (2000: 200-201) lists a set of features that can adequately capture effective vocabulary use in free production. Read identifies the following components: (1) lexical variation, (2) lexical sophistication, (3) lexical density, and (4) the number of errors (if any) in vocabulary use. Lexical variation (also called lexical diversity) refers to the variety of words used in production. A person who is not repetitive in his choice of words is most likely to possess a high degree of lexical variation. According to Read, more proficient learners with larger vocabularies will avoid repeating the same words by using synonyms, subordinates, and other related words, showing a wider range of expression than less advanced learners. Lexical sophistication is equally informative and pertains to

the proportion of relatively uncommon or advanced words to general, everyday vocabulary. A learner who produces *The dog chases the cat* has probably more limited lexical resources than a learner who writes *The greyhound pursues the feline*, even if the message is substantially the same. Therefore, a high percentage of advanced and rare words in a text or conversation (i.e., *greyhound/feline*), as opposed to easy words (i.e., *cat/dog*), is an indication of more sophisticated vocabulary use, intuitively ascribed to more competent writers or speakers. Lexical density is generally described as the proportion of content (or lexical) words to function (or grammatical) words and is one characteristic that discriminates between written and spoken language, as we will discuss in further detail in Section 2.3. Finally, features like spelling mistakes, wrong lexical choice, and errors resulting from L1 interference or derivational morphology can be measured to describe and compute lexical errors in language production, thus addressing the quality and correctness of use of L2 vocabulary knowledge.

Read's list echoes one of the early in-depth reviews of complexity, accuracy and fluency (CAF) initiated by Wolfe-Quintero and colleagues (1998) in their book on L2 writing assessment. In their volume, the authors consider that measures of fluency, accuracy, and complexity in SLA can all relate to the lexicon: lexical fluency measures the quantity of written or spoken output (how many words – tokens, types, verbs, etc. – do learners produce to get their message across?); lexical complexity is manifest in terms of lexical variation and sophistication of L2 learner's productive vocabulary (how diverse and sophisticated the produced vocabulary is?); and lexical accuracy involves the correct use of the target vocabulary (how accurate are

these productions?). To this list, Wolfe-Quintero and colleagues also add the domain of lexical density, and conclude that despite not being systematically investigated across studies, lexical complexity measures “do offer promise as indicators of language development” (p. 115). Ten years later, Bulté and his team (2008) quantify the development of vocabulary through a set of metrics, taking a similar view to Read and Wolfe-Quintero in that a learner’s *lexical proficiency* can be construed in terms of lexical fluency, diversity, sophistication, complexity, and productivity of his L2 use, an approach that resembles ours, as will be seen in the remainder of the chapter.

2.2.3.2.1. Lexical fluency

Although in common parlance fluency is used as a loose cover term, typically equated with global language proficiency, in the field of SLA, fluency has been recognised as a complex phenomenon that can be operationalised in a variety of ways (Freed, 1995b; Segalowitz, 2011). Wolfe-Quintero and colleagues (1998) regard fluency in terms of frequencies and ratios. Fluency frequencies are defined as “the sheer number of words” produced by an individual, and this amount or length of production can be measured by calculating the total number of words or other structural units (i.e., clauses, sentences, T-units, verbs, etc.) contained in a piece of writing or transcribed speech¹⁷. Frequency ratios, on the other hand, consider temporal aspects of fluency and measure the rate of production, which in writing is usually counted as the number of words per minute, and in speech as either

¹⁷ Although Wolfe-Quintero primarily focuses on the measures of CAF in writing, her classification can be extended to spoken language, too.

the number of words per minute or syllables per second (e.g., Arthur, 1979; Griffiths, 1991; Ortega, 1995, but as cited in Wolfe-Quintero et al., 1998).

In studies assessing lexical richness¹⁸, fluency, in the Wolfe-Quintero's sense of frequencies, can also go by other names such as lexical productivity (Bulté et al. 2008; Dewaele & Pavlenko, 2003), lexical output (Read, 2005; Read & Nation, 2006) or vocabulary count (Golonka, 2006). Dewaele and Pavlenko (2003), for instance, suggest that the degree of proficiency in the L2 may determine productivity up to a certain point and consider the total number of words (tokens) as the best measure of productivity, given that words are easier to define than clauses, sentences, or T-units, particularly when analysing spoken discourse. Similarly, Golonka (2006) shows that vocabulary count, as measured by types, is a powerful predictor of oral proficiency gains, proved to be one of the most discriminating factors for gainers and non-gainers.

The number of types was one of the many measures used to capture lexical productivity in the study of Bulté and colleagues (2008) investigating L2 lexical proficiency development in Dutch-speaking learners of French. Furthermore, the authors also evaluated the word-class distribution in learners' production, quantifying the proportion of three lexical classes (nouns, verbs, and adjectives). The rationale behind such inclusion was that increases in lexical productivity might

¹⁸ Although lexical richness has often been used as an equivalent to lexical diversity (Johansson, 2008; Daller et al., 2003), some authors (Malvern et al., 2004; Read, 2000) use this term to refer to other dimensions such as lexical sophistication or complexity.

be related to “lexical growth in one or two content word classes only,” although it remains unclear which lexical class is most likely to account for such developmental fluctuations (p. 10). Examples of studies that approach lexical productivity by counting word classes can be found in Broeder et al. (1993) and Collentine (2004). Broeder et al.’s study suggests that quantitative developments in particular word class categories may have a bearing on the overall development of the lexicon, just as Collentine (2004) who considers the frequencies of seven word classes (nouns, adjectives, verbs, adverbs, prepositions, pronouns and conjunctions), as a metric of the growth of the learners’ lexicon.

Whether measured in types, tokens, or word class frequencies, lexical fluency seems to be a good baseline indicator of L2 written lexical competence, as a number of studies demonstrate that the amount of text produced grows with proficiency (Bulté et al., 2008; Linnarud, 1986; Wolfe-Quintero et al., 1998), but less so for L2 oral development. Read (2005: 13) considers that lexical output by itself (i.e., types and tokens) has “a limited value as a measure of the quality of the learner’s speaking performance,” and must be complemented with additional objective measures of lexical richness, that are presented as follows.

2.2.3.2.2. Lexical density

Another fruitful area for analysis of free lexical production pertains to lexical density. Defined as a measure of the proportion of content words (nouns, verbs, adjectives, and adverbs) to the total number of

Chapter 2

words (word tokens), lexical density is a useful way to distinguish written from oral language (Read, 2000). Ure (1971), who originated this measure, found lexical density to be over 40% in writing and below 40% in speaking, showing that information is usually presented in a more concentrated way in written modality. She also found it to be a good indicator of the characteristics of different textual types and modes of delivery, noting that a prepared speech was far denser than spontaneous talk.

Although it is reasonable to believe that higher scores in lexical density would imply greater language development, it is not entirely clear what developmental prediction is to be expected (Wolfe-Quintero et al., 1998). Some authors consider that lexical density is an aid to learning about ‘information packaging’ (Johansson, 2008; Ure, 1971), since content words primarily convey information: thus, the higher the proportion of content words, the denser the information contained in the text. Others (Hyltestam, 1988; Laufer & Nation; 1995) note that lexical density may not give a clear indication of vocabulary development, with a high index of content words in a text not always reflecting more literate usage. Hyltestam (1988), for instance, proves that high values for lexical density in low-level learners stems from their inability to construct a coherent text, rather than producing a lexically dense discourse, while Laufer and Nation (1995) consider that lexical density may be dependent on the structural (syntactic and cohesive) properties of the text, that do not necessarily measure lexis. Johansson (2008) concludes that lexical density can be less informative than other aspects of vocabulary use (i.e., lexical diversity), particularly when detecting differences between age groups.

With this in mind, numerous variants of lexical density have been proposed, a popular one being that of ‘minor variants’ (Johansson, 2008); that is, calculating density of different parts of speech in the text such as noun density (the ratio of nouns to total number of tokens), or the density of verbs, adjectives, adverbs or pronouns and so on. Examples of studies that address density from the word class perspective can be found in Marsden and David’s (2008), and Collentine (2004). Taking the view that progress in language development can be seen through changes in the part of speech that predominates in production, Marsden and David’s study presents evidence that learners produce a higher concentration of verbs as opposed to nouns in speech as they progress, and that as they start producing more verbs, they also begin to produce more adjectives. Likewise, a higher density of nominal features (nouns and adjective) may be an indicative of semantically dense (i.e., informationally rich) discourse, synonymous with high levels of language mastery, in studies by Biber (1988; 2009) and Collentine (2004). Finally, there is also some evidence pointing to overuse of personal pronouns in L2 learners in contrast with NSs (Shaw & Liu, 1998; Silva, 1993) and further investigation into pronoun density may shed some light on the degree of formality especially in writing, since a text with a low pronoun density is usually associated with greater generality and impersonality and is more characteristic of academic writing (see Biber, 1988).

2.2.3.2.3. Lexical diversity

As mentioned above, lexical diversity means the same as lexical variation: a variety of vocabulary and avoidance of repetition. In other

words, lexical diversity scores give us the range or number of different word types deployed in a text by either a writer or a speaker and demonstrate how likely it is for him to repeat the same words. Like lexical fluency or density, lexical diversity is a typical example of what has been called *intrinsic* (Meara & Bell, 2001) or *text-internal* (Daller, van Hout, & Treffers-Daller, 2003) measurement of vocabulary, in that the assessment is performed in terms of the words contained in the text itself, without relying on any external sources (e.g., word frequency lists).

In SLA research, it is generally assumed that as learners improve in overall language proficiency, they tend to increase the variety of words they use in production. Empirical studies back up this assumption, finding lexical diversity variables to be valid as developmental indices and reporting high correlations between these and other aspects of linguistic proficiency (see Malvern, Richards, Chipere, & Durán, 2004, for a comprehensive review). Malvern and Richards (2004) point out that lexical diversity is indicative of productive vocabulary size, meaning its development implies underlying vocabulary growth. Although lexical diversity has been considered an illuminative indicator of the quality of L2 writing and speaking (Jarvis, 2002; Malvern & Richards, 2002; Yu, 2009), changes in lexical variation may not be uniformly visible across genres and modalities. Thus, Biber (1988; 2009) states that lexical diversity is stronger in academic writing than in conversation, as the processes of planning, revising and editing, which allow for use of a wide range of words, are only possible in the written register.

Traditionally, lexical diversity has been analysed by means of the type-token ratio (TTR), which relates the number of types to the total number of tokens. However, TTR tends to be strongly affected by text length, as the longer the text, the more likely that words will be repeated (e.g., Arnaud, 1984; Broeder et al., 1993; Daller et al., 2003; Malvern & Richards, 1997; Richards, 1987; Vermeer, 2000, 2004). That is to say, a longer text may by default produce a lower TTR value than a shorter text, calling into question the validity of the index and making the comparison of texts varying in length notoriously difficult. To overcome this, some authors (Arnaud, 1992) have suggested limiting each piece of writing or transcribed speech to the length of the smallest sample so that they are all of equal size¹⁹, while others have put forward various alternatives to TTR, attempting to correct the text-length factor. One of these is the Root TTR, most commonly known as the Guiraud's Index (GI). GI minimises the impact of text length by using the square root for relating types and tokens (types/ $\sqrt{\text{tokens}}$). Unlike a number of modified TTR calculations, such as the mean segmental TTR (MSTTR), the corrected TTR (CTTR), the Bilogarithmic TTR (LogTTR), Uber Index, to name but a few²⁰, that have also come into use in studies of L2 vocabulary to address the text-length dependency shortcomings, GI has been found to be the most stable one proving its validity in a handful of studies (Broeder et al., 1993; Vermeer, 2000) and highly reliable in correlations with other dimensions of linguistic complexity (David, 2009).

¹⁹ Arnaud's proposal gleaned considerable criticism as to its validity, since cutting the texts might result in a waste of valuable data.

²⁰ The studies that have addressed the text-length issue are too numerous to list and go beyond the scope of this research, but for a full review, see Malvern et al. (2004).

Chapter 2

In more recent years, Malvern and Richards (1997) have devised a new measure of lexical diversity, called D, which begins to be considered the ‘industry standard’ (McCarthy & Jarvis, 2013). D is based on mathematical transformations and uses a curve-fitting method to represent how the TTR changes in a text through random sampling. To calculate D, a set of mean segmental TTR values is computed for texts of different token size and then matched to a series of curves generated by the formula $TTR = D/N * ((1 + 2*N/D)^{1/2} - 1)$. The D value of the text is the value that produces the best-fitting curve. The primary advantage of D is that it discriminates across a wide range of language learners and users (Malvern & Richards, 2000; Richards & Malvern, 2000), proving to be more informative and theoretically more valid than previous measures. Furthermore, it operates very well with short texts (a minimum of 50 tokens is required to compute a valid D), which makes it potentially useful when dealing with learners who do not produce much.

Although the methodological advantages of D have been extensively confirmed – recent studies report significant positive correlations between D and EFL written and spoken tasks performances (Lu, 2012; Yu, 2009) – and its developers “have put us on the right track” (Jarvis, 2002: 82, but see McCarthy & Jarvis, 2007), different measures can be used to obtain more information. Despite these advances, most researchers, however, agree that the type-token approach, in all its guises, has limited capacity to fully evaluate lexical resources of L2 speakers and is certainly insufficient to assess its appropriateness of use (Jarvis, 2013; Meara & Bell, 2001; Schmitt, 2010). The aforementioned examples *The dog chases the cat* and *The greyhound pursues*

the feline illustrate such limitations, since both utterances consist of 5 tokens and 4 types each and will, therefore, produce the same values when measured through indices of lexical variation based on TTR, which is probably the reason why in some studies these measures have been seen insensitive to changes (Van Hout & Vermeer, 2007).

In this sense, given that a single index has a limited capacity to encompass the construct of lexical diversity (McCarthy & Jarvis, 2007), it seems reasonable to supplement it with other lexical measures to paint a more complete picture of learners' vocabulary knowledge. Vermeer (2000) recommends analysing the words in relation to their difficulty, instead of counting how many times they pop up in the data set. This approach leads us to consider another area for analysis of free writing and speech and is the domain of lexical sophistication.

2.2.3.2.4. Lexical sophistication

Lexical sophistication (also called *rareness*, Arnaud, 1984) reflects the extent to which an individual draws upon difficult or 'sophisticated' vocabulary in what he writes or says. It is usually assessed as the proportion of infrequent words in a text through external resources, which can evaluate one's choice of words based on its frequency in language. Because lexical sophistication measures are based on criteria external to the text itself, they are considered *extrinsic* (Meara & Bell, 2001) or *text-external* (Daller et al., 2003) measures. Wolfe-Quintero et al. (1998) find the measures of lexical sophistication intuitively appealing, being a potential indicator of lexical development. That is,

the better a learner is, the more likely he is to use more advanced vocabulary in production.

Lexical sophistication may be related to vocabulary size (Nation & Webb, 2011; Vermeer, 2000) and writing skills, as it is often associated with the academic register. Nation and Webb consider that “using low-frequency words allows learners to express meaning using more precise terms” (p. 251). Because of its academic orientation, lexical sophistication scores tend to be higher in the written modality than in oral speech and are representative of ‘literate lexicon’ (Nippold, 2006). Within this line of thought, Biber (2006) identifies six features that distinguish academic language from conversational English: 1) Latin and Greek vocabulary, 2) morphologically complex words, 3) nouns, adjectives and prepositions, 4) abstractness, 5) grammatical metaphor, and 6) informational density. For these reasons, along with low-frequency words, studies probing into lexical sophistication have variously assessed use of Greco-Latin words, abstract words, and word length, among others (Berman & Nir, 2010; Berman & Verhoeven, 2002).

One of the most noteworthy indices of lexical sophistication to date is the Lexical Frequency Profile (LFP) developed in 1995 by Laufer and Nation and available for use with the *VocabProfile*²¹ programme. The LFP looks into the proportion of words from different frequency levels. More specifically, it is calculated by segmenting (deconstructing) the lexicon of a text on the basis of four frequency lists: the first 1,000 most frequently used words in English (K1 Words), the second most

²¹ See Section 3.3.5.3.1., for details of how *VocabProfile* operates.

common 1,000 words (K2 Words), the words in the Academic Word List (AWL words), and all remaining vocabulary not found in any of these three lists (Off-list Words). The profile assigned to each text can be given on the basis of word families, types, tokens and percentages. Cobb (2003: 403) states that a typical LFP for NS English is: 70-10-10-10 (i.e., 70% from first 1000, 10% from second 1000, 10% academic, and 10% off-list words) in writing, and 80-5-5-10 in speech. A typical result for advanced EFL learner writing is 88-3-3-6.

In their original study, Laufer and Nation outlined a number of advantages of the LFP over traditional measures of lexical richness²²: (1) it correlated closely with other independent measures of vocabulary size (Nation's *Levels Test*), (2) it discriminated between learners at different levels of proficiency and lastly, (3) it was a reliable and objective diagnostic test of learners' vocabulary use. They concluded that by calculating the proportion of words belonging to different levels of frequency, LFP was able to give a snapshot of learners' lexical resources at that particular point of development and was a reliable index of their productive vocabulary distribution.

The merits of the LFP and the fact that it is easily accessible online have encouraged a steady flow of research, using this profiling method for different purposes: as a measurement of the vocabulary used in textbooks (Milton & Hales, 1997; Tragant, Marsol, Serrano, & Llanes, 2015) or classrooms (Meara, Lightbown, & Halter, 1997), or as a stand-alone assessment instrument of potential TESL trainees (Morris

²² According to the authors, the LFP represents an advance over Lexical Density, Lexical Sophistication, and Lexical Variation.

& Cobb, 2004). In addition to this, LFP has been extensively used to analyse EFL writing development (Laufer, 1994; Muncie, 2002, Lee & Muncie, 2006) or spoken discourse (Ovtcharov, Cobb, & Halter, 2006), sometimes involving native speaker comparisons (Cobb, 2003), or contrasting recognition vocabulary knowledge to productive (Laufer, 1998).

While Laufer and Nation's LFP remains influential, some methodological limitations for its use have already been acknowledged (Meara & Bell, 2001; Meara, 2005; but see Laufer, 2005, for rebuttal): one of them being that the technique is not sensitive enough to distinguish between learners with similar vocabulary sizes. However, despite having several alternative indices of lexical sophistication, such as Guiraud Advanced (Daller et al., 2003; Mollet et al., 2010) or measures using the frequency-band approach, like P_Lex (Meara & Bell, 2001), a number of studies have shown that LFP has strong concurrent validity (e.g., Daller & Xue, 2007; Horst & Collins, 2006; Lee & Muncie, 2006; Morris & Cobb, 2004).

Finally, while the frequency-band approach behind the LFP is worth pursuing, other external indicators can be used to supplement the analysis. In Horst and Collins's (2006) study, for instance, looking at the frequency-based vocabulary acquisition by francophone learners of English, learners made progress in language proficiency during the treatment, yet this progress was not reflected in the LFP analysis. A complementary analysis did show that improvement in productive vocabulary was taking place but occurred in the use of fewer French

cognates, a greater variety of frequent words, and more morphologically complex forms.

2.2.3.2.5. Lexical accuracy

The final domain of free productive vocabulary analysis considered here is the accuracy or appropriateness of use. Defined by Foster and Skehan (1996) as ‘freedom from error’, accuracy measures describe “how well the target language is produced in relation to the rule system of the target language” (Skehan, 1996, as cited in Ellis, 2012: 206). In SLA, accuracy has long been postulated as an indicator of advancedness and is “probably the oldest, most transparent and most consistent construct of the [CAF] triad” (Housen & Kuiken, 2009: 3).

At the vocabulary level, there are several studies that analyse accuracy in terms of the percentage of lexical errors contained in a text. Lexical errors are considered a valuable window into the process of L2 vocabulary acquisition, offering insights on the learners’ lexical competence at different stages of its development (Ellis, 1994). They are considered the most severe of all types of errors and highly damaging to communication, since they affect the meaning of the message and are typically the primary cause of intelligibility (Agustín Llach, 2011). Agustín Llach notes that lexical errors imply a lack of vocabulary knowledge or an inadequate use of words, contributing negatively to language assessment. In her terms, “the quality of written and oral production is influenced by the presence of lexical errors in the composition or discourse. In sum, a piece of language with many lexical errors will obtain a bad score” (p. 103-104).

Vocabulary errors may be of various kinds: sometimes, they may be due to wrong word choice when expressing an intended meaning, or have the right meaning but be in the wrong form; they may also not fit grammatically into the sentence in which they occur, or be simply stylistically inappropriate (Read, 2000). Many of the authors who deal with lexical errors set up their own taxonomies in order to analyse the data in a concise and systematic way. Some examples of such endeavours are Arnaud (1984) or Engber (1995), both tapping into written production, as elicited through compositions. Arnaud (1984) distinguishes spelling mistakes – minor (Ex. *teatcher*) and major (Ex. *scholl*) –, derivation mistakes (*to comparate*), false friends (also deceptive cognates), interference from another foreign language, or errors resulting from confusion between two lexemes (*The teachers learn them maths*) (p. 19). Engber (1995) develops a more elaborate classification scheme distinguishing between errors in the lexical choice (semantic error types in individual lexical items and in collocations) and those in the lexical form (derivational distortions, verb forms, spelling and phonetically related errors). This distinction is similar to that found in James (1998), who in his exhaustive volume *Errors in Language Learning and Use* devotes several pages to the discussion of lexical errors and identifies form- and content-oriented lexical errors.

Lexical accuracy has been seen to correlate highly with the quality of learners written and oral production (Engber, 1995) and reading comprehension (Agustín Llach, 2006), although research in this area is scarce. One inherent weakness of any accuracy measure is subjectivity and difficulty to identify and interpret the error. In fact, the issue of agreement or inter-rater reliability remains one of the most pressing

concerns in this type of analysis. Given that error detection and classification requires subjective judgments, it is highly desirable to have the errors classified and counted by multiple researchers so as to validate the results (Read, 2000).

2.2.3.3. Measuring native-like selection

As we have already mentioned, recent SLA trends have embraced a view of language as an inventory of memorised phrases, chunks or collocations, which are stored in memory as single wholes rather than constructed spontaneously (Wray, 2002; Pawley & Syder, 1983; Foster, 2009; 2014). Although the idea is not new (see, for instance, Peters (1983) on the units of language acquisition in the L1), technological advances in computing have facilitated the work on the topic.

Vocabulary research has witnessed an increasing interest in the study of formulaic language; yet, still little is known about ways in which knowledge of formulas shapes L2 vocabulary acquisition. Despite the objectivity of formulas assessment and lack of straightforward applications to identify the multi-word units, research has shown that there is a substantial amount of formulaic language in both writing and speech (Erman & Warren, 2000; Foster, 2001), and that its use is what distinguishes native from non-native production. Dörnyei, Durow, & Zahran (2004) consider successful acquisition of formulaic sequences to be the function of the interplay of three variables: motivation, language aptitude and sociocultural adaptation, the latter is clearly associated with residence abroad. Following this line of thought, Read (2000) notes that in oral and written production, one fundamental

indicator of L2 performance is the extent to which learners produce naturally-sounding lexical units rather than awkward collocations, achieving ‘native-like selection’ in Pawley and Syder’s words (1983). Foster (2009) argues that further research in this area is necessary, as zooming in on L2 idiomatic expressions (or lack thereof) may provide new insights into the nature of productive vocabulary knowledge and put flesh on other more conventional quantitative metrics. In this sense, the use of multiple measurements becomes essential in order to “triangulate results from more than one approach” and thus achieve more robust findings (Schmitt, 2010: 149). Armed with this greater understanding, we now turn to the final section of this chapter, which discusses the connection between writing and speech and considers in what ways modality impacts language acquisition.

2.3. The relationship between writing and speech

There is a long research tradition in studying language in the written compared with oral modality from diverse pedagogical, linguistic, psychological, and ethnographic perspectives (see review by Sperling, 1996; 2001). On the one hand, the writing-speaking relationship helps to better gauge the complexity of the two different modes of communication and isolate the many factors involved in text construction (Williams, 2012). On the other hand, certain strains of research, particularly from the perspective of L1 and L2 writing, view this connection as a means of identifying what spoken language experiences shape the development of written language and what needs to be learnt and unlearnt about language when shifting from one modality to another (Sperling, 2001).

At the textual level, early work in the field comparing spoken and written modes has often demonstrated that writing is syntactically more complex than speaking and requires more concision, elaboration, and grammatical accuracy. An extensive corpus linguistics analysis of textual characteristics of spoken and written English, undertaken by Douglas Biber in his 1988 volume *Variation Across Speech and Writing* identified a set of features that reliably distinguished between oral and written language across a range of genres, and registers. At the word class level, Biber found that the oral dimension was associated with a high proportion of verbs, adverbs, and 1st and 2nd person pronouns, whereas the literate dimension contained a greater frequency of noun and prepositional phrases. Another text-embedded difference between writing and speech, as we have noted earlier, pertains to lexical density. In an investigation of English language corpora, Ure (1971) found written discourse to be lexically denser than oral discourse (i.e., having a higher degree of information packaging). During the first years of the 2000s, a handful of developmental studies (Berman & Verhoeven, 2002: 183; Berman & Nir, 2010; Strömquist et al., 2002), analysing text-based lexical usage as diagnostic of writing-speech distinctions, demonstrated that written texts consistently score higher than the spoken versions of the same texts in terms of lexical diversity, lexical density, word length, register and abstractness, reflecting “a more elevated and carefully monitored style of expression.” Berman and Verhoeven (2002) explained the impact of modality on lexical usage as due to the distinct cognitive demands and communicative conditions associated with each mode of language production. Their finding gave further evidence that not only writers construct their texts with greater compactness and a higher

degree of grammatical and structural integration, they also draw on a more enlarged vocabulary as compared with when they speak (Hopper, 2001; Olson, 2006).

At the cognitive level, one important distinction between writing and speech has to do with the different social contexts in which each mode of language production is used, as the two modalities tend to be associated with different communicative circumstances and information-processing constraints. Thus, written language (more often than spoken language) generally takes place in the absence of a physically present interlocutor, with the writer and the reader being located at different points in time (Sperling, 2001; Strömquist et al, 2002). The written text is then produced under the assumption that only the finally revised version is to be perceived by the intended recipient. This relative isolation of writing, often described as ‘decontextualized’, contrasts with the spoken language, which takes place in a speaker’s immediate ‘here and now’ surroundings in the presence of an interlocutor to help him shape the message as it unfolds (Strömquist et al., 2002; Sperling, 2001). While spoken language typically allows feedback and is processed under strict online constraints, meaning that “it must be perceived online just as it is produced online” (Strömquist et al., 2002: 46), in writing, these constraints can be lifted. In other words, the inherent benefit for written production in terms of time, planning, monitoring, and revision (e.g., allowing for more control over attentional resources as well as greater opportunities to retrieve a recently acquired form and produce more accurate and complex language (Williams, 2012)), contrasts with the pressure associated with the spoken production. In

speech, the primary concern is to maintain the flow of conversation, a distinction which makes the artefact of modality particularly interesting in language studies.

2.4. Summary

In this first chapter we have provided a brief summary of the state-of-the-art in research on language learning during study abroad with a special focus on vocabulary. In Section 2.1., we have seen that on a backdrop of Europe's turbulent history, many endeavours on the part of European institutions were made to achieve harmonious coexistence and mutual understanding between the countries, being mobility and foreign language learning the part and parcel of their subsequent educational and linguistic policies. This is the socio-political context that has given rise to systematic SA research in Europe testing the efficiency of ERASMUS programmes that differ to some degree from the programmes in other parts of the world (i.e., American exchanges). We have also presented a brief history of SA research from the very early attempts to predict foreign language development based on multiple-skill studies to the collection of papers contained in Freed (1995a) and further contributions such as the 2004 special issue of *Studies in Second Language Acquisition (SSLA)* or volumes edited by DuFon and Churchill (2006), Kinginger (2013), or Pérez-Vidal (2014), among many others.

Also in this chapter, we have described SA as a learning context in relation to SLA and as a function of the interplay of three different parameters, at the macro- and micro-level, and with regard to the

programme-specific design (Pérez-Vidal, 2014b). We have then summarised the impact of SA on linguistic abilities as reported in SA related literature. More time has been spent on vocabulary acquisition, specifically on its receptive and productive facets as well as on knowledge of a formulaic kind. To conclude Section 2.1., we have considered the role that initial proficiency has on vocabulary acquisition in SA contexts, as research evidence indicates that learners with different proficiency level do not always benefit in the same way.

In Section 2.2., we shifted our focus away from SA toward the overview of L2 vocabulary acquisition. First, we have seen that after many years of neglect vocabulary acquisition finally made its way into the applied linguistics arena established itself as a promising area for investigation. Specifically, it was in the early 90s that research on vocabulary gained momentum and has ever since been addressed in a myriad of ways and from various perspectives. We have also seen that although there is still no single overarching theory of vocabulary acquisition, a number of significant assumptions and frameworks have emerged in order to describe and better understand the complex nature of vocabulary knowledge: the components of word knowledge by Richards (1976) and receptive-productive dichotomy (Nation, 1990; 2001), the breadth and depth distinction by Anderson and Freebody (1981), as well as the different models involving these dimensions, Read (1988), Meara (1996a), Wesche and Paribakht (1996), Henriksen, 1999), among others. We then have discussed measures that can be used to evaluate vocabulary content in free writing and speech, particularly in the area of fluency, density, diversity, sophistication and accuracy, as well as from the perspective of formulas. We have seen

Literature review

that by combining quantitative indices of lexical proficiency with qualitative analysis of native-like selection, we may gain a better insight into the vocabulary knowledge of L2 learners and achieve more conclusive findings. Finally, to conclude Chapter 2, we addressed the relationship between written and oral modalities, as it is of particular interest in the present study.

Chapter 3

The study

In the previous chapter, we have seen that despite a wide body of research generally pointing to SA as the learning context that maximises advancement in the foreign language, certain linguistic areas are relatively under-investigated and there are still some open-ended questions with regard to the efficiency of SA across skills and modalities. Oral and written vocabulary knowledge is clearly one of those areas, as few studies have related SA and L2 vocabulary acquisition, and even fewer have looked at the impact of SA on vocabulary in relation to both written and spoken production modes. To the best of our knowledge, only three studies have combined both modalities along the SA line of inquiry (Freed et al., 2003; Pérez-Vidal et al., 2012; Serrano et al. 2012), yet none of them have looked at vocabulary beyond the conventional word counts and lexical diversity indices, suggesting that additional research in this area is sorely needed.

In light of these gaps, the present study expands upon the findings of these earlier contributions and aims at leading us to a better understanding of the understudied effects of SA on EFL learners'

vocabulary, as measured in writing and speech samples. In the following sections, we provide more background to the empirical study developed in this dissertation. First, we present a brief account of the institutional context in which our research has been conducted (Section 3.1) and outline the specific objectives and research questions posed (3.2). We then turn our focus to the methods involved in the study (3.3). Here we cover the design of the study (3.3.1), the description of the participants (3.3.2) and the two learning contexts under investigation (3.3.3). Next, we address the process of data collection (3.3.4), describing the instruments and procedures that were used to evaluate participants' oral and written vocabulary, as well as their global initial proficiency. We also describe the process of data analysis (3.3.5), detailing the transcription procedures and the methodological decisions that were made before conducting statistical analyses. We end this chapter with a summary of the computational tools used to analyse vocabulary in different domains and the specific measures selected (3.3.6).

3.1. The SALA project

The data collected for the present study belong to a large-scale, state-funded research project called 'Study Abroad and Language Acquisition' (SALA), based at a public university located in Barcelona and carried out in collaboration with a second public institution in the Balearic Islands (see Pérez-Vidal, 2014, for a detailed description). This project was developed to track the acquisition of English by Catalan/Spanish undergraduates during SA in contrast with the acquisition in the preceding FI learning context AH. SALA followed

learners at 4 different points in their language development over a period of 2.5 years: (1) at onset, prior to any learning condition; (2) after a period of classroom instruction in their home university, FI; (3) upon return from a 3-month period abroad, SA; and (4) some months later (with no further English instruction) to observe retention. SALA was designed to take advantage of the specific conditions of participants' home institution, where spending 3 months abroad (mostly in English-speaking countries) after a period of formal classroom instruction was a graduation requirement, fully integrated into their degree curriculum. These university-specific characteristics allowed SALA to successfully collect data from three different cohorts of participants over the course of several years, looking at a wide array of skills through oral, written, listening and perception tasks (See Appendix 1 for a detailed list of SALA tests). To further add to an already large pool of informants, SALA also compiled baseline data from native speakers of English. Over the years of its operation, the project has benefited from grants from the Spanish Ministry of Science and Innovation, and the Catalan Government.

The extensive years of work on the SALA project have resulted in a vast string of publications (e.g., Mora, 2008; Pérez-Vidal & Juan-Garau 2011; Pérez-Vidal et al., 2012; Trenchs-Parera, 2009), four doctoral dissertations (Avello, 2013; Barquin, 2012; Lara, 2014; Valls-Ferrer, 2011), and the recent volume *Language Acquisition in Study Abroad and Formal Instruction Contexts* (Pérez-Vidal, 2014a) that we have already mentioned in our review of the literature in Chapter 2. Drawing from the SALA corpus, our study examines the written and oral vocabulary elicited from advanced learners of English through two main tasks: a

written composition and an oral interview, administered during three data collection times. Two global proficiency tasks (grammar and cloze tests) are also included in the analysis so as to gauge learners' initial level of English. In addition to these tasks, we also make use of two questionnaires used to gather information on learners' language backgrounds and the SA conditions (SAC) during their exchange sojourns. We reserve a description of each task and data collection procedures for coming subsections and now we turn our focus to the specific objectives and research questions that have motivated the current study.

3.2. Objectives and research questions

The focus of the present study is to track learners' vocabulary development in written and oral production, separately and in contrast, over time and after two different learning contexts, FI and SA, experienced one after the other. Firstly, we seek to examine the impact of two subsequent learning contexts, FI and SA, experienced by advanced EFL learners on their productive vocabulary both in oral and written mode. We consider learners' productive vocabulary in free writing and speech and analyse it in terms of quantitative measures in the domains of fluency, density, diversity, sophistication, and accuracy, as well as through a qualitative analysis examining learners' native-like selections. We also include baseline data from native speakers of English, as elicited through the same tasks (written composition and oral interview), so as to determine whether learners' productive vocabulary approached that of native speakers in any significant way as a result of FI or SA. Additionally, with NS data in hand, we can

The study

enhance our understanding of the quantitative measures selected for analysis and enrich our interpretation of the changes occurred. Secondly, we aim to determine the amount of gains achieved at each learning context (FI vs. SA) and in each production modality (writing vs. speech), in order to find out whether learners experienced more gains after either FI or SA, and whether these gains were greater in one modality than in the other. Central to this aim is the relationship between the two modalities in relation to vocabulary gains. Finally, we seek to find out if and to what extent initial level of EFL proficiency affects learners' outcomes after FI or SA learning periods. To that end, we explore the impact of onset level in two different ways: as a function of learners' initial vocabulary knowledge – lexical proficiency – (as measured through quantitative measures in each domain) and as a function of their initial global proficiency (as measured through a combined score on grammar and cloze tests).

In order to accomplish these objectives, we formulated three main research questions that guided our analysis and discussion developed in coming chapters. We break down each research question into several sub-questions to address the main question in a more systematic way.

Research Question 1 (RQ1): How does productive vocabulary by advanced EFL learners develop in each modality (writing and speech) longitudinally, after two different learning contexts (FI and SA) and in contrast with NS production?

Chapter 3

RQ1a) Does learners' written and oral productive vocabulary improve significantly over time, and after either the FI or SA learning contexts?

RQ1b) Does learners' written and oral productive vocabulary approximate native-like norms at any point over the observation period in terms of quantitative lexical measures?

RQ1c) Does learners' written and oral productive vocabulary approximate native-like norms at any point over the observation period in terms of qualitative native-like selections?

Research Question 2 (RQ2): In which context (FI vs. SA) and modality (writing vs. speech) do learners accrue greater vocabulary gains?

RQ2a) Do vocabulary gains, if any, in written and oral production modality, respectively, accrue to a larger extent in one context than another, when comparing FI and SA?

RQ2b) Do vocabulary gains, if any, in FI and SA, respectively, accrue to a larger extent in one modality than another, when comparing written and oral production?

Research Question 3 (RQ3): Are there different patterns of vocabulary development and gains associated with learners' onset level?

RQ2a) Are there different patterns of vocabulary development and gains associated with learners' onset level, as measured through initial vocabulary knowledge?

RQ2b) Are there different patterns of vocabulary development and gains associated with learners' onset level, as measured through initial global proficiency?

3.3. Method

Having described the academic context in which our research has taken place and outlined the particular objectives and research questions formulated above, we now turn to the methods involved in the empirical study.

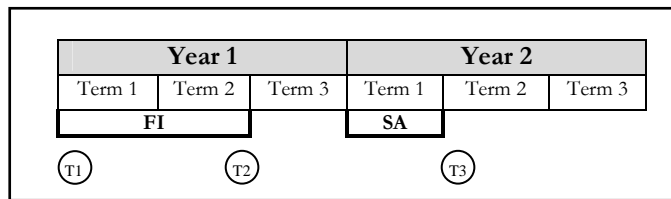
3.3.1. Design

As we have mentioned above, SALA was developed to capitalise on the particular conditions of the university, where students were required to go on a stay abroad following a period of formal classroom instruction. In order to evaluate the linguistic benefits of two different learning conditions (FI followed by SA), a repeated measures within-subjects design was adopted, in which participants' performance in one context was measured against their own performance in a previous context. Although we acknowledge the possibility of task repetition effects when using the same instruments over the different treatment periods, the key advantage of this type of design is that it neutralises the problem of between-subject variability (Milton & Meara, 1995).

Following a longitudinal approach, the empirical study contained in this dissertation comprises three data collection times (T1, T2, T3) over a 15-month period, which included the first two years of the participants' degree. It should be noted that the academic year of students' degree is organised around three terms, with each term

lasting one trimester. Thus, firstly, learners were tested upon their entrance at university (T1). They were tested for the second time (T2) 6 months later, after an 80-hour period of English FI in their home university. The third data collection (T3) took place immediately after students returned from their 3-month SA in an English speaking country. In sum, while the data collected at T1 aimed at establishing the participants' initial level before FI, the T2 functioned both as a post-test for T1 and a pre-test for T3, being T3 a post-test for T2. The diagram of testing times is illustrated in Figure 3.

Figure 3. Diagram of SALA data collection (adapted from the SALA project design, see Pérez-Vidal, 2014b).



3.3.2. Participants

Our participants consisted of two different groups: a group of EFL learners, longitudinal subjects who were recruited at three data collection times so as to examine their vocabulary development across different learning conditions, and a group of native speakers of English, who provided baseline data for comparison, yet were tested on only one occasion.

The group of EFL learners included in our analysis consisted of undergraduate students ($n=30$) enrolled in a language specialisation

degree in Barcelona, with English as their primary foreign language. The sample was mostly composed of females in line with the demographic reality of their degree. A similar distribution pattern was found in the case of native speakers, with females over-represented as compared to males, thus allowing for more control of possible gender effects in both groups, as illustrated in Table 2. Learners' average age at the beginning of the study was 17.93 (range 17-21), with the vast majority (93.3%) having entered the university at the age of 17-18.

Table 2. Participants' demographics.

Participants	n	% Female	% Male
EFL learners	30	80% (n=24)	20% (n=6)
Native speakers	27	77.8% (n=21)	22.2% (n=6)

All students were bilingual in Catalan and Spanish. In their linguistic background questionnaires, the majority reported being Spanish/Catalan bilinguals (76.7%), having been exposed to both languages from birth²³, while the remaining 23.3% grew up in either a Spanish or Catalan household but acquired the second language later in life. As for their foreign language background, they all followed the official curriculum implemented by the Spanish Law of Educational System (LOGSE) and took EFL in primary and secondary schools (mean AOL=7.0). Under the LOGSE curriculum, students completed 12 years of EFL classroom instruction, resulting in a total of 1,050 hours (Pérez-Vidal, 2014). The intensity of exposure to English corresponded to 2.5 hours per week from 8 years (Grade 3) to 15

²³ Catalan is a co-official language, together with Spanish, in Catalonia. It is normal for students in this region to grow up competent in both languages and be considered bilingual.

years of age (Grade 10), and 2 hours a week during Grades 11 and 12, respectively (Muñoz, 2006). Apart from the official hours of English language courses taken at school, the majority of students (86.7%) reported having attended extra-curricular lessons (either at a language school or through private tutoring) at some time during their secondary education.

In order to qualify for higher education at university, students had to pass a university entrance examination, *Prova d'Accés a la Universitat* (PAU). The minimum mark required to enter the translation degree programme during the years of data collection was 6.72 (out of 10) on average²⁴. Additionally, for acceptance into the English translation stream, students had to accredit having at least a B2.1 level within the Common European Framework of Reference for Languages (CEFR). This was done through an entry test in English at the start of the degree.

The group of native speakers (NSs) provided baseline data and was made up of 27 young adults from the USA (16), the UK (7) and Ireland (4). The NSs were also undergraduates enrolled in an exchange programme at different universities in Spain and highly comparable to the EFL learners in terms of age, educational backgrounds and gender (see Table 2). They were tested once, using the same battery of tests as the EFL learners.

²⁴ Accessed via <http://www.upf.edu/universitat/retiment/estudis/tra.html>.

As we have already touched upon, NS productions were used as baseline data for comparison purposes, and to measure the potential development towards native-like norms as a result of the SA experience. Although some authors denounce the use of the idealised native speakers' competence as "the golden benchmark for investigating linguistic development in an additional language" (Ortega, 2014: 34) and aim at exploring new models, in which NS patterns are not invoked, NS data still continue to be a widely used point of reference in research on SLA. When in 1999 Cook proposed his 'multi-competence' approach as a new challenging perspective to tackle the area of adult L2 learning, he justified the adoption of the native-like patterns, as long as it did not make NS language "the measure of final achievement in the L2" (p. 190). A decade later, Foster and Tavakoli (2009) concluded that NS-NNS comparisons completing the same task under the same conditions were important in detecting if there was something other than L2 extra processing costs affecting learners' performance.

All participants took part in the project on a voluntary basis and in accordance with the ethical standards set out by the university, guaranteeing to keep their identities confidential. At the end of each testing session, participants received a financial reward in exchange for their participation.

3.3.3. Learning contexts

As we have already avowed, EFL learners were exposed to two sequential learning contexts: first, receiving English formal instruction

(FI) at their home university during the first two terms of their degree, and then participating in a three-month SA period at the beginning of their second academic year. We describe the two contexts of acquisition as follows:

3.3.3.1. Formal Instruction (FI)

The FI learning context consisted of two English language courses, which were required for all English majors in the first two terms of their first academic year at university. These courses were designed to develop a solid CEFR B2.2 level by the end of the second term and resulted in a total of 80 hours of in-class instruction. Although the courses were divided into two trimesters, the second course was the continuation of the first one in that it followed the same organisation and intensity with the progressive development of contents.

As for the structure of the courses, each of them consisted of lectures and seminar sessions with English as the medium of instruction. While the contents of the lectures covered formal linguistic analysis and morpho-syntactic aspects of the language, the seminar sessions were more practice-oriented and intended to enhance students' general competence of EFL in the four language skills (reading, writing, listening, and speaking). No explicit training in the area of vocabulary was provided during the course. Exposure to English and opportunities for practice were largely restricted to the classroom. There was no further FI during the third term of the first academic year.

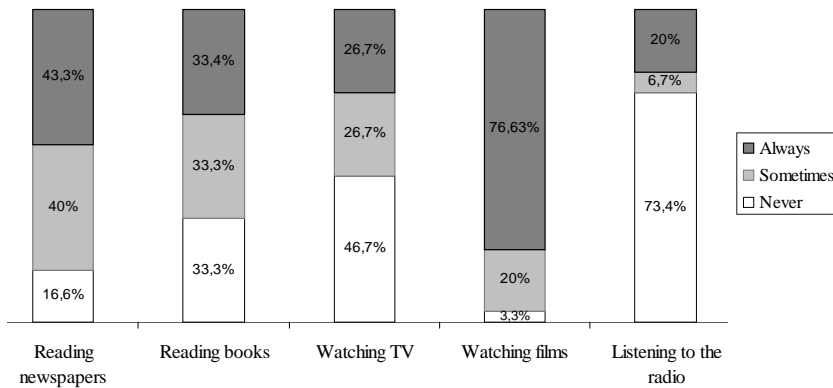
3.3.3.2. Study Abroad (SA)

During their second academic year, all EFL learners embarked upon a compulsory 3-month long (mid-September to mid-December) SA programme in an English-speaking host university. The exchanges were facilitated through the Erasmus scheme, with a minimum 90-day period established for a successful completion of the programme (see Pérez-Vidal (2014) and Beattie (2014), for a comprehensive account of the mobility scheme that inspired the project). The majority of undergraduates went to the UK (93.3%), while a small percentage completed their SA in the USA (6.7%). During this period, they were mainly enrolled in Modern Languages and Humanities departments. As indicated in the Study Abroad Conditions (SAC) questionnaires, it was their first substantial trip abroad, lasting more than a month. Students departed in groups of between 3 and 10 to the same university of destination, with no local professor accompanying them. For one third of the students, the SA was also their first trip to an English-speaking country. The rest reported having travelled abroad for rather short periods of time (usually 2-3 weeks), either for the purpose of studying English or for tourism.

Although no official across-the-board requirements were specified, all students took a minimum of 4 academic courses while abroad (Beattie, 2014). Most of them registered for modules focused on translation and their second foreign languages (German or French), alongside other optional, free-choice courses, depending on their host institution's regulations. According to self-reports, learners received an average of 9 hours of classroom instruction per week.

Overall, students lived in single rooms in university residence halls (60.7%), shared apartments with other students (25%), or stayed with host families (14.3%). In their SAC questionnaires, they reported a fairly high degree of contact with native English speakers (83.3%) as well as with other international students (86.7%). They were also exposed to a greater amount of L2 input through the host country media, as revealed in the questionnaires: watching British and American films (76.63%) was their most common receptive-skill activity (Figure 4).

Figure 4. Learners' contact with English media during SA.



In addition to self-reporting on social interactions with English speakers and the amount of contact with the host country media, students also provided information on their perceived improvement in different linguistic areas of EFL competence. As part of the SAC questionnaires, they were asked to rate their progress in different language skills (speaking, listening, reading, writing, vocabulary, grammar, and overall skills) on a 5-point Likert scale, with 5 staying

for the most improvement and 1 for the least improvement. In their subjective assessments, students felt that their oral-aural skills had improved the most (both with 4.30 out of 5), followed by vocabulary, which was given the third highest rating (4.15). The skills that received the lowest rating were grammar (3.11) and writing (3.19). The data for learners' self-assessed improvement in each skill area is reported in Table 3 below.

Table 3. Learners' mean perceived improvement in EFL proficiency after SA.

Language skills	Rating	5=most 1=least
Overall skills	4.07	
Speaking skills	4.30	
Listening skills	4.30	
Reading skills	3.59	
Writing skills	3.19	
Vocabulary	4.15	
Grammar	3.11	

Once back at home after the SA period, students returned to their home institution and followed the regular translation degree curriculum.

3.3.4. Instruments and procedures

At each data collection time, EFL learners were asked to complete a series of tests designed to measure their general English language competence, with each test focusing on a specific language skill. Learners took the tests in exam-like conditions on the university campus outside of class time. The tests were timed, and the whole battery of tests was completed over the course of 2 hours. No additional material (e.g., dictionaries, grammar books) was permitted

for use during the testing sessions. Baseline data from NSs were collected following the same protocol, using the same instruments and procedures. A complete list of the SALA tests is documented in Appendix 1.

In order to maintain participants' anonymity in the compilation of the corpus, each subject was assigned a unique alpha-numeric code during the first data collection session. The same codes were later used at the consecutive testing times and at all stages of the subsequent data analysis. In our study, we used the written task (argumentative essay) and the oral task (structured interview) to compile written and oral data, and two tests of general EFL proficiency (a sentence-rephrasing task and a cloze test), used to provide additional information on learners' initial level of English. The description of each task is provided in the corresponding sections below.

3.3.4.1. Written task

The written task required participants to write an argumentative essay in response to the following prompt:

Someone who moves to a foreign country should always adopt the customs and way of life of his/her new country, rather than holding on to his/her own customs.

The task was administered in the same conditions at all testing times and followed the same procedures: participants were given a ruled, double-sided sheet of paper and were allowed 30 minutes to write their compositions. No minimum or maximum text length was

specified. All participants wrote their essays by hand within the allotted time period. Although they were allowed to request extra sheets of paper if necessary, none of the essays considered in this study exceeded one double-sided page in length.

We chose to assess productive vocabulary through free written expression, as written essays and compositions have high validity for assessment in higher education, being the best sources to obtain real language and as sufficiently complex as to discriminate between skilled and less skilled learners (Laufer, 1998). Furthermore, they provide very valuable data for vocabulary knowledge and error analysis (Agustín-Llach, 2011), given that a well-written piece of writing is likely to make effective use of vocabulary (Laufer & Nation, 1995). Finally, the specific topic was selected in an attempt to engage students in critical thinking around a phenomenon they would experience first-hand during their SA, as well as assess their intercultural awareness throughout the course of the observation period, as explored in other SALA studies (Merino & Avello, 2014). The decision to maintain the same topic at each measurement time was made in light of evidence that a change in topic may affect learners' written narrative and result in uncontrolled variation in their vocabulary use (Laufer, 1998).

3.3.4.2. Oral task

The oral task consisted of a semi-guided interview, in which participants were asked to act as both interviewers and interviewees. They were presented with a set of seven questions (See Appendix 2) written in the instructions and were randomly assigned 'Student A' or

'Student B' roles. All the questions were different but centred on the same topic: university life.

The oral interviews took place in a quiet room and were digitally recorded. Once the subjects were properly identified before the start of the recording, the interaction was initiated: first, 'Student A' asked 'Student B' the questions one at a time and then they switched roles²⁵. The allotted time for the task was up to 5 minutes. A member of the research team was present as an observer and only intervened to solve any technical problem or inform the test-takers of the approaching time limit. Furthermore, participants were explicitly instructed to perform the interview as if the researcher was not there, favouring a more relaxed atmosphere, as well as a balanced and spontaneous interaction.

The reason for focusing the conversation around students' life at university was that of topic familiarity. As previous work by Foster and colleagues has shown (Foster & Skehan, 1996; Foster & Tavakoli, 2009; Skehan & Foster, 1997; Tavakoli & Foster, 2008), task-takers' familiarity with the content is one of the features of cognitive complexity known to have an impact on L2 performance. Talking in an L2 about well-known information (i.e., oneself or one's personal experience) typically entails a more fluent and accurate production and prevents speakers from running out of ideas. Since we aimed to elicit as natural and close to real-life speech as possible, getting learners to perform a low-complexity task of this kind was a priority.

²⁵ Only the answers to the questions were considered in the analysis.

At this point, it should be noted that as roles were assigned at random, some participants did not repeat the same questions assigned at T1 in the subsequent testing sessions. This was the case on 13 occasions (14.4% of the recorded interviews). Despite these minor issues of inconsistency in task conditions, all questions were linked to the same topic and the resulting speech productions did not vary substantially from ‘Student A’ to ‘Student B’. For these reasons, it was deemed preferable to include these cases, so as to prevent the loss of valuable longitudinal data.

3.3.4.3. General proficiency tests

Two additional SALA internal tests were included in the analysis and consisted of a sentence-rephrasing task and a cloze test. Together, these tests were intended to assess participants’ general EFL competence and helped us in gauging their initial proficiency level. They were administered during the 2-hour group session and participants had 15 minutes to complete each of the tests.

In the rephrasing test, participants were asked to rewrite 20 sentences, given a new initial structure, while preserving the meaning of the original prompt. An example sentence manipulation was included in the instructions. Grading this type of test was not simple, as some of the sentences invited multiple ways of rephrasing to get a similar meaning. Therefore, the decision was made to give full or partial credit for any response, which was grammatically correct and stayed close to the original sentence, following Barquin (2012). For example, in item

Chapter 3

3., full credit was given for both *He slammed the door* and *He slammed the door shut*.

First three items from the sentence-rephrasing task

- | |
|---|
| <p>1. <i>Please don't use the shower after midnight.</i>
<i>Would you mind</i></p> <p>2. <i>The weather was fine at the seaside last Saturday.</i>
<i>We had</i></p> <p>3. <i>Mark shut the door with a slam.</i>
<i>He slammed</i></p> |
|---|

The cloze test presented participants with a 286-word text entitled *The lady who liked the adventure*. The story was divided into three paragraphs and contained 20 gaps. Students had to fill in the gaps with a single acceptable word that best fit the context. Some blanks could have more than one appropriate answer. Again, although one correct answer was possible per gap according to the original text, full or partial credit was assigned for any response that was semantically and grammatically accurate given the specific context. For example, in item (3) below, full credit was awarded for the expected response *try*, while partial credit was given when the response was *put*.

First four items from the cloze test

<p style="text-align: center;"><i>The lady who liked adventure</i></p> <p><i>It was one of those impulse buys that can happen while shopping. Mary Bruce was in London looking for a nice dress (1) she noticed a showroom with a light aircraft for (2) at a terribly reasonable price. Mrs. Bruce went away to (3) on a dress. It did not suit her. The plane (4).</i></p>

The two tests together were considered to be measures of implicit language knowledge: the sentence rephrasing functioning as a focus-on-form task, and the cloze test used as a global test within the SALA battery of tests (See Juan-Garau et al., 2014, for a detailed account of the tests). SLA research has generally favoured a positive view of the cloze test in the evaluation of general language ability, if proven to be valid and reliable, and reported high correlations with writing (Fotos, 1991) or general language ability (Hughes, 1989). In our study, concurrent validity of the sentence-rephrase and cloze tests had been assessed in earlier SALA research (e.g., Juan-Garau et al., 2014), showing acceptable internal consistency for both instruments. Furthermore, previous SALA studies have combined learners' scores on the two tests into a global measure of proficiency, used to operationalise as an independent variable 'initial level' (Barquin, 2012; Lara, 2014). We also adopt this approach in order to assess learners' initial proficiency and use composite rephrasing-cloze scores as an alternative measure to assess their onset level.

3.3.5. Data analysis

In this section, we review the procedures followed in the transcription process of our data. We also detail some of the methodological and editing matters that may have implications for analysis and the subsequent results, and describe the different computational tools (VocabProfile, D_Tools, CLAN, and CLAWS part-of-speech tagger) used for analysis in the present study.

3.3.5.1. Transcription

A total of 117 essays (90 by learners and 27 by native speakers) and 115 interviews²⁶ were selected for the current study. They were transcribed by SALA research assistants, including the author, and later formatted taking into account the construct analysed and the specific requirement of the computational tools used for our intended purposes. With this in mind, three different data sets were created, as illustrated in Appendix 3: a) raw data in a Word document (.doc), b) plain text files (.txt), and finally c) the CLAN-specific format (.cha).

The first stage of transcription involved typing all essays in Word, with auto-correction disabled. This way, the resulting CHAT files would be coding ready. An attempt was made to keep the transcriptions as close to the original handwritten texts as possible. Thus, misspellings and punctuation errors were retained. Finally, the transcripts were double-checked against the original paper essays to ensure accuracy. Interviews were handled in a similar way and were transcribed verbatim preserving every feature of the natural language. Incomprehensible speech was marked with 'xxx'. Lastly, every transcription was double-checked while listening to the recording, and a few minor corrections were applied. Data checking and editing were performed by a single researcher, the author, with the full corpus totalling 27,774 words for compositions and 30,370 words for interviews. Once this set of 'raw' written and oral transcriptions was

²⁶ A few native speakers did not complete the whole test battery. In our sample, two of them were unable to perform the interview task.

generated, multiple copies were made and saved for use with different formatting (e.g., .txt or .cha files).

3.3.5.2. Methodological decisions

In an attempt to enable comparisons with other studies and make future replications possible, it was deemed essential to make explicit which features were included in the analysis, especially those that may affect the final word count. We, therefore, provide clear indications of the standardisation of oral as well as written data used in the present study.

Oral data

As most spoken language is by nature spontaneous and unplanned, our speech samples were untidy and contained a great deal of disfluency. For the sake of consistency with previous studies on vocabulary (e.g., Broeder, Extra, van Hout, Strömquist, & Voinmaa, 1988; Levelt, 1989; Miralpeix, 2008; Richards & Malvern, 2000) and in order to ensure that the total amount of types and tokens produced was neither underestimated nor overestimated, transcribed oral corpus was pruned by excluding false starts, self-repeats, and hesitation markers such as *hm*, *eh*, and *uh*. Thus, given the utterance in (1), the first unfinished sentence and the pause marker *eh* were taken out and are contained in brackets {...} for demonstration purposes.

- (1) {it's similar because... eh} it's similar to what I think because

Another problem was that of lexical inventions or non-words. Even though their use was practically non-existent, with only 0.08% of the total learners' oral corpus, they were also removed from the calculation so that the level of lexical sophistication would not be overestimated (see Appendix 5.2 for a complete list of non-words). Non-words were treated differently for the domain of accuracy and are further explored below. Special attention was given to L1 Catalan/Spanish words that, though infrequent, were not included in the analysis. In most occasions, however, using L1 words (e.g., *Batxillerat*, *paella*, *siesta*) was a communication strategy constrained by the specific local context rather than an indication that the learner was unfamiliar with its English equivalents. As learners had no difficulty making themselves understood and wrong pronunciation was not severe enough to make speech unintelligible, pronunciation errors were disregarded. Examples of pronunciation errors mostly included phonemic substitutions and stress misplacements, as shown in utterances (2)-(4):

- (2) I find here an old friend of the **primary** [pronounced as /'primari/] school
- (3) Well, I think that **classes** [/'kleisis/] are well
- (4) It was a **catholic** [/ka'tɒlik/] school

Abbreviations and shortened forms of words (e.g., *tv* – *television*, *'cos* – *because*, *yeah* – *yes*, *wanna* – *want to*) were converted to their full forms, as were contractions (e.g., *doesn't* – *does not*, *he'll* – *he will*). Compound words were counted as one lexical entry (e.g., *high-school* – *highschool*, *open-minded* – *openminded*), following Broeder et al. (1988), and Miralpeix (2008). Finally, in order to control for the amount of proper nouns

The study

and improve the accuracy and reliability of computational analysis, it was deemed important to follow the criteria offered by Navés (2006). Thus, specific names of people, places, and organisations (e.g., *Claudia*, *Ciutadella*, *Plaça Catalunya*, *UAB*, *UPF*, etc.) were carefully edited and are listed in Appendix 4.

Written data

For the analysis of the written production, we followed the same conventions regarding contractions, lexical inventions, L1 words, hyphenated words and proper nouns. Words with more than one acceptable spelling (e.g., *globalisation* - *globalization*, *burka* - *burkha*) were standardised to maintain consistency. As punctuation errors were not a focus of the present study and they could in fact negatively impact the performance of computational tools, especially those working with plain text format (e.g., D_Tools, VocabProfile, CLAWS), full stops, commas, exclamation points and other punctuation marks were entirely eliminated.

Spelling errors and non-words

Different procedural decisions were made when handling spelling errors and non-words in the data depending on the focus of analysis. For analysis of lexical diversity and sophistication, spelling errors were corrected and included in the final word count. This was done so that a learner who erroneously produced words, such as *foreing* or *althought*, would not be given credit for having a more sophisticated vocabulary

than a peer who correctly produced *foreign* and *although*. While the correct two are considered common words and are classed as first (1K) frequency band by the VocabProfile²⁷ package, the first misspellings, if not corrected, would be counted as infrequent or difficult words (Off-list), as they are simply not recognised by the programme. For analysis of other linguistic characteristics - such as lexical accuracy - spelling errors were meticulously logged and classified according to the nature of the error. Following Arnaud (1984), we distinguished between major and minor spelling errors. Major errors referred to words that were consistently misspelled as a result of lack of orthographic knowledge (e.g., *bebabe*, *enfazice*) or were part of the problem with homophones (*there/their*, *affect/effect*). Minor errors, on the contrary, encompassed slips of the pen that arose as a result of careless handwriting. Slips of the pen were typically “attributable to lapses or inattention on the part of the writer, which would have amended if they had been noticed,” in Montgomery’s terms (1997: 88). We logged spelling errors as slips of the pen when they were high frequency and there was evidence of correct usage elsewhere in the text (e.g., *alway – always*, *peope – people*).

As with spelling errors, non-words were treated differently depending on the focus of the analysis. As we have already mentioned above, non-words were filtered out to render more accurate profiles of participants’ lexical sophistication. Thus, instances such as *conscient* instead of *conscious* or *unrespectful* meaning *disrespectful* would be erroneously tagged as low-frequency Off-list items, had they not been entirely eliminated. For the analysis of lexical accuracy, non-words

²⁷ The description of the VocabProfile programme is given in Section 3.8.3.1.

were, in contrast, left intact and coded as lexical errors. Most non-words were due to L1 transfer and direct borrowings (e.g., *conscient* from the Spanish/Catalan adjective *consciente/conscient*), or resulted from ‘creative’ use morphology, where L1 was not necessarily the source of the error (e.g., *unrespectful*). All spelling errors and non-words are reported in Appendix 5.

3.3.5.3. Computational tools

Here we present the different computational tools selected to analyse learners’ productive vocabulary quantitatively in the domains of fluency, density, diversity, sophistication, and accuracy. We describe the four tools used (VocabProfile, D_Tools, CLAN, and CLAWS part-of-speech tagger) in detail below.

3.3.5.3.1. VocabProfile

Once the editing of participants’ spoken and written production had been completed, we proceeded to obtain the basic estimates of lexical production: the total number of words used, *tokens*, and the number of different words, *types*. This was carried out with Cobb’s (2016)²⁸ version of VocabProfile, an online programme originally based on both Nation and Heatley’s original *Range* (1994), and Laufer and Nation’s (1995) *Lexical Frequency Profile* (LFP).

²⁸ VocabProfile is available at <http://www.lextutor.ca>.

The procedure for running VocabProfile was quite straightforward: plain texts were copied and pasted into the web-based tool, one by one, until the entire corpus had been processed. The tool automatically generated a set of statistics about each text, including the number of types and tokens, and displayed a colour-coded version of the text based on the frequency of each word. In the output text, all words were categorised on the basis of four frequency bands: a) the first 1000 most frequently used words (K1), b) the second 1000 (K2), c) the Academic Word List (AWL), and d) the remainder not found on the other lists (Off-list). These frequency bands were indicative of learners' lexical sophistication (Appendix 6 shows the output of a text submitted to VocabProfile). All words classified as *off-list* were scrutinised to ensure that they were not due to technical flaws. As already mentioned, proper nouns and common words not recognised by the programme (e.g., *ok*, *Tupperware*, *video*, *Internet*) were excluded to prevent being identified as off-list items.

The Greco-Latin cognate index, for its part, was calculated using a recently created VP-Compleat, also available at Cobb's (2016) website. With the Cognates box activated, the tool computed Greco-Latin index (GLI) or 'cognateness', by dividing the Anglo-Saxon words by the words from Greco-Latin origin. When all the files were analysed, the number of types and tokens was recorded for each text, as were the percentages of words from each frequency bands and the cognate ratio. These data were entered into an Excel spreadsheet and then SPSS for statistical analysis.

3.3.5.3.2. D_Tools

Lexical diversity, as calculated by D, was analysed by means of the D_Tools software (Meara & Miralpeix, 2007)²⁹. As described earlier in Chapter 1, the D statistic was originally devised by Malvern and Richards (1997) and computed by *vocd* programme. The decision to compute the index using D_Tools stemmed from practical reasons: whereas *vocd* required specific data formatting, D_Tools accepted plain text files, a format previously used in VocabProfile. This was believed to be a considerable advantage, given that “facilities in computation often lead to more systematic testing” (Miralpeix, 2008: 115). Once installed, a text of interest was selected and loaded into the programme, with a default value of 100 set as a sampling rate³⁰. The programme then ran the analysis and generated the best-fitting curve to describe the lexical richness of the text. This value is reported on the box labelled D_estimate. The functioning of the software is not discussed as part of the analysis, yet can be consulted in Meara and Miralpeix’s (2008) *D_Tools: the Manual*.

In our data, three essays yielded a value above 100. The checking procedure on these statistics confirmed that those cases were anomalous and required special caution. To prevent reducing our sample size by eliminating these values, we deemed it best to replace them with the mean of the variable, as the abnormal values only

²⁹ This software is freely available <http://www.lognostics.co.uk/>.

³⁰ This would make 1600 text samples. For further explanation see *D_Tools: the Manual* available at www.lognostics.co.uk/tools/.

represented less than 5% of our data. The D scores were recorded in Excel and then entered into SPSS for further analysis.

3.3.5.3.3. CLAN

Analysis of lexical accuracy was performed using Computerized Language Analysis (CLAN) software package (MacWhinney, 2000), extensively documented in the manuals and available online³¹. To use the tool, oral and written data were converted into .cha format and edited according to the CHAT transcription conventions. In order to avoid oversights or keep them to a minimum, files were checked regularly by running the CHECK command within the editor. After formatting was verified and reported to be successful, all texts were coded, one by one, following a set of accuracy codes developed by SALA researchers (Barquin, 2012; Lara, 2014; Pérez-Vidal, 2014) and included in Appendix 7. Errors were coded by type including lexical and spelling errors, and are discussed in Section 2.2.3.2.5., pertaining to accuracy analysis. After the entire corpus was coded, the FREQ command was used to run a count of errors in each category and these were entered into Excel and later imported into SPSS for analysis.

3.3.5.3.4. CLAWS part-of-speech tagger

Analysis of different word class categories as part of lexical fluency and density dimensions was carried out using CLAWS C5, the online

³¹ CLAN is downloadable at <http://childes.psy.cmu.edu>.

part-of-speech (POS) tagging programme³², developed at Lancaster University and accessible online. CLAWS takes plain text files as input and assigns to each word in running text its correct grammatical POS in context. The programme outputs a tagged text. The system is robust and reports a high degree of accuracy (96-97%), with an error-rare of 1.5% (Garside & Smith, 1997).

After the text has been entered, we selected a vertically-presented output style, suitable for manual post-editing. Manual post-tagging editing was considered necessary in categories where the tagging system was likely to fail. This process was far from trivial and required following standardised rules of tagging practice. To accomplish this, we checked erroneous assignments of tags against the *Wordclass tagging guidelines*³³. These included borderline cases and ambiguous taggings and are summarised, as follows.

First, proper nouns were deleted, as in VocabProfile. They represented only 1% of the data and were not considered to be a valuable measure of participants' lexical knowledge, as in Granger and Wynne (1999). Multiple-word lexical items or multiwords were tagged by its individual component parts (i.e., *for example*, were tagged as preposition and noun, rather than two adverbs that the programme outputted). Ambiguous taggings denoted those cases where the choice between assigning one tag and another was left open and proved to be unreliable (see *Ambiguity Tags* and *Disambiguation Guide*³⁴). To clarify

³² CLAWS (Constituent Likelihood Automatic Word-tagging System) can be accessed at <http://ucrel.lancs.ac.uk/claws/trial.html>.

³³ Accessed via <http://ucrel.lancs.ac.uk/bnc2/bnc2guide.htm>.

³⁴ Available at <http://ucrel.lancs.ac.uk/bnc2/bnc2guide.htm#pm>.

this, we will summarise the most common instances of ambiguity in our data, pertaining to lexical items, which belong to more than one word class, and are among the most problematic for disambiguation.

Cardinal numbers and ordinal numbers, either written as words or as figures, were eliminated from the analysis, as they were not a representative part of lexical knowledge. The main ambiguity in this category was between the word *one* functioning as a cardinal number (CRD³⁵) and as an indefinite pronoun (PNI). *One* was tagged as a numeral (and then deleted) in a quantifying noun phrase (measurements, numerical expression, etc.), as in (5), and it was tagged as a pronoun, when it replaced an understood noun (6).

(5) from nine o'clock to **one (CRD)** o'clock

(6) abandon their own culture and embrace a new **one (PNI)**

Some borderline phenomena occasionally seemed unsolvable and the suggestions provided by the guidelines somewhat arbitrary. This was the case for the tagging of adjectives (AJ0) versus present (VVG) or past (VVN) participles. To achieve a reasonable degree of consistency in the tagging process, we made a decision in our first calculation of ambiguity and then adhered to our decision strictly, so that all tags would be applied consistently. Thus, when the distinction was unclear and no syntactic context was provided to differentiate both categories, as shown in (8-9), preference was given to adjectives.

(7) When you are **surrounded** by people (**VVN**)

³⁵ We used CLAWS5 Tagset with 62 tags. The whole list of tags is compiled at <http://ucrel.lancs.ac.uk/claws5tags.html>.

- (8) Some people could feel strange and **shocked (AJ0)**
- (9) your mind is **closed (AJ0)**

Finally, features of spoken data prone to error were checked and corrected. These included discourse particles and fillers (e.g., *like, ok*) and were all tagged as interjections.

- (10) they will not be after you, and **like... warning you (VVB)** (initially tagged as a verb by CLAWS)
- (11) they will not be after you, and **like... warning you (ITJ)** (manually corrected and tagged as an interjection)

After tagging, the tagger outputs were recorded in an Excel spreadsheet and then transferred to SPSS to count frequencies. The different word class categories that are considered in the study are further explained in sections 3.3.6.1 and 3.3.6.2.

3.3.6. Measures

Having described the assortment of tools for computational analysis, we now present the different measures selected. We consider a set of quantitative metrics of productive vocabulary in the domain of fluency, density, diversity, sophistication, and accuracy, alongside qualitative analysis of native-like selections (NLSs).

3.3.6.1. Lexical fluency

The first step in assessing the overall oral and written production of vocabulary by learners was to simply look at how many words they produced to complete both tasks at each testing time. To that end, we

calculated the total number of words used (tokens) and the number of the different words (types). Although the word count by itself provided little information on the quality of learners' vocabularies, type and token counts served as the basis to estimate a more robust vocabulary measure, that is, lexical diversity (see Section 3.3.6.3).

As discussed in Chapter 2, improvement in fluency or lexical diversity may be related to (stem from) increases in a few word classes, although it remains unclear which class in particular benefits the most from such increases (Broeder et al., 1993; Bulté et al., 2008). In light of evidence that word class distinctions may play a central role in language production (Bulté et al., 2008; Hatch, 1983, as cited in Broeder et al., 1993), we calculated separate measures for the following lexical categories³⁶ – adjectives, adverbs, nouns, verbs, conjunctions, prepositions, and pronouns – and then grouped them into two major global classes, content words (adjectives, adverbs, nouns, and verbs) and function words (conjunctions, prepositions, and pronouns).

The overall process of word-class tagging was performed using CLAWS C5, in accordance with the procedures spelled out in Section 3.3.5.3.4. Each word in transcribed essays and interviews was tagged with its POS label, identifying an extensive set of word sub-classes. Thus, the category for adjectives, for example, had three different tags representing each linguistic constituent; the unmarked form (AJ0), the

³⁶ As Polinsky (2005: 419) notes, “The very notion of word class has many names in linguistic literature: aside from ‘word class’, the terms ‘lexical category’ and ‘part of speech’ are also often used.” In the present study, we will be using all these terms interchangeably.

comparative adjective (AJC), and the superlative adjective (AJS) forms (See Appendix 8 for a detailed word class taxonomy). Once the different constituents were counted for frequency of occurrence, it was deemed convenient to combine them into major classes. Following this decision, the aforementioned ordinary, comparative and superlatives adjective forms made a single word class for adjectives; just as singular and plural common nouns comprised a single class for nouns. Similarly, verbs covered all existing forms irrespective of person, tense, aspect, or modality (i.e., the different labels for *take*, *takes*, *taking*, *took*, *taken* ultimately represented a single verb class). After all major forms were calculated, ‘nouns’, ‘verbs’, ‘adjectives’ were combined with ‘adverbs’ to form a global category of ‘content words’. As in Collentine (2004: 234), content words were considered as primary parts of speech carrying semantic information and “provided a measurement of the learners’ core lexical base.”

The remaining word classes represented those lexical units that offered learners discursive coherence and included conjunctions, prepositions, and pronouns. The articles, determiners, auxiliary verbs and modals were not counted as separate parts of speech, yet were included in the umbrella category of ‘function words’ together with the conjunctions, prepositions, and pronouns. As with content words, certain function word classes initially broken down into its constituents were grouped into major categories (i.e., personal, reflexive and wh-pronoun forms comprised a single class for pronouns; see Appendix 8). Finally, a few loose particles and isolates³⁷ that were difficult to classify were excluded from analysis, as were

proper nouns and numerals. Interjections were also eliminated from oral interviews, due to the lack of comparative data and literature support for its inclusion.

Table 4. Lexical fluency measures.

Domain	Measure	Abbrev.	Tool
Lexical Fluency	Types	Types	VocabProfile
	Tokens	Tokens	
	Content words	ContW	CLAWS C5
	- Adjectives	Adj	
	- Adverbs	Adv	
	- Nouns	N	
	- Verbs	V	
	Function words	FunctW	
	- Conjunctions	Conj	
	- Prepositions	Prep	
	- Pronouns	Pron	

3.3.6.2. Lexical density

In addition to presenting raw counts of word classes (entailing a simple frequency of different parts of speech), we calculated the density of each word class in relation to the length of the text. This calculation consisted in dividing the particular word class frequency by the total number of tokens produced. As described in Chapter 2, changes in the POS that predominates in production may provide new insights into the nature and organisation of L2 lexis. Furthermore, a consideration of density metrics controlling for text length was motivated by the fact that learners might produce more words at some point over the observation period and, in turn, generate more

³⁷ This type of tags primarily covered the existential *there* construction appearing in *there is/are* clauses (EX0), the infinitive marker *to* (TO0), and the negative particle *not* (XX0).

instances of each word class (as a result of their greater overall fluency rather than showing real gains in specific word classes), which could consequently distort the results (as has been previously reported in Collentine, 2004). Like in lexical fluency, we computed adjectives, adverbs, nouns, verbs, conjunctions, prepositions, and pronouns separately, and then grouped the first four into the global class of content words, and the last three into function words. The standardised word class ratios were converted into percentages.

Table 5. Lexical density measures.

Domain	Measure	Abbrev.	Tool
Lexical Density	Content word density	ContW/w	CLAWS C5
	- Adjective density	Adj/w	
	- Adverb density	Adv/w	
	- Noun density	N/w	
	- Verb density	V/w	
	Function word density	FunctW/w	
	- Conjunction density	Conj/w	
	- Preposition density	Prep/w	
	- Pronoun density	Pron/w	

3.3.6.3. Lexical diversity

The measures of lexical diversity included in the present study were Guiraud's Index (GI) and D. As discussed in Chapter 2, GI was calculated by dividing the number of types by the square root of the number tokens ($\text{typens}/\sqrt{\text{tokens}}$) and was selected over TTR on the basis of empirical results (Daller et al., 2003; van Hout & Vermeer, 2007) pointing to GI as one of the most stable transformations proposed to rectify the dependence on text length. By taking the square root of tokens, the problem of negative correlations with larger

sample size is more likely to be obviated, unlike with a simple TTR. The counting of types and tokens was obtained from the VocabProfile output, and the formula was calculated with the COMPUTE command in SPSS.

In addition to GI, we also computed an alternative measure of lexical diversity index, D. By calculating D, we were better equipped to understand any potential changes in diversity, as D has also been claimed to be more informative than TTR and especially effective within the text length of 100-400 word tokens (Malvern et al., 2004; McCarthy & Jarvis, 2007). Given that our participants produced an average of 232.59 word tokens when writing compositions and 267.46 words when performing the interview across data collection times, both written and oral data fell within the desired range. The D index was calculated by means of D_tools, as explained in Section 3.3.5.3.2. above.

Table 6. Lexical diversity measures.

Domain	Measure	Abbrev.	Tool
Lexical Diversity	Guiraud's Index	GI	VocabProfile
	D	D	D_Tools

3.3.6.4. Lexical sophistication

Lexical sophistication was analysed through one of the most well-known frequency-based method, Lexical Frequency Profile (LFP), following the procedures described in Section 3.3.5.3.1. (See Appendix for the output of VocabProfile classifying learners' vocabulary into four frequency levels). Word frequency profiles were thought to be

especially interesting measures to include in our analysis, as they attempted to capture the kinds of words learners produced distinguishing between ‘frequent’ versus ‘rare’ vocabulary. Furthermore, the use of less frequent words was considered a potential indicator of proficiency, since a number of studies have shown that the proportion of sophisticated and unusual vocabulary tends to increase as learners advance (Laufer, 1994, 1998; Laufer & Nation, 1995). In our study, the results of each frequency level are calculated in percentages.

In addition to looking at word frequencies, we considered another potentially complementary measure to determine what percentage of learners’ lexis was of Greco-Latin origin: that is, to what extent learners relied on words that are cognate with Catalan/Spanish (both Romance languages) in their English productions, given that thousands of words in English come from Romance roots. The calculation of the Greco-Latin index (GLI) was expected to help us broaden our perspective on the effects of cognates on learners’ overall vocabulary knowledge, based on previous research (Horst & Collins, 2006; Granger, 1993; Miralpeix, 2008).

Table 7. Lexical sophistication measures.

Domain	Measure	Abbrev.	Tool
Lexical Sophistication	First 1,000 words	K1	VocabProfile
	1001-2000 words	K2	
	Academic Word List	AWL	
	Not on the lists	Off-List	
	GL index	GLIndx	V_Compleat

3.3.6.5. Lexical accuracy

The final construct explored in the quantitative analysis was that of accuracy. Along with the measures of lexical diversity and frequency profiles used to assess how varied and sophisticated learners' productive vocabulary was, we also included measures of accuracy to gauge how appropriately these words were used. Accuracy was evaluated in terms of the total number of errors per word (E/w), as well as the number of lexical errors per word (LexE/w), measured separately.

To begin, all errors were identified manually and recorded through a system of codes, using CLAN. Errors were coded by type following a comprehensive error classification scheme adapted from other SALA-based studies (Barquin, 2012; Lara, 2014; Pérez-Vidal et al., 2012) and summarised in Appendix 7. In our corpus, lexical errors (LexE/w) primarily included wrong word choice errors, transfer errors, and non-words. A common source of problematic word usage was due to language-to-language literate translation, either of individual lexical items such as Spanish/Catalan-English false friends (i.e., producing *career* to mean *university degree*, or *residence* instead of *dorms/balls of residence*) or large chunks of language and fixed expressions (i.e., *nothing from another world* to mean *nothing special*). The latter, however, were considerably rare. Many word choices stemmed from incomplete collocational knowledge (i.e., confusing English verbs like *make* vs. *do*) or were not context-appropriate. Lexical errors were also used to mark problems with phrasal verbs, as in the example: *I think the best way would be holding with our own customs...* instead of the correct form *holding on to*.

When the nature of the error could not be clearly attributed to lack of vocabulary knowledge, but rather merged with grammatical aspects of language, it was deemed preferable to regard this type as ‘concomitant grammar errors’ (James, 1998: 147) and include them only in the total error count (E/w). In our data, concomitant grammar errors mostly resulted from inappropriate use or formal misselection of a particular POS (e.g., an adjective substituted for a verb such as *broad* instead of *broaden*), as in the example (12):

- (12) people who *is* really interested in the customs and even try to acquire some of them *to broad* their minds

All errors were tallied on the basis of ratios controlling for text length, which consisted in dividing the total number of errors and its lexical subtypes by the total number of tokens. Spelling errors (MjSp/W and MnSp/W) were also converted to ratios, yet calculated independently. All error ratios were expressed in percentages.

An important point to be made here concerns inter-rater agreement. To ensure reliability, error correction was carried out with the help of a second coder, also a member of the SALA research team. An inter-rater reliability analysis was conducted on 10% of randomly selected texts using the Intraclass Correlation Coefficient (ICC) for average measures, with a two-way fixed effects model and the confidence interval set at .95. The results revealed that inter-rater reliability was acceptable, at .809, indicating that error detection and coding criteria were applied consistently across raters.

Table 8. Lexical accuracy measures.

Domain	Measure	Abbrev.	Tool
Accuracy	Total number of errors per word	% TotE/W	CLAN
	- Lexical errors per word	% LexE/W	
	Major spelling errors per word ³⁸	% MjSp/W	
	Minor spelling errors per word (slips of the pen)	% MnSp/W	

3.3.6.6. Native-like selections (NLSs)

Finally, aside from the quantitative analysis in the domains described above, we explored participants' oral and written productive vocabulary qualitatively. To do so, we looked at learners' ability to sound idiomatic when formulating utterances (i.e., whether they were apt to choose the same word combinations as their NS counterparts at any point over the observation period), thus achieving 'native-like selection' (henceforth NLS). As mentioned in Chapter 2, SLA research supports that the capacity to select the right words and expressions in ways preferred by TL community is an integral component of fluent language production (Foster, 2009; Pawley & Syder, 1983; Wray, 2002).

For the purposes of comparison, we compiled an exhaustive data-driven list of the similarities and differences between learners at three testing times and NSs when formulating the same idea. Unlike the study by Foster (2009), where the lexical choices were extracted from a controlled task (cartoon picture prompts), with learners describing exactly the same action, our both oral and written tasks were characterised by a high degree of freedom. For example, while most

³⁸ Major spelling errors and slips of the pen pertain to the written data analysis, only.

The study

learners brought up immigration and racial issues in response to the essay statement (especially before SA), NSs related the topic to their own experiences living and studying abroad in Spain, which then made the search for similar ideas less than straightforward. Similarly, certain interview questions were too specific and at times encouraged participants to recount anecdotes, yielding as a consequence a number of diverse responses. Finally, one last difficulty in our data-driven classification had to do with the treatment of lexical choices that were originally contained in the essay prompt or used in the interview questions. These were examined with special care and specified when appropriate. Despite all these remarks, we were able to locate a handful of comparable word choices (by using the “Find” command in the Word processor) and extract a few general observations that will be covered in Section 4.1.3. of the next chapter.

Chapter 4

Results

This chapter presents the results of the analyses pertaining to the present study and is organised around the three main research questions that we outlined in Chapter 3. Section 4.1 addresses longitudinal development of learners' vocabulary in written and oral production modality over time, after the FI and SA learning contexts respectively, and in contrast with NS production, in order to provide an answer to RQ1. In this section, we also include NS baseline data as a yardstick assessing L2 performance across time and perform quantitative between-groups comparisons, as well as a qualitative analysis of NLSs. In section 4.2 we examine learners' vocabulary gains as a result of the FI and SA contexts within and across both types of production modality (oral and written) in response to RQ2. Finally, in section 4.3 we evaluate the role of initial proficiency level on the patterns of vocabulary development and amount of gains in order to address RQ3. We consider initial level in two different ways: as a function of learners' vocabulary knowledge, based on their scores in the domain of lexical fluency, density, diversity, sophistication, and accuracy, and as a function of their general proficiency, based on their scores on the sentence-rephrase and cloze tests.

To guide this chapter, we restate the RQ and its sub-questions in each section and then present descriptive summaries followed by the results of inferential statistics. The chapter concludes with a summary of the findings.

All statistical data analyses were performed using SPSS (version 19). For all tests, unless otherwise indicated, the alpha level was set at .05. Prior to performing each of the statistical analyses, data were assessed for normality and when found to be non-normal, were treated non-parametrically. Decisions that were made in each case and the tests applied are described in the sections to follow.

4.1. RQ1: Vocabulary development in oral and written production

In this section, we address our first research question, which was formulated in Chapter 3 as follows:

RQ1) How does productive vocabulary by advanced EFL learners develop in each modality (writing and speech) longitudinally, after two different learning contexts (FI and SA) and in contrast with NS production?

RQ1a) Does learners' written and oral productive vocabulary improve significantly over time, and after either the FI or SA learning contexts?

RQ1b) Does learners' written and oral productive vocabulary approximate native-like norms at any point over the observation period in terms of quantitative lexical measures?

RQ1c) Does learners' written and oral productive vocabulary approximate native-like norms at any point over the observation period in terms of qualitative native-like selections?

In response to RQ1, we explored our data in a variety of ways. To begin with, we performed one-way repeated measures ANOVAs, with Bonferroni *post-hoc* tests, in order to analyse to what extent learners' vocabulary changed over time through written and oral data respectively. The aim of the analysis was to determine whether any progress occurred after learning English either in the FI or SA learning contexts, when experiencing them subsequently, and thus provide an answer to our first sub-question (RQ1a). These analyses are detailed in sections 4.1.1.

In reply to our second sub-question on cross-sectional quantitative comparisons between learners' and NSs' productive vocabulary (RQ1b), a series of independent t-tests were conducted to examine in what ways they differed across time and whether they converged at any point over the observation period. These results are reported in Section 4.1.2.

Finally, in order to answer our third sub-question (RQ1c) and compare learners' and NSs' productive vocabulary qualitatively, we compiled a list of word combinations in which learners expressed the same (or similar) ideas as NSs, looking at how native-like their lexical

choices were across all three testing times. A detailed account of these NLSs is provided in Section 4.1.3.

4.1.1. Longitudinal and context-specific changes in productive vocabulary

In this section, we address RQ1a examining longitudinal development in learners' written and oral productive vocabulary, over time and after the FI and SA contexts respectively. For practical reasons, we begin by presenting a descriptive summary of the written production, followed by that of the oral production (see Table 9 and Table 10 respectively). Descriptive statistics include learners' mean vocabulary scores and standard deviations on all measures at three testing times, as well as NS baseline data. We proceed to apply statistical tests after the presentation of the descriptive analysis of the data.

A close look at the mean scores in written production (see Table 9) reveals that overall learners' fluency in terms of types, tokens and word class frequencies dropped from T1 to T2 but increased greatly from T2 to T3, moving toward NS values. The length of learners' compositions ranged from 116 to 452 words, with a mean of 232.59 over the course of the study. Turning to the examination of lexical density, we see a slightly different picture, demonstrating little change over time. As for lexical diversity variables (GI and D), these show a non-linear U-shape curve equivalent to that seen for fluency, that is, dropping from T1 to T2, and then increasing again from T2 to T3. As far as measures of lexical sophistication are concerned, here, although

Results

some differences are apparent, they are not striking. While the K2 frequency band shows a linear decrease over time, the Academic wordlist (AWL) slightly increases at T3, in the direction of NS norms. The Greco-Latin cognate index (GLI) remains quite stable. Finally, lexical accuracy, as assessed by E/w and LexE/w, appears to improve longitudinally, with major spelling errors being more frequent after FI, but declining after SA to reach the lowest values. The minor spelling errors, or slips of the pen (MnSp/w), in contrast, appear to increase by the end of the study. It should be noted that given a relatively larger standard deviation at T3 for both types of spelling errors, there seems to be greater variability in learners' scores after the SA context than at the other two data collection points.

Table 9. Descriptive statistics for written productive vocabulary: means obtained at T1, T2 and T3 for learners ($n=30$) and NS ($n=27$) (SD in parentheses).

Written production		Learners			NSs	
		T1	T2	T3		
Fluency	Tokens	219.40 (76.00)	210.63 (45.30)	267.73 (82.43)	253.37 (68.87)	
	Types	109.70 (25.96)	102.90 (18.09)	128.00 (25.92)	127.15 (26.62)	
	CW	88.07 (28.18)	86.40 (21.46)	108.53 (32.59)	111.89 (28.94)	
	Adj	14.67 (5.10)	15.93 (4.95)	18.40 (6.47)	22.44 (7.03)	
	Adv	8.97 (4.83)	8.90 (4.42)	13.07 (5.69)	11.48 (5.16)	
	N	38.73 (14.54)	36.43 (10.93)	42.93 (13.21)	47.48 (14.52)	
	V	25.7 (9.89)	25.13 (6.84)	34.13 (12.16)	30.48 (9.01)	
	FW	117.30 (43.09)	110.30 (24.77)	140.70 (46.88)	123.04 (36.20)	
	Conj	18.07 (6.91)	17.83 (6.13)	22.73 (9.49)	19.89 (5.84)	
	Prep	22.80 (8.46)	19.90 (5.40)	26.67 (8.86)	29.19 (10.10)	
	Pron	21.90 (11.26)	20.83 (6.98)	28.43 (11.87)	17.96 (8.02)	
	Density	CW/w	40.48 (3.77)	40.91 (3.69)	40.72 (3.82)	44.40 (3.58)
		Adj/w	6.94 (2.19)	7.69 (2.06)	6.96 (1.73)	9.04 (2.23)
Adv/w		4.02 (1.61)	4.18 (1.70)	4.98 (1.84)	4.45 (1.55)	
N/w		17.79 (3.04)	17.17 (2.84)	16.19 (2.82)	18.80 (3.02)	
V/w		11.72 (2.13)	11.87 (1.85)	12.59 (2.04)	12.11 (1.88)	
FW/w		53.16 (3.07)	52.32 (3.82)	52.31 (3.45)	48.35 (3.04)	
Conj/w		8.30 (1.62)	8.41 (1.95)	8.29 (1.71)	7.88 (1.28)	
Prep/w		10.51 (2.09)	9.53 (1.93)	10.04 (2.06)	11.40 (1.79)	
Pron/w		9.73 (3.42)	9.94 (2.83)	10.52 (2.68)	7.10 (2.24)	
Diversity		GI	7.44 (0.77)	7.11 (0.82)	7.87 (0.71)	8.01 (0.87)
	D	66.57 (14.46)	62.73 (15.06)	67.75 (10.42)	66.56 (13.74)	

Written production	Learners			NSs	
	T1	T2	T3		
Sophistication	K1	91.30 (2.36)	91.76 (2.93)	91.26 (2.48)	87.23 (3.61)
	K2	4.16 (1.74)	3.99 (1.92)	3.57 (1.54)	4.97 (1.66)
	AWL	3.47 (1.71)	3.36 (1.79)	4.05 (2.00)	4.94 (1.98)
	OffL	1.06 (0.81)	0.88 (0.85)	1.11 (1.09)	2.85 (1.78)
	GLI	15.83 (3.92)	15.77 (3.78)	15.60 (3.18)	20.15 (4.24)
Accuracy	E/w	7.80 (4.01)	5.99 (2.39)	4.61 (2.64)	0.17 (0.23)
	LexE/w	2.28 (1.50)	1.88 (0.85)	1.40 (0.98)	0.07 (0.15)
	MjSp/w	0.76 (0.63)	0.87 (0.79)	0.69 (0.88)	0.34 (0.42)
	MnSp/w	0.06 (0.15)	0.07 (0.19)	0.20 (0.37)	0.12 (0.24)

Regarding oral production, differences in fluency measures across time are less noticeable: types and tokens show a slight decrease over time, with tokens moving towards NS levels. A similar linear decrease is displayed by word class frequencies (i.e., nouns, function words), presumably due to a shorter text length at T2 and T3. Interestingly, this pattern does not hold true for adverbs, whose proportion, in contrast, increases at T3 reaching NS levels. Once controlled for text length, lexical density measures do not appear to change in any systematic way, yet there are some apparent fluctuations in terms of Adv/w, Noun/w and Conj/w. Overall, learners produced an average of 267.46 words during the oral interview, ranging from 91 to 663 words. On the other hand, lexical diversity measures, GI and D, showed a slight non-linear increase over time, whereas the developmental pattern for lexical sophistication was less straightforward with most notable changes occurring after FI. As we can infer from the mean scores at T2, learners approach target-like language in the measures of K1, K2, OffL and GLI, that is producing a lower percentage of basic vocabulary and incorporating a more advanced lexis in their speech. Finally, lexical accuracy appears to mirror the pattern seen for written data, in that the overall percentage of errors gradually decreases over time, as does the number of lexical

Results

errors per word, although learners still remain far less accurate than NS.

Table 10. Descriptive statistics for oral productive vocabulary: means obtained at T1, T2 and T3 for learners ($n=30$) and NS ($n=25$) (SD in parentheses).

Oral production		Learners			NSs
		T1	T2	T3	
Fluency	Tokens	279.20 (110.56)	262.07 (92.33)	261.10 (107.15)	251.96 (118.13)
	Types	112.90 (32.80)	110.13 (27.44)	109.13 (28.64)	119.44 (38.95)
	CW	97.13 (39.38)	91.37 (32.40)	92.77 (38.14)	98.04 (43.88)
	Adj	15.03 (5.99)	13.60 (5.93)	14.33 (6.79)	14.16 (6.03)
	Adv	17.13 (8.35)	17.17 (7.03)	20.57 (9.90)	20.44 (9.09)
	N	34.97 (18.77)	33.63 (13.46)	30.50 (12.85)	37.44 (18.06)
	V	30.00 (12.43)	26.97 (10.72)	27.37 (11.96)	26.00 (13.85)
	FW	153.40 (62.23)	145.30 (52.33)	141.50 (58.76)	129.76 (62.30)
	Conj	28.03 (12.48)	28.43 (9.70)	25.90 (11.55)	19.60 (10.21)
	Prep	18.53 (10.73)	16.77 (9.58)	16.37 (8.03)	20.84 (13.93)
	Pron	44.20 (17.80)	40.47 (14.15)	41.23 (16.00)	32.40 (15.51)
	Density	CW/w	34.74 (3.30)	34.99 (2.48)	35.60 (2.62)
Adj/w		5.59 (1.68)	5.33 (1.89)	5.48 (1.53)	6.06 (1.98)
Adv/w		6.24 (2.40)	6.47 (1.37)	7.84 (1.83)	8.19 (1.80)
N/w		12.12 (2.65)	12.83 (2.08)	11.74 (1.89)	15.10 (2.00)
V/w		10.78 (1.76)	10.36 (1.82)	10.54 (2.01)	10.03 (1.98)
FW/w		54.68 (3.08)	55.24 (2.61)	54.04 (3.10)	51.18 (2.65)
Conj/w		9.99 (1.89)	11.07 (2.19)	9.91 (2.02)	7.81 (1.88)
Prep/w		6.33 (1.57)	6.08 (1.82)	6.10 (1.47)	7.81 (2.35)
Pron/w		15.91 (2.36)	15.54 (1.90)	15.98 (1.76)	12.84 (1.68)
Diversity	GI	6.78 (0.74)	6.83 (0.62)	6.81 (0.59)	7.57 (0.85)
	D	46.94 (8.60)	45.61 (7.38)	48.63 (7.04)	63.26 (9.10)
Sophistication	K1	94.49 (1.87)	93.35 (1.56)	94.10 (1.71)	91.68 (2.02)
	K2	3.28 (1.22)	3.73 (1.21)	3.49 (1.37)	3.89 (1.53)
	AWL	0.98 (0.65)	1.11 (0.82)	1.13 (0.70)	1.94 (1.17)
	OffL	1.25 (0.97)	1.81 (1.01)	1.28 (0.68)	2.49 (0.99)
	GLI	9.43 (2.13)	10.17 (3.01)	9.57 (2.57)	13.92 (2.86)
Accuracy	E/w	6.39 (2.94)	5.51 (3.09)	4.21 (2.38)	0.44 (0.48)
	LexE/w	1.72 (1.00)	1.30 (0.80)	0.86 (0.71)	0.06 (0.17)

After exploring the data descriptively, we went on to test the observed differences statistically. Given that most variables made up reasonably normal distributions³⁹, we performed a series of one-way repeated measures ANOVAs with *Time* (T1, T2, T3) as the within-subjects

³⁹ The variables that did not meet the assumption of normality either at one or more data collection times were submitted to Friedman non-parametric tests and are marked with (*) in Table 11.

factor to test the significance of longitudinal changes for each measure. We further applied Bonferroni-adjusted pairwise comparisons to determine whether these changes occurred after either formal instruction (FI, between T1 and T2) or study abroad (SA, between T2 and T3). A results summary can be seen in Table 11.

Table 11. RM-ANOVAs: Longitudinal changes with context-specific (FI or SA) differences in written and oral productive vocabulary (all Fs = [2, 28]).

	Written production					Oral production				
	F	<i>p</i>	η^2p	FI	SA	F	<i>p</i>	η^2p	FI	SA
Tokens*	8.359	.001	.37	1.00	.001	.462	.635	.03	--	--
Types	12.616	.000	.47	.568	.000	.219	.805	.02	--	--
CW	8.196	.002	.37	1.00	.001	.395	.677	.03	--	--
Adj	3.389	.048	.20	.962	.226	.863	.433	.06	--	--
Adv*	8.720	.001	.38	1.00	.002	2.119	.139	.13	--	--
N	4.049	.029	.22	1.00	.021	1.330	.281	.09	--	--
V	10.093	.001	.42	1.00	.000	1.398	.264	.09	--	--
FW*	7.112	.003	.34	1.00	.002	.441	.648	.03	--	--
Conj*	4.525	.020	.24	1.00	.031	.991	.384	.07	--	--
Prep*	9.259	.001	.40	.277	.000	.584	.564	.04	--	--
Pron	7.576	.002	.35	1.00	.003	.826	.448	.06	--	--
CW/w*	.145	.866	.01	--	--	1.161	.328	.08	--	--
Adj/w	2.040	.149	.13	--	--	.211	.811	.02	--	--
Adv/w	4.059	.028	.23	1.00	.175	7.615	.002	.35	1.00	.002
N/w*	4.262	.024	.23	.789	.124	2.174	.133	.07	--	--
V/w	1.724	.197	.11	--	--	.732	.490	.05	--	--
FW/w	1.104	.345	.07	--	--	1.318	.284	.09	--	--
Conj/w	.056	.946	.00	--	--	6.041	.007	.30	.008	.033
Prep/w	3.849	.033	.22	.026	.904	.363	.699	.03	--	--
Pron/w	1.048	.364	.07	--	--	.612	.549	.04	--	--
GI	14.583	.000	.51	.062	.000	.166	.848	.01	--	--
D	2.478	.102	.15	--	--	1.795	.185	.11	--	--
K1	.510	.606	.04	--	--	4.090	.028	.23	.034	.131
K2	1.376	.269	.09	--	--	1.345	.277	.09	--	--
AWL	2.137	.137	.13	--	--	.626	.542	.04	--	--
OffL*	1.141	.334	.08	--	--	3.880	.033	.22	.050	.061
GLI*	.076	.927	.01	--	--	.970	.392	.07	--	--
E/w*	13.052	.000	.48	.014	.007	15.684	.000	.53	.360	.052
LexE/w*	4.819	.016	.26	.390	.096	10.482	.000	.43	.052	.056
MjSp/w*	.367	.696	.03	--	--	N/A	N/A	N/A	N/A	N/A

*Variables submitted to non-parametric techniques (Friedman test).

Shaded cells indicate significance.

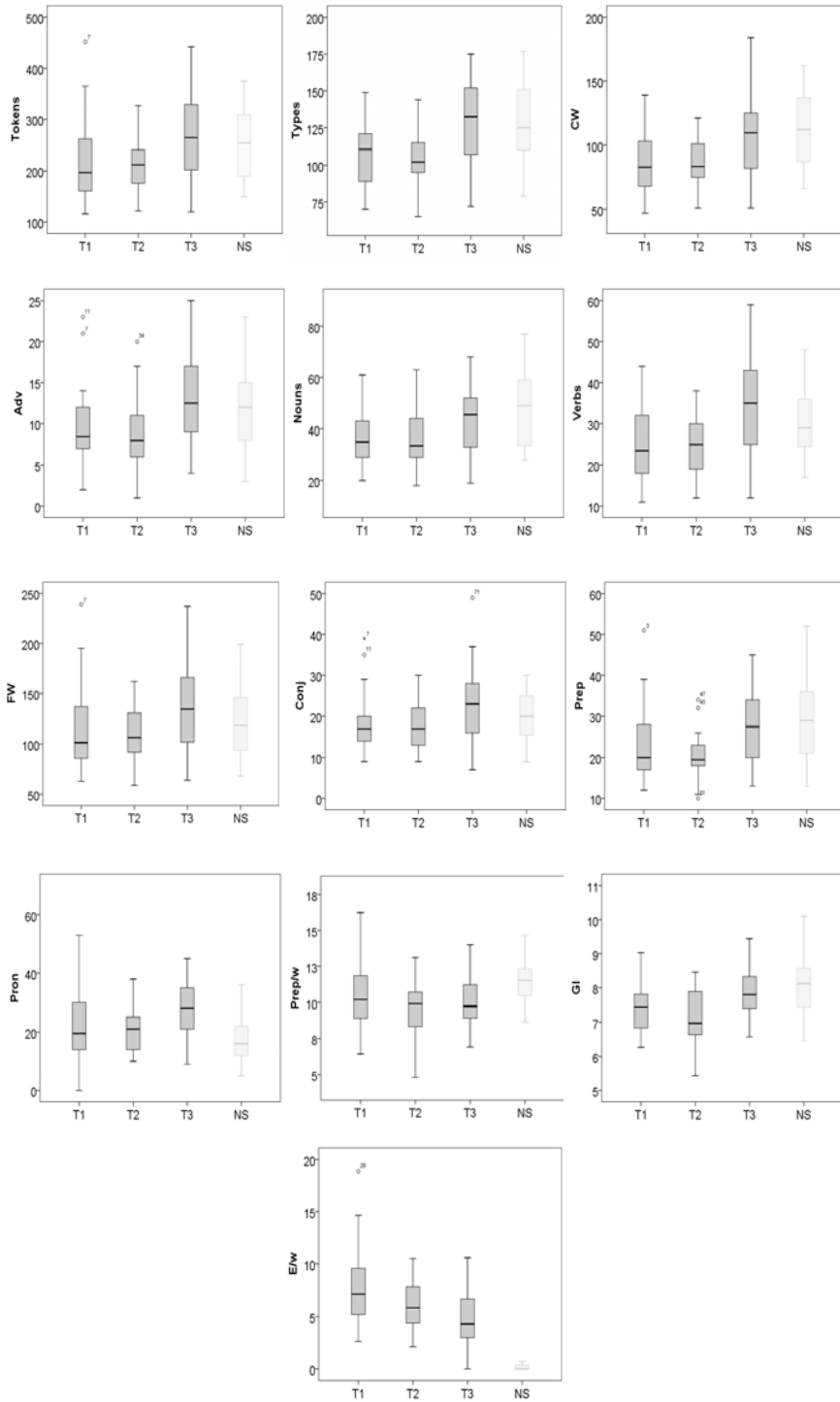
As can be seen in Table 11, within written production, the results of the main ANOVAs revealed a significant main effect of time on all

Results

measures of fluency, adverb, noun and preposition density, lexical diversity, as measured by GI, and accuracy (E/w, LexE/w). The largest main effect was seen for lexical diversity (GI), with partial eta squared (η^2p) of .51 and lexical accuracy (E/w), with η^2p of .48. We did not find statistically significant differences for any of the variables associated with lexical sophistication. Follow-up Bonferroni pairwise comparisons revealed that all measures (with the exception of Prep/w and E/w) were only significant between T2 and T3, suggesting that improvement occurred primarily as a result of the SA experience. Accuracy, as assessed by E/w, showed a consistently linear decrease in both contexts, as we suspected from the descriptive statistics of accuracy scores, while preposition density decreased significantly after FI. Boxplots of significant context-specific changes in written production are provided in Figure 5.

As noted earlier, due to violation of normality in some variables, we applied non-parametric Friedman tests to ascertain our results. These confirmed the previous findings at the same level of significance and are reported as follows: Tokens [$\chi^2(2)=41.57, p=.003$]; Adv [$\chi^2(2)=14.33, p=.001$]; FW [$\chi^2(2)=9.90, p=.007$]; Conj [$\chi^2(2)=6.44, p=.040$]; Prep [$\chi^2(2)=13.98, p=.001$]; OffL [$\chi^2(2)=0.12, p=.941$]; E/w [$\chi^2(2)=19.47, p=.000$]; SpE/w [$\chi^2(2)=1.14, p=.565$]. It should be noted that one of the spelling variables (MnSp/w) was left out of the analysis, as the shape of the distribution showed severe positive skewness and high kurtosis values, which was also confirmed by checking the normal Q-Q plots and K-S tests of normality. MnSp/w is henceforth only evaluated descriptively.

Figure 5. Boxplots of significant context-specific changes by group in written productive vocabulary.

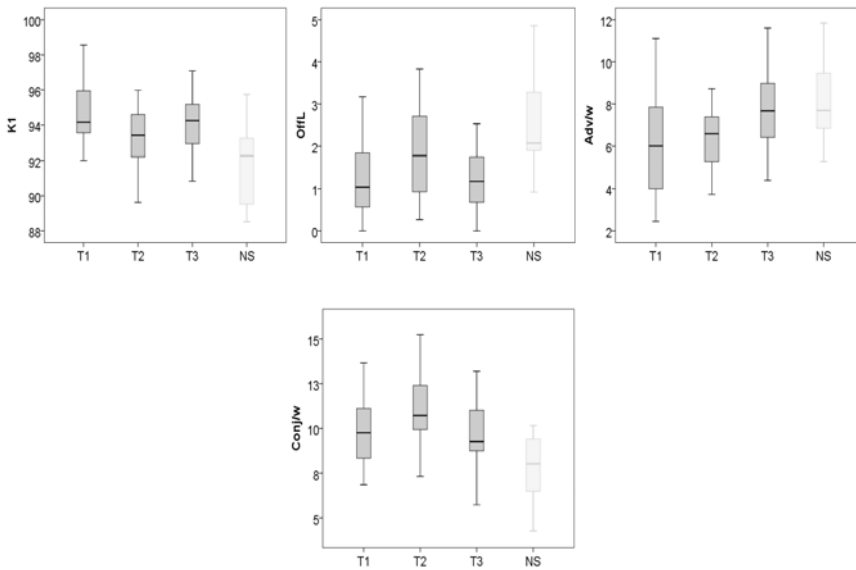


Results

Moving to the results obtained in oral production, the main ANOVAs showed a less prominent effect of time and significance was found in adverb and conjunction density, lexical sophistication, as measured by K1 and OffL, and accuracy (E/w, LexE/w). The largest effect size was observed for accuracy $\eta^2 p=.53$, just as in written production. We did not find statistically significant differences in lexical fluency and diversity, unlike in writing, nor did we find any significant change in sophistication measures of K2, AWL and GLI. Subsequent pairwise comparisons showed that while significant changes in adverb density occurred after the SA, fluctuations in conjunction density were significant in both contexts. In contrast, the change in K1 and OffL bands reached significance after the FI period. The full list of the OffL words used by the learners at T2 can be consulted in Appendix 9 and is covered in greater depth in the discussion section 5.1.1. Finally, with regard to accuracy, measures improved over time, although not in favour of any specific learning context. This indicated that learners gradually reduced the number of errors in speech, a trend that was also true for writing.

Again, for the non-normally distributed variables pertaining to oral data, Friedman tests confirmed the same type of differences (i.e., significant or non-significant) as follows: Tokens [$\chi^2(2)=0.87, p=.648$]; Adv [$\chi^2(2)=2.58, p=.275$]; Noun [$\chi^2(2)=0.90, p=.636$]; FW [$\chi^2(2)=1.06, p=.589$]; Prep [$\chi^2(2)=0.62, p=.733$]; CW/w [$\chi^2(2)=2.47, p=.291$]; GLI [$\chi^2(2)=0.41, p=.814$]; LexE/w [$\chi^2(2)=15.95, p=.000$]. Boxplots illustrating significantly context-sensitive changes in oral production can be seen in Figure 6.

Figure 6. Boxplots of significant context-specific changes by group in oral productive vocabulary.



Thus far, we have looked at learners' vocabulary development in each modality (written and oral production, respectively) over time, and after each of the learning context (FI and SA). The results obtained from descriptive and inferential statistics showed that EFL productive vocabulary in writing and speech develops in somewhat different ways. In written production, we found a significant main effect of time on all fluency measures, three measures of density (Adv/w, N/w, Prep/w), Guiraud's index of lexical diversity and two measures of lexical accuracy (E/w, LexW/w). Results also revealed that improvement in all measures (except for Prep/w and E/w) was significant after the SA period, but not after classroom FI. No significant changes were found in the domain of lexical sophistication. The results for oral production, in contrast, presented a rather different picture. While there were significant changes over time for adverb and conjunction density (Adv/w, Conj/w), two measures of

lexical sophistication (K1, OffL) and accuracy (E/w, LexE/w), only one measure – adverb density (Adv/w) – showed a significant improvement after SA but not after FI. Conversely, the progress made in lexical sophistication occurred as a result of classroom FI, with learners using a significantly larger proportion of more advanced or less frequent words and relying less on basic vocabulary. We did not find any significant changes in the domain of lexical diversity.

4.1.2. Quantitative comparisons with native speakers

Having explored learners' developmental progress in written and oral productive vocabulary respectively over time, and as a consequence of following the FI and SA periods, we conducted further quantitative analyses to complete the response to RQ1b and compare our learners' performance for each modality and lexical domain with that of native speakers at all three data collection times. For that purpose, we carried out independent samples *t*-tests, and in cases where the variables did not meet the assumption of normality, the distribution-free Mann-Whitney U-tests were used (See descriptive statistics reported above in Table 9 and Table 10). The results for independent *t*-tests in written production are summarised in Table 12, and in oral production in Table 13.

In written production, results of *t*-tests revealed that despite scoring below NSs on average, learners improved over time and converged with NSs by the end of the study (see Table 12). More specifically, significant differences were found in most measures of fluency (Types, CW, Adj, Adv, N, Prep) between learners (at T1 and T2) and NS,

which converged in all measures (except for Adj and Pron) at T3, suggesting that the SA context was particularly beneficial for fluency, as opposed to FI. The same was true for lexical diversity, when this was measured through GI. On the other hand, interestingly, this was not the case for D, which did not differ substantially at any testing time. As for lexical sophistication, learners' means for K1, OffL, and GLI diverged significantly from NS values across all three testing times, with AWL coalescing with NS productions at T3 and no longer registering significant differences. Finally, and as expected, learners remained significantly less accurate than NS over the course of the study, yet converged with NS on spelling errors at T3, suggesting that at least one component of the word knowledge like knowing the written form of words was being mastered in more appropriate, target-like ways after SA.

For the variables that did not have a normal distribution, Mann-Whitney U-tests ascertained previous results at the same level of significance: Tokens T1-NS [$z=-1.93$, $p=.054$], Prep T1-NS [$z=-2.66$, $p=.008$], FW/w T1-NS [$z=-4.76$, $p=.000$]; OffL T3-NS [$z=-4.36$, $p=.000$], Adv T3-NS [$z=-.83$, $p=.404$], FW/w [$z=-4.06$, $p=.000$], SpE/w T3-NS [$z=-1.60$, $p=.109$], with the only exception pertaining to Adv T1-NS [$z=-2.12$, $p=.034$], which yielded a non-significant result through parametric statistics, yet with a borderline p value (.063) (see Table 12).

Results

Table 12. Independent-samples *t*-tests comparing written productive vocabulary of learners (at T1, T2 and T3) and NSs.

Written production	T1 vs. NS			T2 vs. NS			T3 vs. NS		
	<i>t</i>	<i>df</i>	<i>p</i>	<i>t</i>	<i>df</i>	<i>p</i>	<i>t</i>	<i>df</i>	<i>p</i>
Tokens	-1.761	55	.084*	-2.736	44.2	.009	.709	55	.481
Types	-2.504	55	.015	-3.978	45.1	.000	.122	55	.903
CW	-3.147	55	.003	-3.744	47.6	.000	-.409	55	.684
Adj	-4.733	47	.000	-4.001	46.1	.000	-2.261	55	.028
Adv	-1.900	55	.063**	-2.034	55	.047	1.098	55	.277*
N	-2.269	55	.027	-3.265	55	.002	-1.238	55	.221
V	-1.901	55	.028	-2.540	55	.014	1.277	55	.207
FW	-.541	55	.591	-1.563	55	.124	1.579	55	.120
Conj	-1.069	55	.290*	-1.293	55	.202	1.378	48.9	.175
Prep	-2.596	55	.012*	-4.261	38.8	.000	-1.003	55	.320
Pron	1.505	55	.138	1.444	55	.154	3.857	55	.000
CW/w	-4.008	55	.000	-3.614	55	.001	-3.734	55	.000
Adj/w	-3.561	55	.001	-2.369	55	.021	-3.943	55	.000
Adv/w	-1.020	55	.312	-.621	55	.537	1.160	55	.251
N/w	-2.447	55	.014	-2.099	55	.040	-3.372	55	.001
V/w	-.714	55	.478	-.477	55	.635	.930	55	.357
FW/w	5.929	55	.000*	4.304	55	.000	4.577	55	.000
Conj/w	1.078	55	.286	1.211	55	.231	1.032	55	.307
Prep/w	-1.730	55	.089	-3.790	55	.000	-2.659	55	.010
Pron/w	3.404	55	.001	4.181	55	.000	5.201	55	.000
GI	-2.587	55	.012	-4.011	55	.000	-.637	55	.527
D	.001	55	.999	1.000	55	.322	.370	55	.713
K1	4.974	44	.000	5.224	55	.000	4.947	55	.000
K2	-1.789	55	.078	-2.047	55	.045	.412	55	.002
AWL	-3.013	55	.004	-3.159	55	.003	-1.684	55	.098
OffL	-4.800	35.6	.000	-5.257	36.4	.000	-4.502	55	.000*
GLI	-3.991	55	.000	-4.127	55	.000	-4.610	55	.000
E/w	10.395	29.2	.000*	13.250	30	.000*	9.179	29.5	.000*
LexE/w	8.047	29.7	.000*	11.499	31.1	.000*	7.371	30.6	.000*
MjSp/w	2.911	55	.005	3.213	45.3	.002	1.961	42.6	.056*

*Variables submitted to non-parametric techniques (Mann-Whitney U Test).

**Between-groups differences in the non-normally distributed adverb (Adv) variable did not reach significance, as revealed by Mann-Whitney U Test.

Shaded cells indicate significance.

In oral production (see Table 13), contrary to what we have observed for writing, *t*-test results revealed little change in learner performance towards native-like standards across the contexts. Overall, learners and NSs differed significantly on the majority of the measures at all testing times, except for fluency where learners consistently converged with NSs (from T1 to T3) on types, tokens and word class counts, with the

exception of conjunctions and pronouns. Furthermore, significant differences were revealed between learners and NSs in all measures of lexical diversity, accuracy and sophistication at all times, except in K2, which remained unchanged throughout the study. Significant Learners/NSs divergence was also found for measures of lexical density, with Adj/w and V/w being the only measures that did not distinguish learners from NSs. The only growth toward NS usage was observed for adverb density (Adv/w), which approached NS values at T3 to the degree that it was no longer statistically distinguishable from NS performance.

The non-normally distributed variables in oral data when submitted to a Mann-Whitney analysis revealed the same type (significant or non-significant) of difference: Tokens T1-NS [$z=-0.91$, $p=.361$], GLI T1-NS [$z=-5.24$, $p=.000$], Prep T1-NS [$z=-0.72$, $p=.472$], FW T1-NS [$z=-1.49$, $p=.137$]. Tokens T2-NS [$z=-0.70$, $p=.483$], Prep T2-NS [$z=-1.36$, $p=.173$], Noun/w T2-NS [$z=-3.83$, $p=.000$]. Prep T3-NS [$z=-1.38$, $p=.168$], Adv T3-NS [$z=-0.24$, $p=.813$], CW/w T3-NS [$z=-4.60$, $p=.000$].

Table 13. Independent-samples *t*-tests: comparing oral productive vocabulary of learners (at T1, T2 and T3) and NSs.

Oral production	T1 vs. NS			T2 vs. NS			T3 vs. NS		
	<i>t</i>	<i>df</i>	<i>p</i>	<i>t</i>	<i>df</i>	<i>p</i>	<i>t</i>	<i>df</i>	<i>p</i>
Tokens	.882	53	.382*	.356	53	.723*	.301	53	.765
Types	-.676	53	.502	-1.037	53	.305	-1.129	53	.264
CW	-.081	53	.936	-.648	53	.520	-.477	53	.635
Adj	.537	53	.594	-.346	53	.731	.099	53	.921
Adv	-1.404	53	.166	-1.505	53	.138	.049	53	.961*
N	-.495	53	.623	-.895	53	.375	-1.661	53	.103
V	1.128	53	.264	.292	53	.772	.393	53	.696
FW	1.402	53	.167*	1.006	53	.319	.718	53	.476
Conj	2.707	53	.009	3.285	53	.002	2.123	53	.038
Prep	-.693	53	.491*	-1.280	53	.206*	-1.488	53	.143*
Pron	2.593	53	.012	2.015	53	.049	2.067	53	.044

Results

Oral production	T1 vs. NS			T2 vs. NS			T3 vs. NS		
	<i>t</i>	<i>df</i>	<i>p</i>	<i>t</i>	<i>df</i>	<i>p</i>	<i>t</i>	<i>df</i>	<i>p</i>
CW/w	-5.619	53	.000	-6.279	53	.000	-5.247	53	.000*
Adj/w	-.953	53	.345	-1.407	53	.165	-1.234	53	.223
Adv/w	-3.351	53	.001	-4.004	53	.000	-.707	53	.483
N/w	-4.618	53	.000	-4.101	53	.000*	-6.383	53	.000
V/w	1.503	53	.139	.645	53	.521	.954	53	.344
FW/w	4.463	53	.000	5.711	53	.000	3.645	53	.001
Conj/w	4.276	53	.000	5.856	53	.000	3.955	53	.000
Prep/w	-2.783	53	.007	-3.070	53	.003	-3.278	53	.002
Pron/w	5.609	51.9	.000	5.519	53	.000	6.714	53	.000
GI	-3.709	53	.000	-3.729	53	.000	-3.913	53	.000
D	-6.824	53	.000	-7.942	53	.000	-6.720	53	.000
K1	5.358	53	.000	3.462	53	.001	4.820	53	.000
K2	-1.661	53	.103	-.451	53	.654	-1.040	53	.303
AWL	-3.664	36.1	.001	-3.000	41.7	.005	-3.026	37.8	.004
OffL	-4.668	53	.000	-2.472	53	.017	-5.183	41.2	.000
GLI	-6.668	53	.000*	-4.715	53	.000	-5.948	53	.000
E/w	10.934	30.9	.000	8.848	30.7	.000	8.467	31.9	.000
LexE/w*	8.918	31	.000*	8.307	32.2	.000	5.961	33	.000*

*Variables submitted to non-parametric techniques (Mann-Whitney U Test).

Shaded cells indicate significance.

To summarise our findings as to the quantitative comparisons between learner and NS productive vocabulary across time and for each modality, we observe the following picture. In written production, learners converged with NSs specifically after SA, no longer being statistically distinguishable in the domain of fluency (except for Adj and Pron), lexical diversity index of Guiraud (GI), lexical sophistication, as seen through the academic word list (AWL), and lexical accuracy in terms of major spelling errors (MjSp/w). In oral production, in contrast, the only measure that moved toward NS rates from T1 to T3 was adverb density (Adv/w). The learners converged with the native speakers on the adverb density measure only after SA.

4.1.3. Qualitative comparisons with native speakers

In order to provide an answer to the sub-question on qualitative comparisons of learner and native performance (RQ1c), we explored learners' native-like selections (NLSs) in written and oral productions qualitatively. That is, we compared learners' word choices at each testing time with those of NSs when describing the same thing. To get a more presentable and manageable overview of the analysis, we distinguished between gradual approximation to NS usage (longitudinal NLSs), context-specific approximation in FI and SA, respectively (context-specific NLSs), and lack of changes at any point over the observation period (lack of changes). For the sake of consistency with previous results, we present the analysis of written production followed by that of the oral, detailing the most representative examples for each case.

Longitudinal NLSs (written production):

Looking at the lexical choices in learners' essays over time, one of the most visible changes that can be observed is a gradual increase in the use of impersonal forms. As shown in Table 14, learners make more use of the *it*-cleft construction, mostly based on '*it* + copula "*be*" + adjective' pattern at T2 and T3 than at T1, suggesting that both contexts seem to be equally beneficial for the acquisition of generality, a feature that is typical of academic writing and also highly frequent in the NS corpus. Likewise, by the end of the study learners rely more on the pronoun *one* – the most common substitute for personal pronouns (I,

you, he) in general statements – thus aligning more closely with the NSs.

Table 14. Frequency of use of impersonal forms in written production.

Written production	Number of uses				
	Impersonal forms	T1	T2	T3	NSs
It + copula “be” + adj. (important, essential, vital, etc.)		23	34	35	52
Use of the pronoun “one”		3	4	8	48

Although in both cases (especially with the reference to the pronoun *one*) NSs make a substantially wider use of impersonal language, overall learners’ essays give the impression of being more formal by the end of the observation period. Table 15 illustrates these examples.

Table 15. Excerpts from learners’ and NSs’ written compositions illustrating the use of impersonal forms.

T1	T2	T3	NSs
it is important not to forget your own customs...	it is very important to get used to their customs...	it is impossible to forget one’s culture...	it is necessary to adapt to the customs...
one should be free to live the way...	one cannot forget his origins...	one has to know the new culture...	one should make an effort to keep an open mind...

Context-specific NLSs (written production):

As for the approximation to NS usage specifically after either of the two learning periods (FI and SA), we noticed the following patterns. In FI, we did not find any remarkable examples. It appears that learners made a more frequent use of the phrasal verb *hold on to* at T2 than at T1 or T3, although the number of incorrect uses, either

producing the base verb *hold* without the required preposition/particle (e.g., *hold their customs and way of life*) or selecting a wrong one (e.g., *hold with our own customs*, *holding to your habits*), was equally high at all times (see Table 16 for the frequency of correct and incorrect uses). Taking into account that the right form of the phrasal verb was actually used in the prompt, we cannot infer that a more accurate usage was due to learners' better knowledge of the compound as a result of FI in the classroom. A higher number of the correct form at T2 might also be related to learners being more concentrated on the task and paying more attention to the wording of the composition topic.

Table 16. Frequency of correct and incorrect uses of *hold on to* in written production.

Written production	Number of uses				
	Phrasal verb <i>hold on to</i>	T1	T2	T3	NSs
Correct usage		1	7	4	13
Incorrect usage (due to 1 or 2)*:		9	11	11	0

*1) incomplete form (*hold* or *hold on*); 2) in combination with other prepositions (*hold with*, *hold to*, etc.)

Regarding SA, the most noteworthy NLSs were found in the use of adverbs modifying adjectives. At T3, learners incorporated a higher number of adverbs into their lexicon, used to intensify or moderate adjectives (e.g., *really difficult*, *entirely different*), and thus equalling NS usage (see Table 17).

Table 17. Frequency of use of adverbs modifying adjectives in written production.

Written production	Number of uses				
	Emphasis	T1	T2	T3	NSs
Adv. (-ly) + Adj. (<i>really difficult</i> , <i>entirely different</i>)		8	6	20	20

As shown in Table 17 and 18, not only did learners rely more on modifying adverbs after SA, the range of the adverbs used was also more extensive (no longer limited to the forms *really* and *completely*, but involving other variants like *equally*, *extremely*, or *strongly*, and thus resembling NS patterns).

Table 18. Examples of adverbs modifying adjectives in learner and NS written compositions.

T1	T2	T3	NSs
really difficult	really substantial	really hard	entirely different
really interested	really different	completely different	seemingly disparate
really interested	really significant	exactly the same	undoubtedly difficult
really important	really difficult	equally valid	vaguely familiar
really important	completely adapted	extremely attached	perfectly natural
highly important	completely changed	strongly important	extremely profitable

Considering that the choice of intensifier combinations is generally limited by collocational restrictions, it seems that SA fosters better knowledge of collocations for expressive purposes in written modality.

Lack of changes (written production):

Despite the longitudinal and context-specific approximations to NS selections described above, there is evidence that learners tend to overuse basic vocabulary, as opposed to NSs. To illustrate, learners are more likely to select high frequency topic-related verbs like *lose* and *forget* (with 28 uses even at T3), unlike NS, whose choices are full of more demanding low-frequency variants (e.g., *discard*, *disregard*, *forego*).

Table 19 summarises these selections.

Table 19. Frequency of topic-related verbs in written production I.

Written production	Number of uses			
	Topic-related verbs I	T1	T2	T3
lose (identity, culture, customs, roots)	6	9	8	4
forget (culture, customs, roots)	16	7	20	6
give up (culture, customs, beliefs)	0	0	1	5
abandon (culture, customs, values)	2	3	1	3
refuse (culture, customs, past)	0	1	3	0
reject (customs, country)	0	0	1	1
leave (customs, culture) (behind)	6	3	4	2
leave (customs) apart	1	1	0	0
quit (way of life)	1	0	0	0
drop, discard, disregard, forego, surrender, take away,	0	0	0	8
terminate, negate (customs, roots)				

Similarly, when describing the process of integration within the new receiving community, learners show a clear preference for the verb *accept*, while most NSs make use of a large set of synonyms to enrich their writing (e.g., *embrace*, *immerse*, *adjust*, *assimilate*, *adhere* and *absorb*), as seen in Table 20.

Table 20. Frequency of topic-related verbs in written production II.

Written production	Number of uses			
	Topic-related verbs II	T1	T2	T3
embrace (customs, culture)	0	1	0	6
accept (customs, traditions)	4	1	7	0
immerse in (the new culture)	1	0	0	6
follow (some of the local customs)	0	2	1	2
settle in to (a different way of life)	0	0	1	2
adjust to (local customs)	0	0	0	5
assimilate into (a new country)	0	0	0	2
adhere to (new customs)	0	0	0	2
absorb (the new culture)	0	0	0	2
accommodate (all aspects of one's customs)	0	0	0	1

Another example where learners systematically differ from NSs has to do with L1 transfer. For instance, when paraphrasing the idea of going abroad, most learners feel inclined to use the determiner *another*,

producing word strings like *move to another country*. NSs, in contrast, prefer the adjective *new* over *another* and are most likely to write *move to a new country*. The explanation for these differences may lie in learners' literally translating from Catalan/Spanish *anar a un altre país/ir a otro país*, where using *new* instead of *another*, while not ungrammatical, is quite clumsy. As we can see in Table 21, the use of *another* becomes less frequent at T3, while for the NS group it is practically non-existent (with 3 uses only).

Table 21. Frequency of use of *another* and *new* as examples of L1 transfer in written production.

Written production	Number of uses				
	Another vs. New	T1	T2	T3	NSs
move to another country ⁴⁰		16	22	15	3
move to a new country		4	3	4	18

We now proceed to describe learners' NLSs pertaining to oral production.

Longitudinal NLSs (oral production):

Turning to the analysis of NLSs in oral production, we observe that learners gradually approach NS usage as they use fewer words and collocations resulting from literal translation from L1, or involving deceptive cognates (false friends). Examples of such selections are displayed in Table 22 below.

⁴⁰Other verbs (e.g., *immigrate, emigrate, live, study, stay*) collocating with *to/in another/a new country* were also found in the data and, therefore, are included on the list.

Table 22. Frequency of use of deceptive cognates and literal translation from L1 in oral production.

Oral production	Number of uses				
	Cognates/L1 Transfer	T1	T2	T3	NSs
career (instead of degree)		10	5	0	0
degree		9	6	8	2
programme		0	1	0	2
to lunch/do the lunch (instead of have lunch)		5	1	0	0
have lunch		7	6	15	0
pack/bring (in) one's own food, lunch, sandwich		1	0	0	11

In Table 22, we can see that learners progressively decrease their use of the false friend *career* meaning *university degree* (*carrera* in Spanish/Catalan), replacing it with true equivalents (e.g., *degree*, *programme*). In a similar vein, learners collocate the noun *lunch* with *have* instead of *do*, gaining in accuracy especially at T3 (after SA). It is interesting to note that NS selections are more narrowly defined, as they resort to the verbs *pack* or *bring in* specifying the type of food they usually eat during their lunch break.

Context-specific NLSs (oral production):

As for the movement toward NS selections specifically after FI, we did not find any noteworthy instances where learners at T2 framed their speech in more native-like ways. The SA context, conversely, enabled learners to acquire greater idiomaticity in terms of lexicalised phrases functioning as pause fillers or through emphasis expressed by the use of adverbs, adjusting learner speech to NS patterns.

For instance, as can be seen in Table 23, learners select considerably more idiomatic fillers such as *I'd say* or *I mean* at T3 than at T1 or T2,

Results

approaching NS usage. Also, there seems to be a tendency at T3 for some learners to make more use of adverbs of emphasis (*really*, *quite*, *particularly*), collocating them with *like* and *enjoy*, rather than simply producing *I like*, and thus moving in the direction of NS choices.

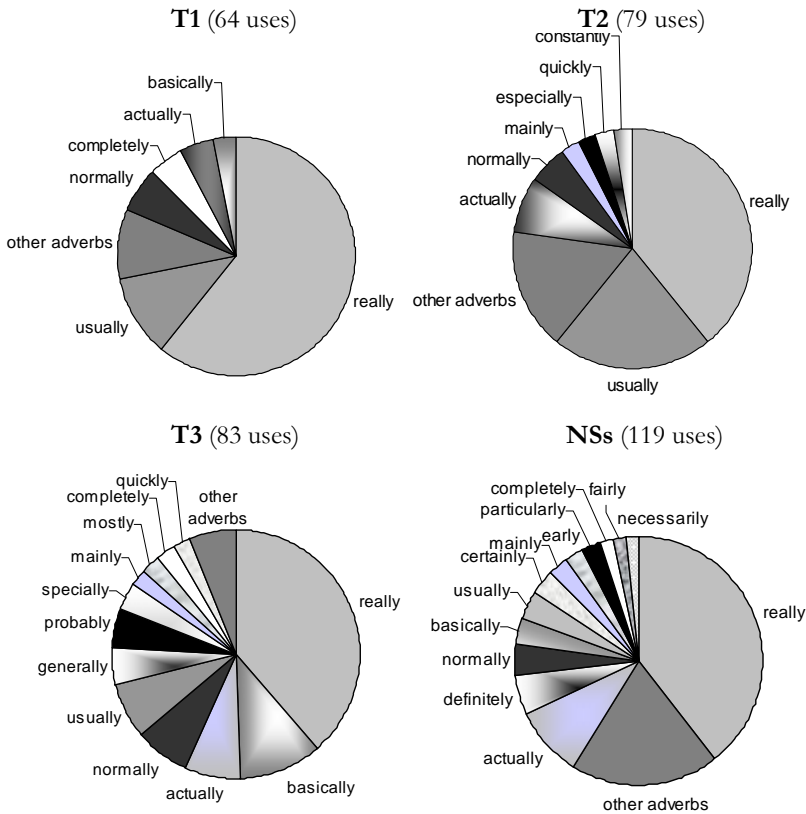
Table 23. Frequency of use of emphasis in oral production.

Oral production	Number of uses				
	Pause fillers and Emphasis	T1	T2	T3	NSs
I would say/I'd say		1	1	6	12
I mean		1	2	14	6
I really (quite, particularly) like/enjoy (X class)		3	1	6	14
I like (X class)		54	50	48	22

In view of our previous statistical results of quantitative comparisons between learners and NSs (see Section 4.1.2.), pointing to the adverb density as the only measure in oral production that improves significantly after SA and converges with NS values, we examined the type of adverbs used in more detail. For reasons of scope and space, we limited our search to all adverbs ending in *-ly*. Figure 7 presents the proportion of the different adverbs contained in the oral data (see next page).

As shown in the charts, learners exhibit an increasing range of adverbs at T3, which in turn resembles more closely NS patterns. The total number of adverbs ending in *-ly* also approaches NS values at T3 (with 83 uses), pointing to SA as the context that to some degree induced these changes.

Figure 7. Circle charts denoting the proportion of different adverbs used in oral production. (Total number of adverbs in parentheses).*



Lack of changes (oral production):

Finally, as in written data, we also found evidence that some learners' lexical choices were resistant to change at all three testing times. Compared to NS use, for instance, learners were likely to select the verbs *work* and *study*, while NSs combined the verb *do* with the nouns *work* and *study*, producing delexicalised phrasings such as *do some work*

* Other adverbs refer to the adverbs that were used only once in each group. They were mostly low frequency adverbs (e.g., *deeply, slowly, individually, firstly, academically, hopefully, barely*, etc.).

or *do a lot of study*. NSs were also more likely to select more specific vocabulary (e.g., producing chunks like *writing papers, assignment, projects, readings*, etc. when describing their study habits). Learners' lexical choices, on the contrary, lacked specificity and remained somewhat vague across time (e.g., the collocation *do homework* was overused after both FI and SA, as shown in Table 24).

Table 24. Frequency of use of delexicalised and specific verb combinations in oral production.

Oral production	Number of uses				
	Delexicalised verbs	T1	T2	T3	NSs
do (a lot of, more...) work		4	5	6	8
do a lot of study(-ing)		0	0	0	4
work (more, harder, at home...)		7	12	12	2
study (more, a lot...)		10	15	11	7
	Specificity				
do homework		2	9	8	3
do/have/write papers, assignments, projects, readings...		2	1	1	8

In summing up our look at the qualitative comparisons between learners and NSs in terms NLSs for each production mode, and as a function of the FI or SA (or both) learning contexts, we can offer a few observations. In written production, learners gradually approached NSs in their use of impersonal forms and gained greater generality, which is a distinctive characteristic of academic writing, after both FI and SA learning contexts. For its part, the SA period was especially beneficial for the development of intensifiers, as seen through the use of adverbs modifying adjectives in alignment with NS norms. On general grounds, however, learners did not expand their lexical repertoire to native-like standards and continued to select more frequent and therefore less specific vocabulary. With regard to oral production, learners made more accurate word choices and relied less

on literal translations from L1 after both FI and SA learning contexts. They also showed greater idiomaticity in terms of lexicalised fillers and adverb diversity, especially after SA. Despite these changes, learners avoided the use of collocations with delexicalised verbs and overused more general lexical items, unlike NSs.

4.2. RQ2: Vocabulary gains across contexts and modalities

This section explores to what extent gains in written and oral productive vocabulary differ from each other as a function of learning context and production modality and tackles RQ2, which was previously stated in Chapter 3 as follows.

RQ2) In which context (FI vs. SA) and modality (writing vs. speech) do learners accrue greater gains?

RQ2a) Do gains, if any, in writing and speech, respectively, accrue to a larger extent in one context than another, when comparing FI and SA?

RQ2b) Do gains, if any, in FI and SA, respectively, accrue to a larger extent in one modality than another, when comparing writing and speech?

Here, for clarity's sake, we address the first sub-question (RQ2a) in Section 4.3.1., and the second sub-question (RQ2b) in section 4.3.2.

4.2.1. Results for vocabulary gains across contexts

In order to analyse possible differences in the amount of gains between learning contexts (FI vs. SA) in written production on the

Results

one hand and in oral production on the other in response to RQ2a, we set up the following analyses. First, we calculated the amount of gains obtained during FI, *FI gains*, by subtracting T1 scores from T2 scores ($T2 - T1$), and gains experienced during SA, *SA gains*, by subtracting T3 scores from T2 scores ($T3 - T2$). We then ran a series of paired-samples *t*-tests to assess the resulting gains within each modality statistically. Gains for all lexical measures were found to follow a normal distribution according to K-S tests of normality, except for the conjunction frequency (Conj.) in oral data, which was additionally assessed through the non-parametric Wilcoxon Signed Rank tests. We considered each production modality independently. Results are displayed for both modalities in Table 25.

Table 25. Paired-samples *t*-tests comparing FI vs. SA mean gains in written and oral production.

	Written production					Oral production				
	FI gains	SA gains	<i>t</i> (29)	<i>p</i>	<i>r</i>	FI gains	SA gains	<i>t</i> (29)	<i>p</i>	<i>r</i>
Tokens	-8.77 (79.30)	+57.10 (75.53)	-2.956	.006	.48	-17.13 (96.66)	-.97 (79.41)	-.691	.495	.13
Types	+6.80 (27.71)	+25.10 (26.91)	-3.966	.000	.59	-2.77 (25.42)	-1.00 (22.00)	-.274	.786	.05
CW	-1.67 (31.26)	+22.13 (30.15)	-2.645	.013	.44	-5.77 (35.26)	1.40 (28.86)	-.802	.429	.15
Adj	+1.27 (6.87)	+2.47 (7.32)	-.554	.584	.10	-1.43 (6.07)	+7.73 (6.20)	-1.224	.231	.22
Adv	-.07 (5.20)	+4.17 (5.87)	-2.487	.019	.42	+0.3 (8.50)	3.40 (8.89)	-1.506	.143	.27
N	-2.30 (14.96)	+6.50 (12.30)	-2.169	.038	.37	-1.33 (18.54)	-3.13 (11.46)	.416	.681	.08
V	-.57 (10.44)	+9.00 (10.98)	-3.088	.004	.50	-3.03 (10.17)	+4.40 (9.49)	-1.140	.264	.21
FW	-7.00 (45.18)	+30.40 (43.46)	-2.891	.007	.47	-8.10 (54.26)	-3.80 (44.44)	-.331	.743	.06
Conj	-.23 (7.70)	+4.90 (9.77)	-1.859	.073	.33	+4.0 (11.15)	-2.53 (9.71)	1.056	.300*	.19
Prep	-2.90 (9.12)	+6.77 (8.49)	-3.699	.001	.57	-1.77 (10.27)	-.40 (7.49)	-.526	.603	.10
Pron	-1.07 (11.76)	+7.60 (11.23)	-2.411	.022	.41	-3.73 (15.75)	+7.77 (11.46)	-1.179	.248	.21

Chapter 4

	Written production					Oral production				
	FI gains	SA gains	<i>t</i> (29)	<i>p</i>	<i>r</i>	FI gains	SA gains	<i>t</i> (29)	<i>p</i>	<i>r</i>
CW/w	+43 (4.29)	-19 (3.56)	.494	.625	.09	+25 (4.03)	+61 (3.03)	-.323	.749	.06
Adj/w	+74 (2.69)	-73 (2.05)	1.975	.058	.34	-.27 (2.21)	+15 (2.19)	-.594	.557	.11
Adv/w	+16 (1.74)	+80 (2.21)	-.983	.334	.18	+23 (2.62)	1.37 (1.93)	-1.824	.079	.32
N/w	-.62 (2.96)	-.98 (2.51)	.437	.667	.08	+71 (3.04)	-1.09 (2.18)	2.417	.022	.41
V/w	+15 (2.32)	+72 (2.34)	-.842	.407	.15	-.43 (1.96)	+19 (2.51)	-.852	.401	.16
FW/w	-.84 (4.24)	-.00 (3.60)	-.639	.528	.12	+57 (3.41)	-1.20 (3.98)	1.510	.142	.27
Conj/w	+11 (2.16)	-.12 (2.18)	.340	.737	.06	+1.08 (1.80)	-1.17 (2.35)	3.439	.002	.54
Prep/w	-.98 (1.90)	+51 (2.63)	-2.114	.043	.37	-.25 (1.74)	+03 (1.86)	-.485	.631	.09
Pron/w	+21 (3.54)	+58 (2.74)	-.379	.707	.07	-.36 (2.53)	+44 (2.29)	-1.092	.284	.20
GI	-.33 (.75)	+76 (.76)	-4.825	.000	.67	+05 (.51)	-.02 (.54)	.516	.610	.10
D	-3.84 (14.17)	+5.02 (13.32)	-2.201	.036	.38	-1.33 (7.71)	3.01 (8.57)	-1.719	.096	.30
K1	+46 (2.71)	-.50 (2.90)	1.027	.313	.19	-1.15 (2.32)	+75 (1.96)	-2.883	.007	.47
K2	-.17 (2.12)	-.42 (1.83)	.406	.687	.08	+45 (1.59)	-.24 (1.50)	1.597	.121	.28
AWL	-.11 (1.72)	+69 (2.04)	-1.284	.209	.23	+13 (.89)	+02 (.91)	.359	.722	.07
OffL	-.18 (1.04)	+23 (.99)	-1.512	.141	.27	+57 (1.24)	-.54 (1.20)	2.835	.008	.47
GLI	-.00 (.03)	-.00 (.03)	.110	.913	.02	+01 (.03)	-.01 (.03)	1.406	.170	.25
E/w	-1.81 (3.22)	-1.39 (2.29)	-.529	.601	.10	-.89 (3.04)	-1.30 (2.82)	.411	.684	.08
LexE/w	-.40 (1.42)	-.48 (1.15)	.191	.850	.04	-.42 (.90)	-.44 (.97)	.098	.923	.02
MjSp/w	+11 (1.12)	-.18 (1.11)	.794	.434	.15	N/A	N/A	N/A	N/A	.13

Note: positive values (+) represent gains, negative values (-) indicate losses.

*Variable (Conj. in oral data) submitted to non-parametric techniques (Wilcoxon Signed Rank).

Shaded cells indicate significance.

As can be seen in Table 25, within the written production modality SA gains were significantly larger than FI gains for most fluency measures (tokens, types, CW, Adv, Nouns, Verbs, FW, Prep, Pron), followed by preposition density (Prep/w) and lexical diversity (GI and D) indices.

Results

The effect sizes⁴², as seen through r , ranged from medium to large, with the largest effect size found for GI, $r=.67$, accounting for almost 45% of the total variance. A greater amount of gains accrued while abroad than those obtained at home suggested that SA was more beneficial to learners' writing than the FI context. Furthermore, these gains indicated a trend toward NS usage, as reported earlier in Table 12. We did not register any significant differences for the size of gains in lexical sophistication and accuracy.

Within the oral production modality, the pattern for gains in most domains was the reverse: there was no significant difference when comparing FI and SA, in fluency and lexical diversity gains, contrary to what was found for writing. We did find significantly larger gains in noun density and conjunction density resulting from the FI learning context, with the effect sizes between moderate and large. However, unlike with noun density, gains in conjunction density during FI did not translate into a shift towards native-like usage, as learners tended to overuse coordination and subordination connectives in comparison to NS baseline rates (see Table 10). Gains between the two learning contexts were also significant in the domain of lexical sophistication, as measured by K1 and OffL, with medium-large effect sizes, $r=.47$ in both cases. More specifically, learners obtained larger gains in sophisticated vocabulary during FI than during SA, becoming more target-like in the L2. We did not register any significant differences for the size of gains in the remaining measures of lexical sophistication such as K2, AWL, GLI, and accuracy. Wilcoxon Signed Rank tests

⁴² Following Morris (2008), the effect size (r) was calculated using this equation: $r = \sqrt{t^2 / (t^2 + df)}$.

confirmed non-significant results for conjunction frequency in oral production Conj [$z=-1.442, p=.149$].

In light of the t -test results, we can consider that in written production SA was more beneficial to learners' vocabulary acquisition than FI, as seen in the progress made in lexical fluency and diversity. Conversely, the results for oral production indicate that greater gains were obtained mostly during the FI period, as seen in noun density and lexical sophistication.

4.2.2. Results for vocabulary gains across modalities

The next step we took in addressing RQ2b consisted in determining the amount of gains as a function of the type of production modality. To that end, we again used *FI gains*, calculated as T2 scores – T1 scores, and *SA gains*, calculated as T3 scores – T2 scores but this time we related them across production modalities (written vs. oral) for each learning context independently. In other words, FI gains in written production were compared to FI gains in oral production, as SA gains in written production were contrasted against SA gains in oral production. Subsequently, we conducted a set of paired-samples t -tests on these gains, which are presented in Table 26.

Table 26. Paired-samples t -tests comparing written vs. oral modality gains in each learning context.

	FI gains					SA gains				
	Written mode	Oral mode	$t(29)$	p	r	Written mode	Oral mode	$t(29)$	p	r
Tokens	-8.77 (79.30)	-17.13 (96.66)	.359	.722	.07	+57.10 (75.53)	-.97 (79.41)	3.211	.003	.51
Types	+6.80 (27.71)	-2.77 (25.42)	-.574	.570	.11	+25.10 (26.91)	-1.00 (22.00)	4.812	.000	.67
CW	-1.67 (31.26)	-5.77 (35.26)	.453	.654	.08	+22.13 (30.15)	+1.40 (28.86)	2.857	.008	.47

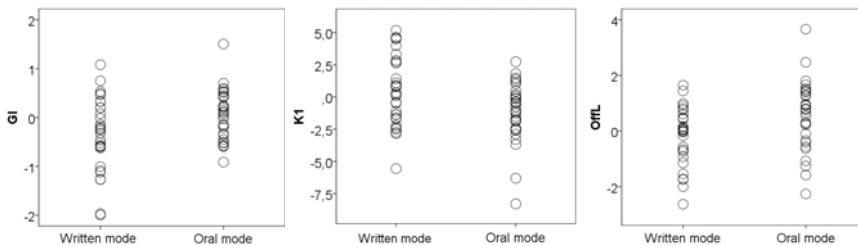
Results

	FI gains					SA gains				
	Written mode	Oral mode	<i>t</i> (29)	<i>p</i>	<i>r</i>	Written mode	Oral mode	<i>t</i> (29)	<i>p</i>	<i>r</i>
Adj	+1.27 (6.87)	-1.43 (6.07)	1.738	.093	.31	+2.47 (7.32)	+73 (6.20)	1.177	.249	.21
Adv	-.07 (5.20)	+.03 (8.50)	-.054	.957	.01	+4.17 (5.87)	+3.40 (8.89)	.379	.707	.07
N	-2.30 (14.96)	-1.33 (18.54)	-.213	.833	.04	+6.50 (12.30)	-3.13 (11.46)	3.517	.001	.55
V	-.57 (10.44)	-3.03 (10.17)	.815	.422	.15	+9.00 (10.98)	+40 (9.49)	3.431	.002	.54
FW	-7.00 (45.18)	-8.10 (54.26)	.086	.932	.02	+30.40 (43.46)	-3.80 (44.44)	3.553	.001	.55
Conj	-.23 (7.70)	+.40 (11.15)	-.256	.800*	.08	+4.90 (9.77)	-2.53 (9.71)	3.863	.001	.58
Prep	-2.90 (9.12)	-1.77 (10.27)	-.431	.669	.08	+6.77 (8.49)	-.40 (7.49)	4.092	.000	.61
Pron	-1.07 (11.76)	-3.73 (15.75)	.718	.478	.13	+7.60 (11.23)	+77 (11.46)	2.547	.016	.43
CW/w	+.43 (4.29)	+.25 (4.03)	.169	.867	.03	-.19 (3.56)	+.61 (3.03)	.944	.353	.17
Adj/w	+.74 (2.69)	-.27 (2.21)	1.580	.125	.28	-.73 (2.05)	+.15 (2.19)	-1.931	.063	.34
Adv/w	+.16 (1.74)	+.23 (2.62)	-.139	.890	.03	+.80 (2.21)	1.37 (1.93)	-1.078	.290	.20
N/w	-.62 (2.96)	+.71 (3.04)	-1.748	.091	.31	-.98 (2.51)	-1.09 (2.18)	.189	.851	.04
V/w	+.15 (2.32)	-.43 (1.96)	1.010	.321	.18	+.72 (2.34)	+.19 (2.51)	.981	.335	.18
FW/w	-.84 (4.24)	+.57 (3.41)	1.387	.176	.25	-.00 (3.60)	-1.20 (3.98)	1.250	.221	.23
Conj/w	+.11 (2.16)	1.08 (1.80)	-1.884	.070	.33	-.12 (2.18)	-1.17 (2.35)	2.006	.054	.35
Prep/w	-.98 (1.90)	-.25 (1.74)	-1.410	.169	.25	+.51 (2.63)	+.03 (1.86)	.938	.356	.17
Pron/w	+.21 (3.54)	-.36 (2.53)	.809	.425	.15	+.58 (2.74)	+.44 (2.29)	.195	.846	.04
GI	-.33 (.75)	+.05 (.51)	-2.270	.031	.39	+.76 (.76)	-.02 (.54)	5.099	.000	.69
D	-3.84 (14.17)	-1.33 (7.71)	-.810	.425	.15	5.02 (13.32)	3.01 (8.57)	.706	.486	.13
K1	+.46 (2.71)	-1.15 (2.32)	2.341	.026	.40	-.50 (2.90)	+.75 (1.96)	-1.943	.062	.34
K2	-.17 (2.12)	+.45 (1.59)	-1.188	.244	.23	-.42 (1.83)	-.24 (1.50)	-.398	.693	.07
AWL	-.11 (1.72)	+.13 (.89)	-.607	.548	.11	+.69 (2.04)	+.02 (.91)	1.640	.112	.29
OffL	-.18 (1.04)	+.57 (1.24)	-2.490	.019	.42	+.23 (.99)	-.54 (1.20)	1.454	.160	.24
GLI	-.00 (.03)	+.01 (.03)	-1.309	.201	.24	-.00 (.03)	-.01 (.03)	.533	.598	.10
E/w	-1.81 (3.22)	-.89 (3.04)	-1.179	.248	.21	-1.39 (2.29)	-1.30 (2.82)	-.157	.876	.03
LexE/w	-.40 (1.42)	-.42 (.90)	.040	.969	.01	-.48 (1.15)	-.44 (.97)	-.129	.898	.02

Note: positive values (+) represent gains, negative values (-) indicate losses. *-see p. 158

By looking at the results summary in Table 26, we see that during FI the size of gains between written and oral production modalities does not seem to differ substantially, except for lexical diversity (GI) and sophistication (K1, OffL). On average, participants obtained more gains in speaking than in writing with moderate-large effect sizes (r values ranging from .39 to .42). We did not find any significant differences in gains for the remaining variables. Wilcoxon Signed Rank tests confirmed non-significant results for the non-normally distributed conjunction frequency in oral data Conj. [$z=-1.184$, $p=.237$]. Scatterplots of individual gains scores in written and oral production modes are provided in Figure 8 (for FI gains) and Figure 9 (for SA gains).

Figure 8. Scatterplots of individual FI gains in written vs. oral production modalities (significant changes).

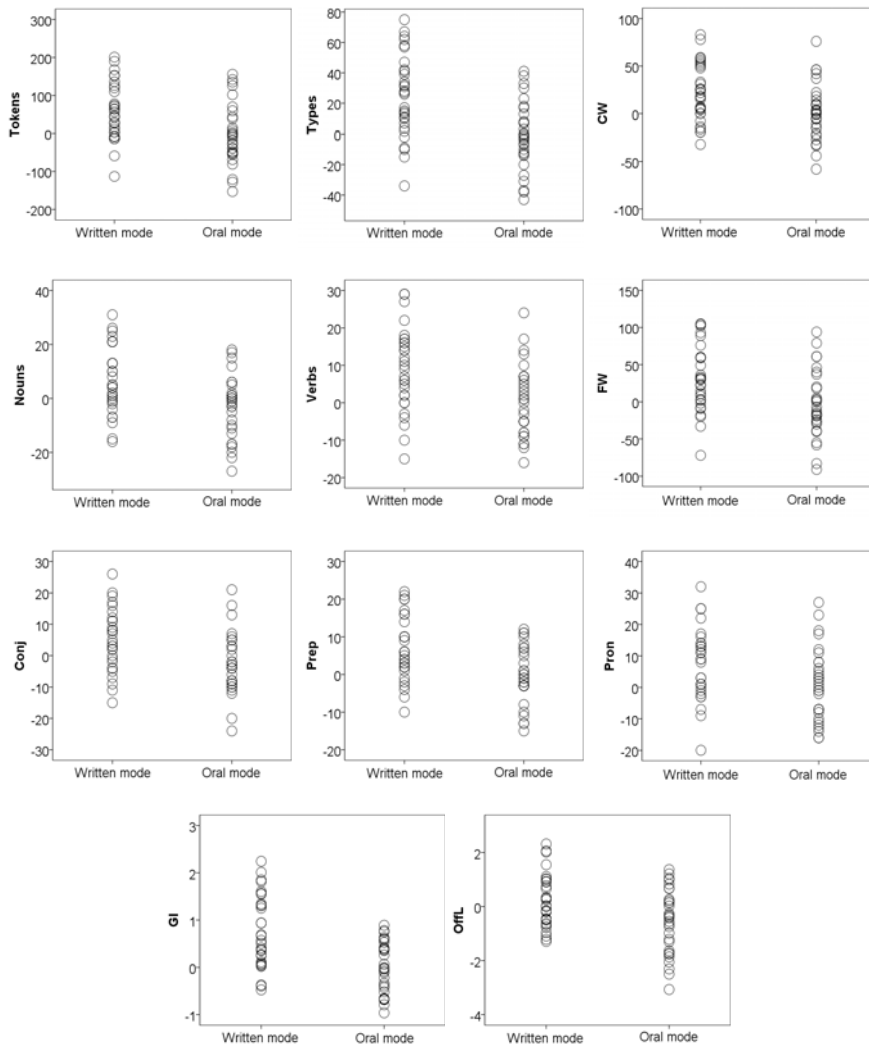


During SA, a rather different pattern is observed, with more gains in writing than in speaking, primarily for most fluency measures (with quite large effect sizes). This is also the case for lexical diversity (GI), which accrues greater gains in writing than in speaking, showing the largest effect size of $r=.69$.

*Variable (Conj. in oral data) submitted to non-parametric techniques (Wilcoxon Signed Rank). Shaded cells indicate significance.

Results

Figure 9. Scatterplots of individual SA gains in written vs. oral production modalities (significant changes).

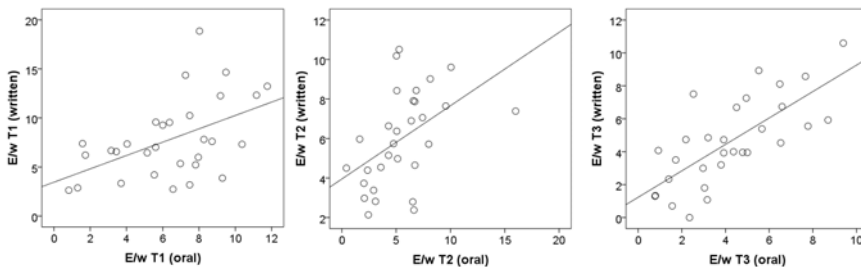


To further look into the relationship between the written and oral production modalities and complete these final analyses, we opted to run a series of *Pearson r* correlations on learners' written and oral production lexical scores. This was done so as to uncover possible relationships between the two modes from a within-subjects perspective: for instance, learners producing more varied and

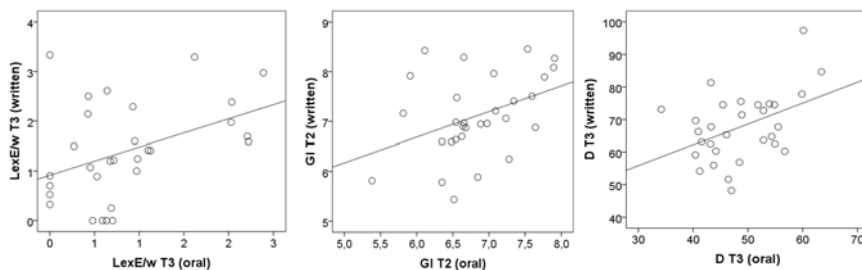
sophisticated vocabulary in writing might also be making a more advanced use of lexis when speaking, which would suggest a comparable degree of competence in lexical richness across the two production modes.

Analyses revealed that there were significant correlations between written and oral scores in the domain of accuracy and lexical diversity. As far as accuracy is concerned, written and oral production scores were significantly correlated at T1 (E/w, $r=.50$, $p=.005$), T2 (E/w, $r=.48$, $p=.007$), and T3 (E/w, $r=.73$, $p=.000$; LexE/w, $r=.41$, $p=.024$), suggesting that more accurate learners in writing were also more accurate in speaking at three testing times, and especially at T3. Significant correlations were also found between oral and written lexical diversity at T2 (GI, $r=.39$, $p=.032$) and T3 (D, $r=.43$, $p=.017$), implying a similar level of competence in lexical variation for both modes after FI and SA learning contexts. As can be seen in scatterplots (see Figure 10), the strength of these correlations was moderate, except for accuracy (E/w at T1 and T3), where it was rather strong.

Figure 10. Scatterplots showing significant associations between written and oral vocabulary gains.



Results



In reply to our question on vocabulary gains across written and oral production, we can see contrasting findings according to linguistic modality. While during the FI period greater gains were obtained in oral production in terms of lexical diversity and sophistication (as seen in GI, K1, OffL), in SA significantly larger gains were achieved in written modality and were found in the area of fluency (Tokens, Types, CW, Nouns, Verbs, FW, Conj, Prep, Pron) followed by lexical diversity and sophistication measures (GI and OffL). Finally, with the correlations results pointing to positive relationships between written and oral modalities in two dimensions (accuracy and diversity), we can consider that although the tasks used to elicit written and oral data are different in nature and may involve different underlying processes in language production, learners seem to be using more varied and accurate vocabulary irrespective of the production modality across time.

4.3. RQ3: The role of initial level

In this section, we explore the role of initial proficiency level on lexical development in FI and SA contexts, in response to RQ3, which was formulated as follows:

RQ3) Are there different patterns of vocabulary development and gains associated with learners' onset level?

RQ3a) Are there different patterns of vocabulary development and gains associated with learners' onset level, as measured through initial vocabulary knowledge?

RQ3b) Are there different patterns of vocabulary development and gains associated with learners' onset level, as measured through initial global proficiency?

To that end, we considered initial level of proficiency through several approaches. First, we measured learners' initial level based on T1 vocabulary scores in each of the domains in question (fluency, density, diversity, sophistication and accuracy) in response to RQ3a (Section 4.3.1.). Secondly, we assessed their level through initial global proficiency, using a combined score on the cloze and rephrasing tests at T1 in order to answer RQ3b (Section 4.3.2.).

4.3.1. The role of initial vocabulary knowledge

In order to begin the examination of the effect of initial level as a function of learners' initial vocabulary knowledge, our first step was to divide learners into high and low initial level groups based on their lexical scores at T1. For reasons of space, we selected two measures from each domain, which have shown to discriminate in various statistical analyses (e.g., RM-ANOVAs, *t*-tests), and include tokens, types, CW/w, FW/w, GI, D, AWL, OffL, and E/w. We then split learner groups using K-means cluster analysis. Descriptive statistics for each group are provided in Table 27 for written data and Table 28 for oral data.

Results

Table 27. Lexical proficiency mean scores of high/low initial level groups in written production (SD in parentheses).

Written production		<i>n</i> high, low	High initial level			Low initial level		
			T1	T2	T3	T1	T2	T3
Fluency	Tokens	14,16	273.80 (71.97)	221.20 (45.60)	292.87 (71.25)	165.00 (21.07)	200.07 (43.96)	242.60 (87.43)
	Types	15,15	129.60 (20.91)	107.60 (20.79)	133.40 (19.45)	89.80 (10.47)	98.20 (14.09)	122.60 (30.83)
Density	CW/w	12,18	44.12 (2.33)	42.22 (3.81)	42.59 (3.09)	38.06 (2.64)	40.04 (3.43)	39.48 (3.82)
	FW/w	15,15	50.56 (1.73)	51.49 (3.58)	50.30 (3.18)	55.76 (1.42)	53.15 (3.99)	54.32 (2.42)
Diversity	GI	11,19	8.25 (.46)	7.86 (.50)	8.31 (.69)	6.98 (.45)	6.67 (.62)	7.62 (.61)
	D	14,16	79.09 (8.14)	68.92 (15.00)	69.92 (11.92)	55.60 (8.48)	57.31 (13.28)	65.84 (8.86)
Sophistication	AWL	13,17	5.07 (1.01)	4.10 (1.88)	5.33 (1.86)	2.24 (.92)	2.80 (1.54)	3.07 (1.51)
	OffL	13,17	1.83 (.49)	1.12 (1.15)	1.17 (1.24)	.48 (.41)	.70 (.48)	1.07 (1.00)
Accuracy	E/w	14,16	4.82 (1.62)	4.95 (2.05)	3.64 (2.42)	10.79 (3.42)	7.04 (2.31)	5.57 (2.56)

Table 28. Lexical proficiency mean scores of high/low initial level groups in oral production (SD in parentheses).

Oral production		<i>n</i> high, low	High initial level			Low initial level		
			T1	T2	T3	T1	T2	T3
Fluency	Tokens	16,14	365.53 (87.26)	303.20 (91.61)	282.73 (126.21)	192.87 (41.62)	220.93 (75.09)	239.47 (82.77)
	Types	16,14	138.73 (23.85)	123.60 (25.99)	115.20 (30.87)	87.07 (15.14)	96.67 (22.26)	103.07 (25.83)
Density	CW/w	14,16	37.58 (2.08)	35.22 (2.58)	36.55 (2.83)	32.25 (1.80)	34.78 (2.45)	34.76 (2.17)
	FW/w	15,15	52.25 (1.71)	54.20 (2.73)	53.19 (3.04)	57.10 (2.02)	56.29 (2.08)	54.90 (3.01)
Diversity	GI	18,12	7.27 (.43)	7.12 (.49)	6.97 (.64)	6.03 (.37)	6.40 (.56)	6.56 (.43)
	D	18,12	52.65 (5.35)	48.10 (5.42)	51.12 (6.70)	38.38 (4.20)	41.89 (8.54)	44.89 (6.01)
Sophistication	AWL	13,17	1.56 (.44)	1.23 (.73)	1.26 (.74)	.53 (.38)	1.02 (.89)	1.04 (.68)
	OffL	13,17	2.18 (.62)	1.99 (1.05)	1.28 (.74)	.54 (.42)	1.68 (1.00)	1.27 (.65)
Accuracy	E/w	13,17	3.67 (1.85)	4.34 (3.99)	2.87 (1.70)	8.48 (1.55)	6.40 (1.84)	5.23 (2.36)

Descriptive statistics revealed that the distributions were normal, despite a low number of participants in each group. Therefore, we

conducted a mixed-design ANOVA with *Time* as within-subjects factors and *Initial level* (high, low), as the between-subjects factor (assumption of sphericity and homogeneity of intercorrelations met for all variables). These analyses yielded significant *Time* x *Initial level* interaction for all measures in oral production [for all cases $F(2,56) > 4.644$, $p < .018$] and most measures in written production [in all cases $F(2,56) > 2.416$, $p < .025$], except for GI and AWL, where no interactions reached significance. Within-subjects contrasts further revealed that the interaction was significant in all cases between T1 and T2, but not between T2 and T3, pointing to a more impactful interaction effects as a result of the FI learning context. To confirm these impressions, we next conducted separate one-way repeated measures ANOVAs for each of the two initial level groups. Results of these tests can be consulted in Table 29.

Table 29. Changes in lexical proficiency scores of high/low initial level learners across contexts.

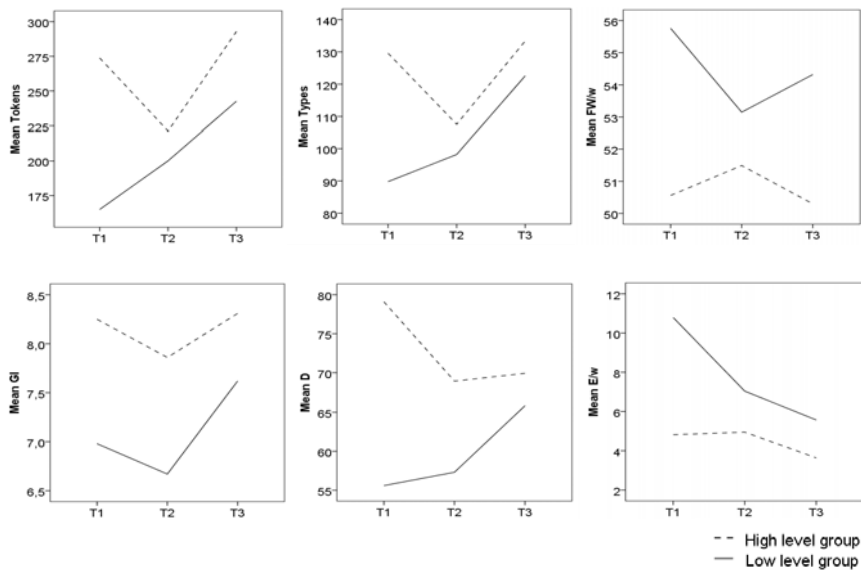
		Written production						Oral production					
		df	F	<i>p</i>	η^2p	FI	SA	df	F	<i>p</i>	η^2p	FI	SA
Tokens	High	2,13	10.446	.002	.62	.095	.002	2,13	3.043	.082	.32	.101	1.00
	Low	2,13	10.032	.002	.61	.014	.219	2,13	3.580	.058	.36	.376	.869
Types	High	2,13	10.983	.002	.63	.001	.033	2,13	5.172	.022	.44	.102	.462
	Low	2,13	10.551	.002	.62	.177	.033	2,13	3.564	.058	.35	.239	.740
CW/w	High	2,10	1.091	.373	.18	.506	1.00	2,12	2.685	.109	.31	.095	.466
	Low	2,16	3.034	.076	.28	.078	1.00	2,14	7.160	.007	.51	.008	1.00
FW/w	High	2,13	1.848	.197	.22	1.00	.198	2,13	2.098	.162	.24	.192	1.00
	Low	2,13	5.693	.017	.47	.049	.925	2,13	3.400	.065	.34	.651	.302
GI	High	2,9	3.243	.087	.41	.271	.175	2,16	2.027	.164	.22	.432	.692
	Low	2,17	14.984	.000	.63	.347	.000	2,10	7.100	.012	.59	.051	1.00
D	High	2,12	4.060	.045	.40	.042	1.00	2,16	5.358	.017	.40	.017	.303
	Low	2,14	8.033	.005	.53	1.00	.027	2,10	8.787	.006	.64	.401	1.00
AWL	High	2,11	3.224	.079	.37	.191	.084	2,11	2.305	.146	.30	.309	1.00
	Low	2,15	3.184	.070	.30	.388	1.00	2,15	4.810	.024	.39	.123	1.00
OffL	High	2,11	1.972	.185	.26	.183	1.00	2,11	4.532	.037	.45	.198	1.00
	Low	2,15	3.246	.067	.30	.477	.431	2,15	19.515	.000	.72	.000	.516
E/w	High	2,13	1.779	.208	.22	1.00	.233	2,11	2.079	.171	.27	1.00	.416
	Low	2,13	23.083	.000	.78	.001	.034	2,15	29.381	.000	.80	.002	.204

Shaded cells indicate significance.

Results

Through these results, we found that in written production the high-level group experienced improvement in fluency after SA producing significantly more tokens but sharply declining in the number of types and the lexical diversity index D after the FI context. The low-level group, in contrast, showed sustained improvement across the contexts on several measures: on one hand, increasing significantly in the number of tokens, making fewer errors per word (with the largest effect size of $\eta^2 p=.78$) and reducing the proportion of function words (FW/w) in the direction of NS levels over the FI context, and, on the other hand, producing more types and showing a significant jump in both measures of lexical diversity (with quite large effect sizes) after SA. Line graphs in Figure 11 visualise these changes with learners, arranged by high/low initial level.

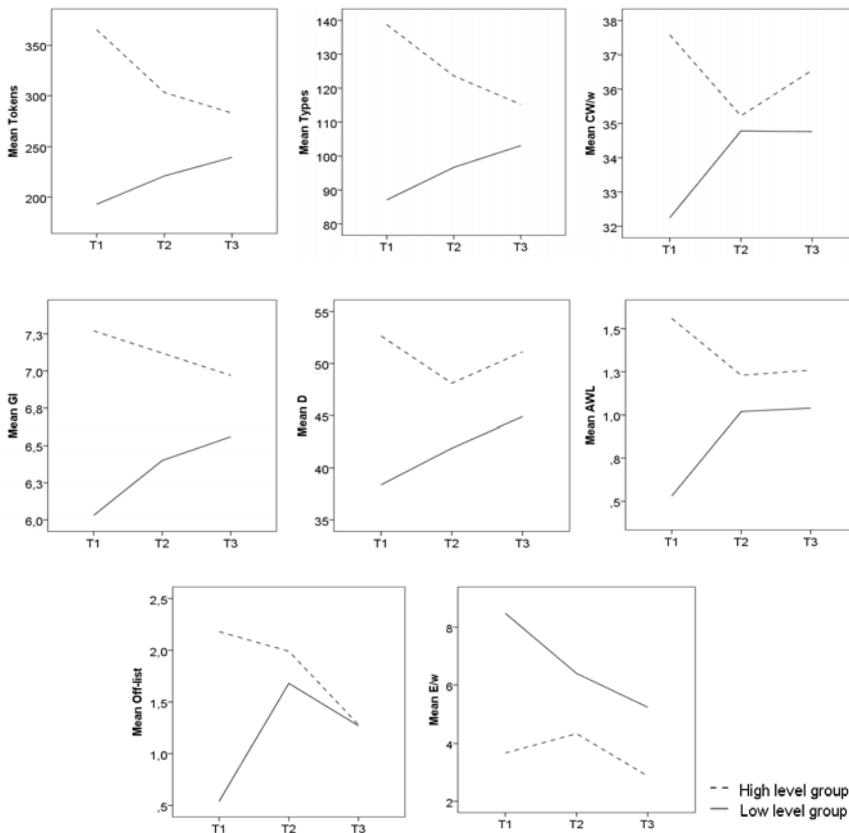
Figure 11. Line graphs comparing written lexical development over time of high/low initial level learners. Significant changes.



Chapter 4

In oral production, the high-level group showed a significant drop in lexical diversity, as measured by D, after the FI period, while the low-level group significantly increased in the proportion of content words (CW/w), the amount of sophisticated vocabulary (OffL) and produced fewer errors per word (with very large effect sizes). Low scoring learners also showed steady improvement over time in lexical diversity (both GI and D), and academic vocabulary (AWL) measures, making considerable headway by the end of the treatment. A graphical summary of mean scores across groups and times for oral data is provided in Figure 12.

Figure 12. Line graphs comparing oral lexical development over time of high/low initial level learners. Significant changes.



Results

To complement these analyses, we also conducted independent-samples *t*-tests in an attempt to find out whether the low-level group caught up with the high-level group by T3. Results of *t*-tests in writing revealed that scores were inherently significantly different at the beginning of the study (T1), with the low scorers converging with the high scorers on the two measures of lexical density (CW/w and FW/w) after FI (T2), and one measure of lexical diversity (D) after SA (T3). Low scoring learners also caught up with the high scorers at both T2 and T3 fluency (Types and Tokens), and sophistication (OffL) measures. However, the high-level group continued to score significantly higher than the low-level group on accuracy [for T2 and T3 $t(28) > -6.112, p < .043$], academic vocabulary [for T2 and T3 $t(28) > -7.075, p < .048$], and Guiraud's Index [for T2 and T3 $t(28) > -7.425, p < .009$].

In speaking, results of *t*-tests at T2 and T3 revealed the low-level group converged with the high-level group after FI (T2) on lexical sophistication (AWL, OffL) and lexical density (CW/w) measures, maintaining non-significant distinctions also after the SA period (T3). To the list of measures that were not significantly different for low scorers and high scorers after SA (T3), we add fluency measures (Types and Tokens), function word density (FW/w), and Guiraud's Index. Despite making substantially more progress, the low scorers did not catch up with the high-level group on accuracy [for T2 and T3 $t(28) > 1.892, p < .005$] and lexical diversity, when this was measured by D [for T2 and T3 $t(28) > 2.445, p < .021$].

Finally, the last step we took in tackling RQ3a was to determine whether learners obtained significantly different outcomes after each individual context as a function of initial vocabulary knowledge. To do so, FI gains and SA gains in each of the measure on focus were submitted to independent samples *t*-tests where *Initial level* (based on their T1 vocabulary scores) acted as the grouping variable comparing the high-level group to the low-level group. All of the measures were found to be normally distributed across the groups. Each production mode was assessed separately and the results are reported in Table 30.

Table 30. Independent-samples *t*-tests comparing context-specific gains for high/low initial level groups.

	Grup	Written production						Oral production					
		Gains FI			Gains SA			Gains FI			Gains SA		
	Mean	<i>t</i>	<i>p</i>	Mean	<i>t</i>	<i>p</i>	Mean	<i>t</i>	<i>p</i>	Mean	<i>t</i>	<i>p</i>	
Tokens	High	-52.60 (85.34)	3.597	.002	+71.67 (64.41)	-1.059	.299	-62.33 (102.60)	2.861	.008	-20.47 (89.41)	1.365	.183
	Low	+35.07 (40.32)			+42.53 (84.93)			+28.07 (66.70)			+18.53 (65.22)		
Types	High	-22.00 (29.07)	3.557	.001	+25.80 (21.47)	-1.140	.890	-15.13 (24.97)	3.013	.005	-8.40 (21.58)	1.926	.064
	Low	+8.40 (15.82)			+24.40 (32.22)			+9.60 (19.68)			+6.40 (20.50)		
CW/w	High	-1.91 (4.49)	2.686	.012	+3.7 (3.38)	-0.696	.492	-2.36 (3.66)	-4.137	.000	1.34 (3.31)	1.235	.227
	Low	+1.99 (3.45)			-.56 (3.73)			+2.53 (2.80)			-.02 (2.70)		
FW/w	High	.93 (4.09)	-2.486	.019	-1.18 (2.30)	1.872	.072	1.95 (3.75)	2.391	.024	-1.00 (4.83)	.262	.795
	Low	-2.62 (3.72)			1.18 (4.31)			-.81 (2.44)			-1.39 (3.07)		
GI	High	-.39 (.69)	.312	.757	+.95 (.76)	1.814	.080	-.16 (.44)	-3.178	.004	-.14 (.49)	-1.537	.136
	Low	-.30 (.79)			+.44 (.69)			+.37 (.46)			+.16 (.60)		
D	High	-10.17 (13.41)	2.487	.019	+1.00 (14.63)	1.587	.124	-4.55 (6.12)	-3.227	.003	+3.02 (7.39)	.004	.997
	Low	+1.71 (12.73)			+8.54 (11.37)			+3.51 (7.51)			+3.01 (10.46)		
AWL	High	-.97 (1.72)	2.656	.013	+1.23 (2.19)	-1.290	.208	-.33 (.67)	-2.728	.011	+.03 (.82)	.025	.980
	Low	+.56 (1.44)			+.27 (1.78)			+.48 (.90)			+.02 (1.00)		

Results

		Written production						Oral production					
		Gains FI			Gains SA			Gains FI			Gains SA		
	Grup	Mean	<i>t</i>	<i>p</i>	Mean	<i>t</i>	<i>p</i>	Mean	<i>t</i>	<i>p</i>	Mean	<i>t</i>	<i>p</i>
OffL	High	-.71	2.703	.024	+.05	.882	.384	-.19	3.388	.002	-.71	.699	.491
		(1.25)			(.96)			(1.19)			(1.27)		
	Low	+.22			+.37			+1.14			-.40		
		(.62)			(1.00)			(.96)			(1.16)		
E/w	High	-.75	4.047	.000	-1.46	-.322	.750	+6.7	-2.710	.011	-1.46	.272	.788
		(2.40)			(2.26)			(3.47)			(3.32)		
	Low	-5.30			-1.14			-2.08			-1.17		
		(3.25)			(2.56)			(2.06)			(2.47)		

Shaded cells indicate significance.

Analysis revealed that within written production there were statistically significant differences across FI gains scores, indicating that high- and low-level groups gained differently in all lexical measures (except for GI) after the FI period. This was also the case for the outcomes within oral production, with statistically significant results between groups in all lexical variables in terms of FI gains. Nevertheless, it is often claimed that with a small sample size, even if the distribution is normal, results should be considered statistically significant if the *p* value is less .01. If we are to keep to this pattern and only consider the results below the cut-off of .01 as significant, we limit statistically significant differences pertaining to FI gains to the domains of fluency (Types, Tokens) and accuracy (E/w) as for writing, and in fluency (Types, Tokens), content word density (CW/w), lexical richness (GI and D), and sophistication (OffL) as for speaking. This adjustment, however, does not represent change, confirming that low-level learners tend to reap greater benefits than their high-level peers, particularly in the FI context, despite beginning at a relative disadvantage. As for SA gains, we did not find any significant differences between high- and low-level groups, which could plausibly be attributed to the fact that the initial level groupings were based on

learners' T1 scores. Had we classified learners into high/low groups according to their T2 scores, as it was done in previous SALA studies (Lara, 2014; Valls-Ferrer, 2011), we may have obtained different outcomes.

In wrapping up our look at the effect of initial vocabulary knowledge on learners' productive vocabulary development over time, after either FI or SA, and in terms of gains, we found that the pattern of development based on a selection of lexical measures (Tokens, Types, CW/w, FW/w, GI, D, AWL, OffL, E/w) was indeed significantly different for students with low onset level. More specifically, low-level learners were able to catch up to high initial level participants after either FI or SA (or both) in all measures (except for E/w, AWL, GI in writing, and E/w and D in speaking), showing a greater propensity towards gain than their high-level counterparts.

4.3.2. The role of initial general proficiency

Our second approach when evaluating the impact of onset level consisted in grouping learners according to their initial global proficiency, as measured through a composite score on a rephrasing task and a cloze test⁴⁴. Before grouping learners and proceeding with the analysis, we explored the relationship between productive vocabulary outcomes, and the composite rephrasing-cloze score in the EFL corpus, so as to evaluate to the degree of correspondence

⁴⁴ A preliminary analysis of the relationship between cloze and grammar scores yielded a strong, positive correlation between the two variables [$r_s=.798$, $n=117$, $p<.0001$], justifying our intention to combine the two scores.

Results

between these measures and the extent to which they were capturing different aspects of L2 proficiency. To quantify this relationship, we ran Spearman's non-parametric rank correlation coefficients (r_s), as a graphical exploration of the grammar and cloze variables revealed a non-normal, positively skewed distribution of the scores. Learners' rephrasing-cloze score and vocabulary measures were correlated at all three testing times and for each production modality separately (see Table 31 for correlation analysis in written data, and Table 32 for correlation analysis in oral data).

Table 31. Correlations between grammar-cloze scores and lexical proficiency measures in written production ($n=90$).

	Fluency		Density		Diversity		Sophistication		Accuracy	
	r_s		r_s		r_s		r_s		r_s	
Tokens	.008	CW/w	.147	GI	.262*	K1	-.174	E/w	-.676**	
Types	.121	Adj/w	.036	D	.309**	K2	-.022	LexE/w	-.438**	
CW	.058	Adv/w	.260*			AWL	.255*	SpE/w	-.196	
Adj	.085	N/w	-.023			OffL	.389	Slips/w	.038	
Adv	.240*	V/w	-.019			GLI	.230*			
N	.043	FW/w	-.209*							
V	.023	Conj/w	-.139							
FW	-.049	Prep/w	.059							
Conj	-.095	Pron/w	-.141							
Prep	.110									
Pron	-.086									

* Correlation is significant at the .05 level (2-tailed). ** Correlation is significant at the .01 level (2-tailed). Shaded values indicate significance.

Table 32. Correlations between grammar-cloze score and lexical proficiency measures in oral production ($n=90$).

	Fluency		Density		Diversity		Sophistication		Accuracy	
	r_s		r_s		r_s		r_s		r_s	
Tokens	.048	CW/w	-.018	GI	.187	K1	-.014	E/w	-.669**	
Types	.110	Adj/w	-.115	D	.115	K2	-.126	LexE/w	-.447**	
CW	.042	Adv/w	.291**			AWL	.019			
Adj	-.025	N/w	-.177			OffL	.096			
Adv	.187	V/w	.062			GLI	-.012			
N	-.025	FW/w	-.205							
V	.065	Conj/w	-.395**							
FW	.002	Prep/w	-.019							
Conj	-.135	Pron/w	.109							
Prep	.011									
Pron	.078									

* Correlation is significant at the .05 level (2-tailed). ** Correlation is significant at the .01 level (2-tailed). Shaded values indicate significance.

From the report of correlations, we can infer that the rephrasing and cloze scores correlated better with the written vocabulary measures than with the oral. In written production, we found significant positive correlations in adverb frequency (Adv) and density (Adv/w), and a negative correlation in FW density, indicating that learners with higher general proficiency used more adverbs and fewer function words in composing written discourse. Both lexical diversity indices, GI and D, correlated positively with the rephrasing and cloze scores, suggesting that high scoring learners were also more lexically diverse. The same was true for two lexical sophistication measures, AWL and GLL, with higher scoring learners producing a higher proportion of academic vocabulary and Greco-Latin derived words, approaching NS usage. We found strong negative correlations between the composite score and accuracy (E/w and LexE/w). This indicated that participants who scored high on the rephrasing and cloze tests were also more accurate in writing.

In oral production, just as in written, we found a strong negative relationship between the composite rephrasing-cloze score and accuracy (E/w, LexE/w), indicating that learners with higher general proficiency produced more accurate vocabulary. On the other hand, the composite score was negatively correlated with conjunction density and positively correlated with adverb density, suggesting that higher general proficiency learners produced more adverbs and fewer conjunctions in their speech. We found no significant correlations in any other area on focus. In sum, we did not find a strong relationship between global proficiency and vocabulary knowledge in our data, except for accuracy.

Results

In order to begin the examination of the effect of initial level as a function of learners' general English proficiency, we grouped learners according to their composite score on the rephrasing and cloze tests at T1. These high- and low-level groups were created using K-means cluster analysis. The low group scored a mean of 5.52 (sd=1.78, range=3.00-7.00) out of 20, and the high group obtained a mean of 10.04 (sd=2.34, range=8.00-15.00).

In order to confirm that the group arrangement was reliable and further look into the developmental pattern of learners' general competence when grouped into high and low onset level, we ran independent samples *t*-test between the two groups (all variables met the assumptions of normality, checked with K-S tests, and homogeneity of variance, checked with Levene's test) and one-way repeated measures ANOVAs with focused contrasts (Bonferroni-adjusted), to identify significant differences between data collection times⁴⁴. The results of independent-samples *t*-tests confirmed that low scorers on the rephrasing and cloze tests were inherently significantly different from high scorers at T1 [$t(28)=-7.452$, $p=.000$] and the results of the main ANOVA revealed a significant main effect of time on these scores. The effect size was rather large for both groups, $\eta^2 p=.70$, suggesting that both groups experienced significant overall improvement. Following Bonferroni-adjusted pairwise comparisons revealed that the low initial proficiency group experienced significant

⁴⁵ Data on the rephrasing and cloze tests were also collected at three testing times, facilitating a longitudinal approach to the exploration of L2 general English competence and thus allowing for the use of repeated-measures design with *post hoc* tests.

jumps in scores especially after the AH context (see Table 33 for the developmental pattern of each group).

Table 33. RM-ANOVAs: Longitudinal changes in grammar-cloze scores of learners arranged by high ($n=11$) and low ($n=19$) initial global proficiency. (all Fs = [2,18]).

Group	T1	T2	T3	F	p	η^2p	FI	SA
Low	5.52 (1.78)	7.13 (1.51)	7.38 (1.53)	21.530	.000	.70	.002	.065
High	10.04 (2.34)	11.71 (2.62)	12.18 (2.11)	5.021	.006	.70	.115	1.00

Shaded values indicate significance.

In light of these results, we went on to evaluate the impact of initial global proficiency on learners' productive vocabulary and determine whether there was a significant interaction between Time and Initial level. For that purpose, a mixed between-within subjects ANOVA with *Time* as the within-subjects factor and *Initial level* as the between-subjects factor was conducted to test changes in all lexical measures (all variable distributions were found to be normal across groups). However no significant interactions were found, suggesting that there were no systematic differences in the developmental patterns of each group, as explored through the composite rephrasing-cloze measure. We also tested for between-groups differences in high- and low-level groups submitting FI gains and SA gains to independent samples *t*-test with *Initial level* (based on the combined grammar-cloze score) as the grouping variable, so as to find out whether or not learners with different initial levels gained differently to a significant degree. However, we did not find any significant difference between the two groups.

In view of these results and finding no significant interactions, we cannot claim any advantage of any one group over the other. However, considering that rephrasing-cloze scores and lexical measures were found to be modestly correlated, as shown in Table 31 and Table 32, the lack of differences between groups differing in onset level may be simply attributable to a relatively loose relationship between general ability and vocabulary knowledge, which in turn represent different dimensions of L2 proficiency.

4.4. Summary

To summarise, we provide a brief overview of the findings surrounding the three main research questions that have shaped the present study and have been reported at length in Chapter 4. We follow the same order in which the results were presented.

In summing up our look at the longitudinal and context-specific (FI and SA) development of EFL lexicon in written and oral production, respectively, and in contrast with NS performance, and thus provide an answer to RQ1, we found mixed findings according to linguistic modality. In written production, significant improvement was found in the domain of fluency and lexical diversity, and occurred as a result of SA, also moving toward more native-like rates. In oral production, in contrast, significant changes occurred in lexical sophistication, yet took place after FI. The only measure in oral production that improved significantly after SA and converged with NS values was adverb density. Accuracy measures in both written and oral production improved gradually over time, with spelling errors

converging with NS standards after the SA period. In relation to NLSs, learners approached NS usage gradually, benefiting from both FI and SA, in terms of the use of impersonal forms in writing and greater lexical accuracy (e.g., false friends) in speech. After SA in particular, their writing also incorporated more idiomatic intensifiers (e.g., adverbs modifying adjectives), and their speech was richer in lexicalised fillers and target-like adverbs.

Regarding RQ2, which contrasted the amount of gains achieved in two different contexts (FI vs. SA) and modalities (writing vs. speech), we found that progress did not occur in a similar way for both modalities in either the FI or SA learning contexts. Thus, SA appeared to be more beneficial to learners' writing than FI, as seen in the progress made in lexical fluency and diversity, while FI, conversely, induced greater improvement in oral production, particularly in the domain of noun density and lexical sophistication. On the other hand, when the two modalities were contrasted, we were able to see that for writing greater gains took place after the SA period and corresponded to the area of fluency and lexical diversity. In oral mode, in contrast, larger gains accrued during the FI period and occurred in the domain of lexical diversity and sophistication.

Finally, our last research question (RQ3) inquired into the different patterns of vocabulary development and gains associated with learners' initial level. Results showed that the pattern of lexical development was indeed significantly different for students with low onset level, when the effect of initial level was assessed as a function of learners' T1 vocabulary knowledge in different lexical domains. Low-level

Results

learners showed a greater propensity towards gain than their high-level counterparts and were also able to catch up to high initial level participants after either FI or SA (or both). However, when we examined the effect of initial level based on learners' global proficiency through a combined grammar-cloze score, we were unable to detect any significant group differences and thus cannot claim any benefit for any one group over the other. Nonetheless, finding relatively modest correlations between vocabulary knowledge and general proficiency, as measured by the composite index, perhaps it is not surprising that the rephrasing and cloze tests failed to capture any difference between the two groups.

Chapter 5

Discussion

In the previous chapter, we presented the results obtained in relation to our three research questions originally outlined in Chapter 3. We examined written and oral lexical development through a composition and an interview task performed by 30 advanced EFL learners before and after FI and SA learning contexts, respectively. Along with learner data, we also considered oral and written productions by 27 native English speakers elicited through the same tasks, as the baseline data. In the present chapter, we further elaborate on the results obtained and discuss them with references to the literature reviewed in Chapter 2. This chapter is again organised around the three main research questions and following the same order as in the previous chapters.

5.1. Vocabulary development in oral and written production

We begin our discussion by addressing the first research question regarding EFL written and oral lexical development over time, after the FI and SA learning contexts, and as compared to NS baseline data through quantitative and qualitative analyses. More specifically, our first research question was formulated as follows:

RQ1) How does productive vocabulary by advanced EFL learners develop in each modality (writing and speech) over time, after two different learning contexts (FI and SA) and in contrast with NS production?

In order to answer this question, we looked at longitudinal progress after the consecutive FI and SA periods, on the one hand, and performed NS-NNS comparisons, on the other, as they required different statistical techniques or involved qualitative analysis. In the sections to follow, we review the main findings and consider how they compare with the relevant empirical literature in our discussion of the three sub-questions (RQ1a, RQ1b and RQ3c).

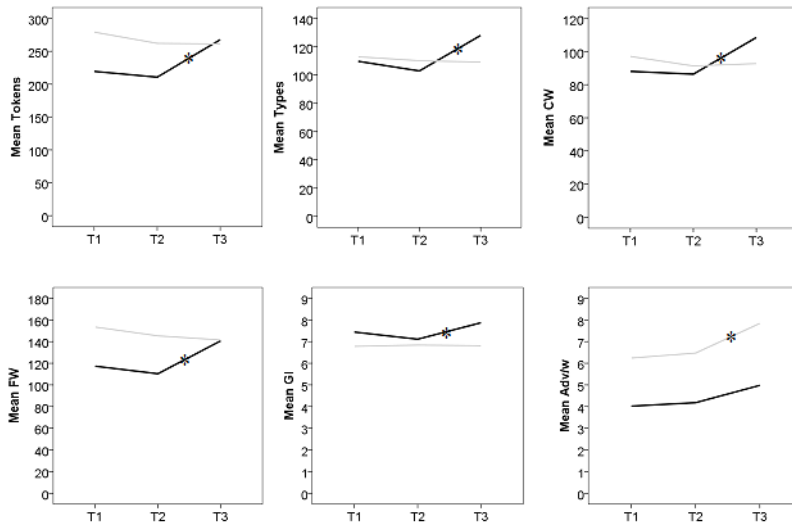
5.1.1. The effects of FI and SA on productive vocabulary

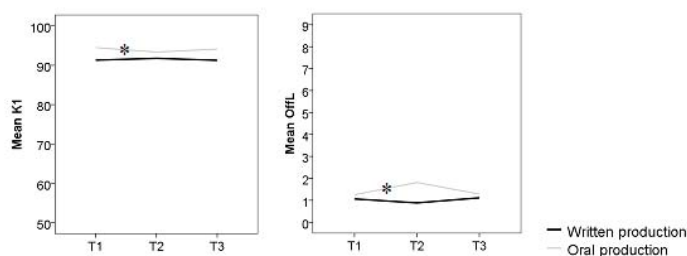
In response to sub-question RQ1a, *Does learners' written and oral productive vocabulary improve significantly over time, and after either the FI or SA learning contexts?*, we explored EFL vocabulary development in both written and oral productions by considering changes in the domains of lexical fluency, density, diversity, sophistication and accuracy. On the basis of the analysis reported in Chapter 4, we can offer an affirmative answer to this question: the 30 EFL learners' oral and written productive vocabulary *did* improve significantly over time, although the pattern of development for each modality and language context was somewhat different. As can be seen in a graphical summary of the developmental patterns in oral and written vocabulary (Figure 13), SA (from T2 to T3) proved to be highly beneficial for the development of learners' written productive vocabulary, specifically in the domain of

Discussion

fluency (Tokens, Types, CW, FW and all the constituents), and lexical diversity (GI), as well as for oral productive vocabulary development, although only in terms of adverb density (Adv/w). FI (from T1 to T2), on the contrary, had less bearing than SA and showed but a modest effect on improvement in oral productive vocabulary, especially as far as lexical sophistication (K1 and OffL) is concerned. In what follows, we begin our discussion with written production considering each domain in turn, as they all capture different aspects of lexical proficiency and merit detailed attention. We then discuss the results of the oral production data, following the order of the preceding chapter.

Figure 13. Line graphs depicting significant context-sensitive changes in written and oral productive vocabulary over the course of the study.





(*) indicates significant improvement.

Written production

In written production, regarding the domain of fluency, we found that learners' compositions increased in fluency over time, and that students produced significantly more tokens, types, content and function words, and word categories (nouns, adjectives, prepositions, etc.) at T3 than at T1. We also found that these changes were not linear, but rather U-shaped, as the mean scores for all fluency variables (except for adjectives) decreased after the FI period and increased again upon return from SA. When we evaluated these changes statistically, we found that the drop in fluency after FI was not significant but that the improvement in fluency after SA was. These results were in line with the findings from previous SALA-based studies assessing the quantity of written output in terms of words (Barquin, 2012; Perez-Vidal & Juan-Garau, 2011; Pérez-Vidal et al., 2012) and with the results from studies drawing on data from different populations (e.g., Sasaki, 2007; Golonka, 2006), which also found that SA students wrote significantly longer L2 texts, or produced a significantly greater number of word types. As far as word class frequencies are concerned, no study to our knowledge has applied this

Discussion

kind of analysis in SA contexts on written data. The only reference we have found is Collentine (2004). Although centred on oral conversational discourse captured through OPI, he reported that after SA students were able to generate more unique words (primarily adjectives and nouns) simply by virtue of speaking more fluently (i.e., producing more words per minute). On the basis of our own findings, it would seem as if spending a period abroad enables learners to produce more content in response to a given task, suggesting that they may access their lexical repertoire more easily (Wolfe-Quintero et al., 1998). Therefore, we can consider written fluency, as measured through types, tokens and word class frequencies, to be one of the lexical proficiency dimensions that flourishes in an SA environment.

In the domain of lexical density, the developmental trend showed minimal changes. We only found significant longitudinal changes in the variables of adverb, noun, and preposition density. The density of adverbs in learners' compositions substantially increased from T1 to T3, moving toward NS baseline rates. Noun density, on the contrary, significantly decreased over time (from 17.79 to 16.19) and this time did not represent an approximation to target-like language as they went in the opposite direction (18.80). One possible explanation for this drop is learners' increasing reliance on pronouns: although not statistically significant, learners showed a steady increase in pronoun density (from 9.73 to 9.94 to 10.52, over the course of the study), contrasting NS norms (7.10). The overuse of personal pronouns in our study was unsurprising, as previous literature has often found that L2 learners use substantially more pronouns than NSs (Barquin, 2012; Shaw & Liu, 1998; Silva, 1993). Finally, preposition density was

statistically significant from T1 to T3, with a significant drop from T1 to T2 during FI, again moving away from NS values. It is interesting to note that the co-occurrence of specific parts of speech in writing, predominantly nouns, adjectives and prepositions, has been found to be indicative of informationally rich or semantically dense discourse in the literature (Biber, 1988; Collentine, 2004). Thus, in the only previous study to consider semantic density in FI and SA contexts, Collentine (2004) found that the group following FI in an AH university tended to produce more informationally dense discourse than the SA group. In our case, both the density of nouns and prepositions appeared to decrease over time, suggesting that neither of the contexts were sufficiently fruitful to bring about the desired outcomes in these areas and generate a lexically denser writing. On the whole, lexical density measures did not appear to be meaningful descriptors of lexical proficiency in our corpus and may have lacked sensitivity to detect subtle differences when discriminating between similar profiles, an observation which was also made in previous research (Johansson, 2008; Laufer & Nation, 1995; Wolfe-Quintero et al., 1998).

In the domain of lexical diversity, as evaluated by Guiraud's Index (GI) and D, we have mixed results depending on how diversity was measured. On the one hand, the mean scores for diversity, as measured by GI, increased significantly over time and after the SA experience, suggesting that learners were using a greater variety of word types in their T3 essays and that it was the SA period that led to this improvement. The observation that written lexical diversity improves as a result of SA is in line with Barquin (2012) and Pérez-

Vidal et al. (2012), following an analogous sample of the SALA corpus, and Serrano et al. (2012)⁴⁶, who examine the effects of SA on learners studying overseas for longer periods (up to a full academic year). On the other hand, lexical diversity scores, as measured by D, revealed no significant improvement after either context, experiencing a slight drop after FI (from 66.57 to 62.73) and then increasing again after SA, to slightly superior initial T1 values (67.75). These results might have been interpreted as a lack of progress on the part of the learners, had we not included a NS baseline reference. With the inclusion of NS data, we were able to see that learners performed similarly to NSs (66.56) from the beginning of the study, implying that the D measure had a clear ceiling effect and was too broad to capture underlying NS-NNS differences. Furthermore, Meara and Miralpeix's (2008: 6) caveat that D may not perform satisfactorily at advanced levels⁴⁷, as the D values higher than 50 (typically produced by highly competent L2 learners) "are not so easy to distinguish from each other," appears to suggest that D was not a meaningful descriptor of written lexical diversity in our data. Therefore, here we feel more inclined to interpret the results based on the Guiraud's index of lexical diversity, following Van Hout and Vermeer's (2007: 114) observation that a square root formula like GI may be "a happy medium between doing nothing to the number of tokens (TTR) and applying too strong a transformation [...] that levels out all relevant differences."

⁴⁶ Although Serrano reported significant gains in lexical diversity (in terms of GI) in advanced EFL learners after a year abroad, she found no significant improvement for the same students in the same domain (GI) after the first three months of their sojourn.

⁴⁷ The fact that D does not always discriminate well between groups with obvious differences in vocabulary has been acknowledged in several studies (Daller et al., 2003, Van Hout and Vermeer, 2007, Jarvis, 2002).

In the domain of lexical sophistication, we observed no significant changes for either the LFP measures or the Greco-Latin Index (GLI). Although learners seemed to be using a slightly more sophisticated vocabulary in their essays by the end of the study (from 3.47 at T1 to 4.05 at T3 for the academic words (AWL), and from 1.06 at T1 to 1.11 at T3 for the off-list words (OffL)), they were still far below the values recorded for NSs (4.94 for AWL and 2.85 for OffL). This lack of improvement in lexical sophistication coincides with the results of Laufer and Paribakht (1998), who in their attempt to examine the effect of learning context on passive and active vocabularies, found that residence abroad was not conducive to better free productive vocabulary knowledge, as measured by the ‘beyond 2000’ index (i.e., the sum of percentages from the academic and off-list word lists in participants’ compositions) (see Laufer, 1995). They found that advanced students following classroom instruction at home (EFL learners in Israel) outperformed students learning L2 in an immersion context (ESL learners in Canada) in free productive vocabularies, but the latter had significantly larger passive vocabularies. The authors explained this paradox not as SA inefficiency to propel gains in the domain of lexical sophistication, but rather as new evidence for different developmental patterns of active and passive vocabulary knowledge in different language-learning contexts. Using a different sample of the SALA corpus, Barquin (2012) also reported nonsignificant results for lexical sophistication, as assessed via either the use of rare words (Advanced Guiraud 1000) or via noun or verb hyponymy. She attributed this lack of progress to the demanding nature of argumentative writing, which generally requires attention to many other details. In fact, Murphy and Roca de Larios (2010) report

that finding the right words to express the intended meaning in the L2 is particularly challenging for L2 learners, as the newly acquired lexis might be very difficult to access when restricted by the essay topic. Likewise, we did not register significant changes for GLI. The percentage of the words from Greco-Latin origin remained stable throughout the study with minimal variation (from 15.83 to 15.77 to 15.60) and fell well below the percentage recorded for NSs (20.15) that reflected rather literate and academic language, following Nippold's (2006) and Biber's (2006) features of academic written register. In one previous study to consider development in GLI, albeit not in an immersion context, Horst and Collins (2006) found that L1 French learners of English following intensive ESL programmes used fewer French cognates in their written narratives by the end of the study. However, one important difference between our study and Horst and Collins's is that their participants were 11-12-year-olds and had a beginner level of proficiency, whereas our participants were adult advanced learners specializing in foreign languages.

The fact that this index remains stable both after acquiring the language in AH and SA contexts makes evident that, on the one hand, the formal instruction received has not particularly focused on academic vocabulary and that, on the other, having been in contact with English abroad has not resulted in an increase of words from Anglo-Saxon origin. It could have been the case that students with L1 Romance languages, as the ones in this study, who may tend to use a high number of cognates, slowly replaced them by other Anglo-Saxon words (e.g., *career* for *degree*, *sociable* for *easy-going*, *metro* for *subway*) (see, however, Section 5.1.2. on native-like selections).

Finally, in the domain of lexical accuracy, we found that the number of errors (E/w) in the learners' compositions decreased significantly over time and after both FI and SA learning contexts. We also found that the proportion of lexical errors (LexE/w) dropped significantly from T1 to T3, as learners appeared to make fewer lexical mistakes gradually. Although these changes did not reach significance at either FI or SA, both learning contexts, experienced one after the other, caused a positive impact on learners' appropriateness of use, confirming the *Combination and Complementarity of Contexts Hypothesis* put forward in the SALA project (Pérez-Vidal & Juan-Garau, 2011). A marked improvement in lexical accuracy as a result of SA has also been reported in a study of a similar demographic by Llanes and Muñoz (2009). The authors provided evidence of more accurate oral productions in terms of significantly fewer lexical errors even after very short stays (3-4 weeks) abroad. As for spelling, no changes were found for either major (MjSp/w) or minor (MnSp/w) errors, despite moderate, albeit non-significant improvement in MjSp/w at T3 (.69), in the direction of NS norms (.34).

Oral production

As far as oral production is concerned, in the domain of fluency we observed quite a consistent lack of significant changes over time and after the FI and SA contexts. At first glance, the lack of progress in this domain was somewhat surprising, as oral fluency has been shown to highly benefit in SA contexts in a number of studies (e.g., Freed, 1995b; Llanes & Muñoz, 2009; Segalowitz & Freed, 2004; Serrano et al., 2012; Valls-Ferrer, 2011; Valls-Ferrer & Mora, 2014). However, the

discrepancies between the results of these studies and those of the current work could be attributed to the lack of uniform definition of what constitutes fluency and the different measures used to operationalise the construct in SLA research (Segalowitz, 2011). While most of the aforementioned studies have focused on temporal variables of fluency (e.g., speech rate, pauses or hesitation phenomena), as we discussed in Chapter 2, our measures were based on simple word counts (types, tokens and word class frequencies) without taking into account time restrictions. To the extent that fluency is defined as a mere quantity of speech and measured by the total number of words produced during the task irrespective of time, research is sparse and inconclusive. Thus, although our findings do not fully coincide with those of Freed and colleagues (2003)⁴⁸, they do fall in line with Segalowitz and Freed's (2004) study, where oral fluency is measured through similar metrics. Segalowitz and Freed also reported nonsignificant results for fluency, when this was quantified as the total number of words (i.e., tokens) produced in OPI, finding no advantage of SA over FI in this respect. Had we regarded fluency as fluidity or smoothness of speech and included a temporal-based measure, we may have obtained different results and may have been able to detect significant changes as a result of SA, based on previous research in this dimension (Pérez-Vidal et al. 2012; Valls-Ferrer, 2011; Valls-Ferrer & Mora, 2014). Another possible explanation for learners' lexical fluency remaining constant over the observation period may be attributed to their already advanced English proficiency at the start of

⁴⁸ As discussed in Chapter 2, in prior work to consider quantity of speech after SA contexts, Freed and colleagues (2003) reported a positive effect of SA experience on fluency, as SA students spoke significantly more and faster than their AH peers.

the study, which might have left little room for improvement (but see further discussion in Section 5.3.).

In the domain of oral lexical density, we did not register significant changes with the exception of adverb and conjunction density. The mean scores for adverb density (Adv/w) increased significantly over time in the direction of NS norms and this improvement occurred during the SA period and not during the FI period. The conjunction density, however, followed a different pattern and showed a significant jump after FI, while decreasing again significantly after SA to the initial T1 level. Although it is unlikely that a single variable may adequately depict the differential impact of the learning contexts on the acquisition of content and function words, a higher concentration of adverbs can be interpreted as a sign of improvement, considering that learners' speech was becoming more native-like by the end of the study, in tune with Biber's (1988) observation that adverbs are highly frequent in spoken and conversational discourse. Furthermore, following Bulté et al.'s (2008) argument that development in lexical proficiency may be related to "lexical growth in one or two content word classes only," it seems that it is the adverb class that has accounted for this development in the oral production data of our advanced learners (p. 10). Although we are unaware of any other study where significant improvement in lexical density was seen through gains in adverbs, Marsden and David (2008) also posit that the proportions of certain word classes change with learning and advancing in proficiency. Their research indicates that with increased language level learners produce more verbs than nouns in speech, and as they begin to use more verbs, they also incorporate more adjectives.

Discussion

With regards to lexical diversity, there were no significant changes for either GI or D measures. The mean scores registered for the two indices remained well below the level of NSs even at T3 (6.81 for GI and 48.63 for D, in comparison to the baseline rates of 7.57 and 63.26 recorded for NSs), with no advantage for SA over and above FI in this dimension of lexical proficiency. The lack of progress in oral lexical diversity, as measured by GI, was consistent with the previous findings of SALA-based studies, such as Pérez-Vidal and Juan-Garau (2011), Pérez-Vidal et al. (2012), Lara (2014), in which SA was not found to be particularly helpful for this domain. One exception is the study by Serrano et al. (2012), following a similar demographic, where significant improvement was observed for oral lexical diversity based on GI, as a result of a stay in the target language country in advanced level adult learners. It should be noted, however, that the sample they drew from was considerably smaller than ours (14 participants) and the instrument used to elicit oral speech was more cognitively demanding, as it consisted of an oral narrative task. Regarding the lack of improvement in D, our findings do not fully coincide with those by Foster (2009), where studying abroad cultivated a more enriched lexicon than following classroom instruction at home, and to a highly significant degree. Nonetheless, one notable difference between her study and ours is again the instrument: whereas Foster made use of oral narrative tasks to elicit L2 speech productions (i.e., cartoon picture prompts like in Serrano et al. (2012)), the task from which our data were compiled was a semi-guided informal interview. In this sense, a common observation in SLA research that task design and complexity can affect L2 performance in a number of ways (Gilabert, 2007; Robinson & Gilabert, 2007; Tavakoli & Foster, 2008) is very

likely to account for the different results between the aforementioned studies. Another possible interpretation that is worth mentioning here has to do with modality-related differences. Although there was a significant increase in lexical diversity in terms of GI in written production, suggesting that learners were using a more varied and less repetitive vocabulary, this improvement did not hold true for oral production. The lack of change in oral lexical diversity could be plausibly attributed to the inherent differences between oral and written production modes. That is, since learners do not have enough time to plan and monitor their speech and they are under pressure to retrieve lexical content more rapidly to keep up with the conversation, it is difficult for them to demonstrate the full breadth and range of their vocabularies when speaking, as opposed to writing.

In the domain of lexical sophistication, fewer words were used in K1, the first frequency band containing the 1,000 most frequent words in English, and more from OffL, the least frequent vocabulary. For the first time, this improvement reached significance specifically after FI and not after SA, as learners used a substantially lower percentage of K1 words and a significantly greater amount of sophisticated vocabulary (OffL) in their speech samples at T2 than at T1 or T3. This change also represented an approximation to native-like behaviour. A closer inspection of the off-list words used at T2 revealed a good proportion of lexis related to education (*high-school, homework, exam, semester, seminar, translator-interpreter, professor, curriculum, specialise*), everyday objects (*television, photocopy, printer, bookstores, certificate*), city places (*downtown, countryside, outskirts, surroundings*), or quality adjectives (*talkative, embarrassing, easy-going, shy*), to give some examples. The

Discussion

complete list of words pertaining to the off-list band in T2 interviews is organised per topic, and is reproduced in full in Appendix 9.

Although we lack comparative studies investigating the differential effects of learning contexts on lexical sophistication in oral production, Laufer and Paribakht's (1998) results, albeit for written expression, are likely the product of a similar process, as gains in sophisticated (academic + offlist) vocabulary by EFL FI group in their study surpassed those of the ESL SA group. That is, sometimes learners may acquire a more sophisticated vocabulary through form-focused instruction than through naturalistic exposure, as textbooks may contain more academic and advanced words than the input they may receive in an immersion setting. Just as in writing, we did not register significant changes for GLI in oral production. Likewise, the percentage of learners' GLI remained well below NS scores across the three data collection times, indicating that also in speech they relied less on the cognate words from Greco-Latin origin than their NS peers. It should be noted however that recent trends in vocabulary assessment (e.g., Bardel, Gudmundson, & Lindqvist, 2012) advocate for new methods that take into account additional factors other than from frequency (i.e., cognates or thematic vocabulary) when evaluating lexical sophistication. To put it another way, words such as *semester* or *seminar* perhaps should not be considered sophisticated if we extend its use to the academic context, in which they are rather highly frequent.

Finally, in the domain of accuracy, significant improvement was found for all measures (E/w and LexE/w) from T1 to T3, although further pairwise comparisons revealed that these changes were not attributable

to either FI or SA learning contexts. These findings fit well with prior work of Llanes and Muñoz (2009), who presented evidence of lexical accuracy development in oral production even after very short stays (3-4 weeks) abroad, with learners showing a significant decrease in the number of lexical errors produced by the end of the study. Although in our case the improvement occurred was not context-specific, together FI and SA were equally effective to cultivate a more accurate use of vocabulary, accumulating gains in both contexts, as indicated by the large effect sizes reported in Chapter 4.

5.1.2. Comparisons with native speakers

In order to gain a better understanding of all changes in all domains, both quantitatively and qualitatively, we employed NS baseline data as a point of reference. The use of NSs as the benchmark for assessing L2 development has been extensively denounced in SLA research (e.g., Cook, 1999; Ortega, 2014). In our study, we are aware of the negative connotations that may be associated with NS-NNS comparisons; however, we do not intend to portray L2 writing and speech in negative terms, as achieving a high competence in a foreign language gives us nothing but respect for our participants. Here we use NS baseline data for lack of a better means of comparison with learners at an advanced level, and under the assumption that they may provide special insight into L2 behaviour across time.

First, we considered quantitative differences in lexical proficiency in each domain, to respond RQ1b: *Does learners' written and oral productive vocabulary approximate native-like norms at any point over the observation period*

Discussion

in terms of quantitative lexical measures? Overall, the results showed that while written productive vocabulary showed a greater approximation to target-like norms than oral productive vocabulary, the convergence between learners and native speakers took place after SA and not FI in both production modes.

In written production, we found that learners converged with NSs in terms of fluency (Tokens, Types, CW, Adv, N, V, Prep), lexical diversity (GI), sophistication (AWL), and spelling (MjSp/w) after the SA context, but that the NSs remained significantly different from learners in accuracy and lexical density. The fact that T3 learners were no longer statistically distinguishable from NSs in most measures of fluency (except for adjective and pronoun frequencies) corroborated our finding that SA was especially fruitful for this domain in written productive vocabulary. This was particularly encouraging in light of all the previous research that has documented inherent NS-NNS distinctions in this area, with NNSs being less productive than NSs when composing a written text, and thus producing substantially shorter compositions (Silva, 1993). In his overview of research into L1-L2 differences, Silva also reported that L2 writers used fewer adjectives, prepositions and prepositional phrases but more pronouns and conjunctions, an observation that echoes our results, as there is quite a consistent lack of changes in the domain of lexical density. As for lexical richness, in view of ample evidence that L2 writing generally has less lexical variety, specificity, and sophistication than L1 writing – presumably due to a lack of lexical resources (Crossley & McNamara, 2009; Hinkel, 2003; Linnarud, 1986; Silva, 1993) – the convergence with NSs on the Guiraud's Index of lexical diversity (GI) and

academic vocabulary (AWL) indicated that improvement was, in fact, tangible in these areas and that learners' lexis was becoming more native-like precisely after SA. One final measure where NS-NNS differences disappeared was the number of major spelling errors per word (MjSp/w), being the only accuracy dimension that reflected target-like behaviour at T3, in tune with Barquin (2012). All in all, the convergence with NSs on these measures only at T3 is further evidence that a considerable progress in written productive vocabulary occurred after the SA period.

In oral production, NS-NNS differences were far more pronounced and learners remained significantly different from NSs across contexts in most domains: lexical diversity, sophistication, accuracy, and most density measures. These results came as no surprise, as despite SLA research lacking a knowledge base on quantitative L1-L2 comparisons in speech production (Read, 2000), as opposed to writing, learners reportedly demonstrate less lexical variety and sophistication overall (Crossley & McNamara, 2009). The only measure that did show NS-NNS convergence in our data was adverb density (Adv/w), as having differed significantly from native speakers at T1 and T2, learners no longer showed significant differences with them at T3. This indicated that SA was particularly beneficial in achieving greater adverb density and making learners adjust their speech to the target-like pattern, at least in this word class. In the absence of research on vocabulary development in different contexts of acquisition involving L1-L2 comparisons, our finding may only be related to a few studies. To provide an example, a SALA-based study by Trenchs-Parera (2009), examining dysfluency phenomena, reported significant increases in

lexical fillers after SA, making L2 speech appear lexically richer and more fluent. It is interesting to note that the lexical fillers included in Trenchs-Parera's study were of two types: lexicalised phrases (e.g., *I don't know, it's like, you know, and stuff*) and single words (e.g., *so, well, like*), the latter mostly represented by adverbs. The author argued that lexical fillers were used as helpful crutches in oral discourse, as happens with formulaic speech in language (Nattinger & DeCarrico, 1992), and that its growth could be related to the growth in lexical repertoire, as in Milton and Meara (1995). This is a position we also share when interpreting the increase in adverb density in our oral data.

Finally, we explored learners' vocabulary qualitatively in order to address RQ1c: *Does learners' written and oral productive vocabulary approximate native-like norms at any point over the observation period in terms of qualitative native-like selections?* The key observations that we were able to extract for each production mode respectively are discussed below.

In written production, we noted a slight tendency for learners to produce more general statements (through *it*-cleft constructions or the use of the pronoun *one*) by the end of the study. Yet, NS uses of these features were far more frequent than those of learners even at T3. Given that *it*-cleft is considered to be an advanced construction that marks the text for a formal register (McCarthy, 1994), and that the frequency rates of *it*-cleft in academic writing (Scollon, 1994, as cited in Hinkel, 2003) are particularly low in L2 texts, we were provided with further evidence that learners were writing more idiomatic and native-like texts. As for context-specific NLSs, after SA we noted that learners' writing incorporated more idiomatic intensifiers through the

use of adverbs modifying adjectives (e.g., *really hard*, *completely different*), adjusting their expressiveness to an NS pattern and confirming previous observations that natural exposure to language SA leads to a more native-like idiomaticity (Foster, 2009; Siyanova & Schmitt, 2008). Despite this improvement, learners, however, consistently relied on rather restricted lexical repertoire and tended to overuse basic vocabulary (e.g., *forget-lose*, *accept-follow*) unlike NSs, whose choices were more elaborate (e.g., *discard-forego*, *embrace-adjust*) and contained predominantly rare word types with numerous possible variants to alternate. This was consistent with a previous string of research that has pointed to lexical sophistication as a key difference between L1 and L2 writing (Hinkel, 2003; Shaw & Liu, 1998; Silva, 1993).

In oral production, we noticed that learners' word choices approached native-like usage gradually in terms of accuracy, as there were fewer collocations resulting from L1 transfer (e.g., *do lunch* instead of *have lunch*) or involving deceptive cognates (e.g., *career* instead of *degree*) after both FI and SA periods than at the beginning of the study. The latter observation corroborated Horst and Collins's (2006) finding that overreliance on cognates to compensate for lexical gaps might be replaced by the use of Anglo-Saxon words with proficiency, as seen in the significant decrease of French cognates in their Francophone ESL learners' productions. Regarding context-specific NLSs, we found that learners' speech was richer in lexicalised fillers (e.g., *I'd say* or *I mean*) and target-like adverbs (*really*, *actually*, *basically*) after SA than after FI, moving in the direction of NS choices. This finding fits well with the quite early SA-related work of Regan (1995) and Raupach (1984), in which SA was found to propel greater production of lexical fillers and

enhanced L2 fluency, replacing hesitations with formulaic expressions. The SALA publication by Trenchs (2009), as mentioned above, has also obtained similar results in these domains, finding a greater flow of L2 discourse upon return from a sojourn abroad, as elicited through the same interview as the one used in this study. In spite of these examples of approximation to target-like norms, learners' productive range of lexis was comparatively small and consisted largely of high-frequency general vocabulary even after SA, which was in tune with previous literature constantly arriving at this finding, even for proficient L2 learners. Although referring to writing skills at the time, Hinkel's (2003: 276) observation that "texts written by NNSs frequently rely on a limited lexical repertoire that results in vague and less sophisticated prose relative to that of NSs" seems to hold true in this case.

5.2. Vocabulary gains across contexts and modalities

Our second research question inquired into the amount of gains accrued for each learning context and production modality and read as follows:

RQ2) In which context (FI vs. SA) and modality (writing vs. speech) do learners accrue greater vocabulary gains?

We tackled this question in two ways: first, we focused on the comparison of the gains obtained after each learning context (FI vs. SA), and then we compared the gains achieved for each production modality (writing vs. speech). In what follows, we comment on the

results obtained and discuss them with references to previous works reviewed in Chapter 2.

5.2.1. On the learning context

In response to our first sub-question RQ2a, *Do vocabulary gains, if any, in written and oral production modality, respectively, accrue to a larger extent in one context than another, when comparing FI and SA?*, both contexts of acquisition, FI and SA, were shown to induce substantial gains on written and oral vocabulary, although the progress experienced by learners in each production modality differed. In fact, our results revealed that the context that proved to be the most beneficial for written productive vocabulary was SA, as greater gains were obtained in lexical fluency (Tokens, Types, CW, Adv, Nouns, Verbs, FW, Prep, Pron), preposition density (Prep/w) and lexical diversity (GI and D) at T3 than at T2. On the contrary, the context that led to the greatest progress in oral productive vocabulary was FI, as superior gains were found in lexical sophistication (K1 and OffL) and noun density (N/w) at T2 than at T3. Our results are similar to those of previous SALA-based studies, assessing the differential effects of learning contexts on learners' progress in oral and written skills. For example, Pérez-Vidal and colleagues (2012) found that progress did not take place in a similar way in both learning contexts for both modalities, and that oral lexical complexity improved at home and written lexical complexity abroad, while accuracy was the only domain that showed steady improvement. Like Pérez-Vidal et al., we may speculate that written productive vocabulary gains achieved as a result of the SA experience may be explained by the fact that during their time abroad, not only

Discussion

were learners attending lectures at the host university and submitting academic assignments, most of them were also writing personal SA diaries, at the request of their home institution. A practice of this kind in the TL community, offering greater intensity of exposure and massive opportunities for interaction, might have enhanced their lexical proficiency in the written modality already cultivated in the grammar-focused English classes. On the other hand, the positive impact of FI, particularly for oral lexical sophistication gains, can be attributed to the role of teaching methods. In other words, although speaking is typically not the most practised skill in a formal classroom setting, the teaching methods that prevailed at the students' home university, where English was the medium of instruction at all times, seemed to exert enough influence to improve their oral lexical proficiency. As noted elsewhere in the literature, form-focused instruction with literate textual support may enable a more sophisticated vocabulary use than exposure to high-frequency everyday vocabulary, generally offered in naturalistic settings. The observation that lexical sophistication may be especially amenable to progress in FI than in SA was also reported by Laufer and Paribakht (1998), who provided evidence for different developmental patterns of vocabulary in different language-learning contexts. Taken together, our results seem to support Hokansson's (2000) argument that the skills most practised in one context are not always the ones to improve the most. Ultimately, our findings also suggest that both the FI and SA learning contexts may complement each other, as seen in the domain of accuracy, which can be tentatively explained along the lines of DeKeyser's (2007) tenet that "there can be knowledge transfer from one learning context to another," a view further taken into

consideration in the SALA project's *Combination and Complementarity of Contexts Hypothesis* (Pérez-Vidal et al., 2012: 230, Pérez-Vidal, 2014b). We will further discuss the nuances of each production mode in the following section.

5.2.2. On the production modality

After answering RQ2a, we proceeded to the second set of comparisons in order to answer RQ2b, *Do vocabulary gains, if any, in FI and SA, respectively, accrue to a larger extent in one modality than another, when comparing written and oral production?* When the two modalities were contrasted, we found that the gains achieved during FI were significantly greater in speaking than in writing, as seen in the domain of lexical sophistication (K1 and OffL), while in SA the gains were significantly greater in writing than in speaking, as seen in fluency (Tokens, Types, CW, N, V, FW, Conj, Prep, Pron) and lexical diversity (GI). This finding was again consistent with Pérez-Vidal et al. (2012) and can be explained along the same lines as the RQ2a results just discussed. On the other hand, greater progress in lexical fluency and diversity in writing than in speech may have also been due to the task planning factor, as the opportunity for planning, reviewing and editing required to achieve precision in the choice of lexis is only afforded in the written mode (Biber, 2009; Williams, 2012). As discussed in Chapter 2, the different processing constraints and communicative conditions associated with each mode of language production (writing vs. speech) have a marked effect on the lexicon used in texts differing in modalities. Following Strömquist et al. (2012: 47), re-using the same words in a conversation may be just a natural strategy for learners “to

reduce the cognitive load imposed by the on-line constraints of spoken language,” since a wider range of vocabulary is only available to them in writing, where the on-line constraints are relaxed.

Finally, greater gains in the domain of lexical sophistication in oral production than in written production can be interpreted as the product of exposure to classroom and degree-related vocabulary (e.g., *high-school, homework, exam, translator, interpreter, professor*, etc., as can be seen in Appendix 9), which propelled growth of lexical items of this type. As we are unaware of any other study to document changes in lexical sophistication in the spoken versus written modes across different contexts of acquisition, further research in this domain would not be misplaced.

Although very few studies have examined the effects of SA on lexical development in the two modes of production, Serrano et al.’s (2012) study is one of the few to find significant improvement in lexical diversity in speaking and no parallel progress in writing at the end of a semester abroad. However, their results are not easily comparable to ours due to the differences in the sample size and the task used, which demonstrates that a contrast of modalities in SLA and SA research is an area worth further investigation.

5.3. The role of initial level

Finally, our last research question was dedicated to measuring the impact of initial level on vocabulary development and gains across the contexts and modalities and was posed as follows:

RQ3) Are there different patterns of vocabulary development and gains associated with learners' onset level?

We found that the patterns of development and the gains obtained *were* different, although the impact of differences depended on the ways in which we measured onset level in the first place. The nuances of these findings are further discussed in the subsections that follow.

5.3.1. Initial vocabulary knowledge

In relation to RQ3a, *Are there different patterns of vocabulary development and sizes of gains associated with learners' onset level, as measured through initial vocabulary knowledge?*, we can offer a positive answer, as groups with different initial vocabularies did show different patterns of development and amount of gains over the two contexts. More specifically, learners with lower initial levels showed a greater propensity towards gain than more advanced students and caught up with the latter after either FI or SA (or both) in all measures (except for E/w, AWL, GI in writing, and E/w and D in speaking). This finding matches up well with the *Threshold Hypothesis* presented in Chapter 2, in that a certain threshold level of proficiency is required for learners to fully benefit from the opportunities for SLA in SA contexts, but that once learners are over this threshold, higher level learners will improve relatively less due to the normal learning curve. Following Carroll (1967), “many low-aptitude students are able to compensate by *diligent study and practice* or because of special opportunities such as *study abroad*” (p. 1, emphasis added). In our study, the diligent study and practice provided during FI served as a

levelling experience between learners of different proficiency levels, as low-level learners showed marked improvement already at T2, while advanced learners either plateaued or even regressed to only improve at T3 after the SA experience (see Chapter 4, Figures 11 and 12). At the level of vocabulary, the studies by Milton and Meara (1995) and Llanes and Muñoz (2009) pointed to similar results as those obtained here: low-level learners tend to gain more when compared to more proficient learners. Both studies provided empirical evidence that learners with lower initial levels benefited more from SA, showing greater propensity toward gain in vocabulary growth (Milton & Meara, 1995) and lexical accuracy (Llanes & Muñoz, 2009), than their more advanced peers.

5.3.2. Initial general proficiency

Finally, in reply to RQ2b, *Are there different patterns of vocabulary development and gains associated with learners' onset level, as measured through initial global proficiency?*, we are more likely to answer *no*, as we did not find any compelling evidence to suggest the contrary. That is, when we examined the effect of initial level through learners' initial global proficiency, as assessed by a combined rephrasing-cloze score, we did not detect any significant group differences, finding no benefit for any one group over the other. Our lack of significant results where initial global proficiency is concerned is consistent with those of Barquin (2012) and Lara (2014). Although Barquin (2012) did find a significant impact of initial grammatical proficiency on the perceived quality of L2 essays, no systematic differences were found on the quantitative textual characteristics of writing, part of which involved analysis of

lexical diversity and sophistication. The composite score on the SALA grammar and cloze tests was neither a significant predictor of gains in speech production elicited through a role-play task, in Lara's (2014) study. Taking the results altogether and considering that the grammar-cloze scores and the lexical proficiency measures were found to be only modestly correlated, our results seem to confirm that general linguistic ability and vocabulary knowledge are overlapping but separate competences (Chapelle, 1994; Read, 2000: 101), and that the cloze is not a truly lexical measure; a definitive understanding of the nature of this relationship goes beyond the scope of the present study and is left aside to be addressed in further research.

5.4. Summary

In this chapter we have discussed our main findings, and how they fit with the background literature. We commented on the effects of the FI and SA learning contexts on vocabulary development in writing and speech for each lexical domain in quantitative and qualitative terms, and offered some possible explanations for the changes occurred. Our results were consistent with DeKeyser (2007), and Pérez-Vidal and Juan-Garau's (2011) hypothesis that both contexts of acquisition – FI and SA – are complementary in the process of L2 learning, although the patterns of development varied across the context and modality. With reference to the context and modality, we also provided different reasons for the significant changes (or lack thereof) observed. Thus, the different teaching methods specific to each context as well as the modality-related characteristics in terms of information-processing constraints are some of the explanations that

may have accounted for our findings. We concluded our discussion of the results by commenting on the impact of initial level when it was measured using two different approaches. Our findings, when approaching initial level as a function of learners' initial vocabulary knowledge, added support to the *Threshold Hypothesis*, as low-level learners eventually experienced greater gains than their more proficient counterparts. Nonetheless, this trend did not hold true for initial level, when this was assessed on the basis of general language proficiency, albeit being consistent with previous SALA-based research (Barquin, 2012; Lara, 2014).

We discussed our findings in relation to previous research with special emphasis on studies with a similar population and tapping into change taken place during SA periods in English-speaking countries. In sum, our results mesh well with most previous research efforts, especially those pertaining to the SALA project (Barquin, 2012; Lara, 2014; Pérez-Vidal et al., 2012), in that the progress experienced by learners after both the FI and SA in oral and written productive vocabulary differed (Pérez-Vidal et al., 2012) and that initial vocabulary knowledge, unlike global initial proficiency (Barquin, 2012; Lara, 2014), was robustly impactful on ultimate gains and maximisation of the SA period. In the following final chapter, we outline some conclusions from the findings we have discussed at length here. We also recognise some limitations to this study and suggest potential avenues for future research.

Chapter 6

Conclusions and future research

In this dissertation, we have attempted to shed light on the impact of a 3-month SA period experienced after a previous period of FI on vocabulary acquisition in oral and written production, an area of research which has to date received little coverage in the field of SA and SLA. To that end, a combination of quantitative and qualitative metrics was employed targeting different aspects of lexical proficiency: fluency, density, diversity, sophistication, and accuracy, and L2 idiomaticity through native-like selections. To our knowledge, this is the first study to systematically address the differential effects of FI and SA experiences on vocabulary in both writing and speech samples produced by the same participants. An approach of this nature represents an important contribution to the SA research as it offers a more complete picture of the type of changes that different learning contexts may trigger in oral and written production at the lexical level.

The data presented in this study are also valuable for the field of lexical acquisition: apart from a couple of research papers published some years ago (Milton & Meara, 1995; Ife et al., 2000), the very few studies available have concentrated on very specific aspects of

vocabulary learning (e.g., the acquisition of meaning of two verbs or prepositions; as in Ryan & Lafford, 1992 and Lafford & Ryan, 1995; or lexical diversity; as in Foster, 2009; or word-associations; as in Fitzpatrick, 2012). In addition, quite often the number of participants in these studies is not high (e.g., Fitzpatrick, 2012 is a case study) or the groups compared are totally independent (e.g., Foster 2009 analyses a group in Teheran and a group in London). This often makes it difficult to see ‘the broad picture’ of how lexis develops in different learning environments. Analysing the vocabulary produced by exactly the same participants in different contexts can help us to better gauge the pros and cons of different lexical environments regarding learners’ output. This may, in turn, be particularly helpful in determining what is most needed in different contexts to ensure that learners benefit from the experiences and improve their vocabulary knowledge. In addition, findings from this dissertation can be compared in the future with ongoing research on lexical acquisition in virtual environments (Milton et al., 2012), a ‘halfway solution’ between formal instruction and immersion to expose learners to real language use.

Our findings in relation to the effects of two consecutive learning contexts, FI and SA, on lexical proficiency in writing and speech have provided evidence for different developmental patterns depending on the learning context and the production modality under scrutiny. In written production, we found that learners’ vocabulary improved significantly over the observation period and that this improvement occurred predominantly after the SA. The SA learning context (i.e., a rich language environment that presumably provides intense exposure

to TL input and massive opportunities for practice and interaction), has been found to have the potential to enhance L2 learners' fluency and lexical diversity. Furthermore, during SA we noted a number of significant trends toward NS values in terms of academic vocabulary, spelling, fluency, and diversity and a greater number of native-like selections. In oral production, longitudinal progress was less pronounced and affected lexical sophistication, which improved after FI, and adverb density, which increased after SA to converge with NS values. The only domain that showed significant improvement irrespective of the context or modality was that of lexical accuracy, as learners' written and oral vocabulary was much more accurate and contained consistently fewer false cognates, as seen through NLSs.

As regards the comparisons of gains as function of learning contexts and modalities, our findings support two claims: 1) both contexts may supplement each other (i.e., as seen in the domain of accuracy), lending support to DeKeyser's (2007) and Perez-Vidal (2014b) combination hypothesis, although with some exception which leads us to the following point; 2) while it is fair to say that SA benefits written productive vocabulary to a greater extent than oral productive vocabulary, lexical sophistication may be enhanced, conversely, with classroom teaching. However, the latter observation may also be considered as an artefact of the modality, in that the cognitive constraints and online processing factors of spoken communication may simply not allow L2 learners to choose a more sophisticated word, or use synonyms to vary their vocabulary, as compared to the written mode, which generally entails better opportunities to plan and review production, as well as more generous time allowance.

As for the role of initial level, our results pointed to an advantage for those students who had poorer vocabulary knowledge at the beginning of the study over those who were at a higher level, in accordance with the normal learning curve phenomenon identified in previous vocabulary-related studies in the SA context (Llanes & Muñoz, 2009; Milton & Meara, 1995). However, we did not find any compelling evidence for the same trend when we examined learners' initial level as a function of their general English proficiency.

The present study is not without its limitations and, as is common in empirical research, these should be acknowledged and considered when planning future studies in this area. One noteworthy shortcoming, for instance, derived from methodology, and more specifically, the task employed. We remember from Chapter 4 that the correlational analysis of gains between written and oral production modes revealed a strong positive relationship between the two modalities at least in lexical diversity and accuracy. However, had we used the same task topic to elicit oral and written vocabulary, the comparability of the two modes would have been more rigorous and effective, and any task effect would have been avoided. That is, had we not been limited by the data available in the SALA corpus, making students speak and write about their *University life* or having them discuss the same essay prompt in the oral interview might have put our results in a different light. In a similar vein, an additional test on receptive or organisational vocabulary knowledge, such as the ones used in Milton and Meara (1995) or Ife and colleagues (2000), or Foster's (2014) receptive NLS knowledge test, would have been an ideal supplement to add to our knowledge of the relationship between

receptive-productive mastery of vocabulary, discussed in the literature review. A tool of this nature would have provided more information on testees' lexical knowledge, as in free production learners may use lexical avoidance strategies to cover their vocabulary knowledge gaps.

Another limitation we would like to highlight here is related to the analysis conducted. As reviewed earlier, oral fluency has often stood out as the main winner with regard to other linguistic skills when examining the value of SA in the advancement of L2 competence, yet such findings generally referred to the temporal aspects of fluency. Due to time constraints, the inclusion of fluency measures tapping into how rapidly speech was delivered (i.e., also an indication of how quickly learners access the vocabulary they use) was not possible. Nevertheless, on the basis of previous SALA findings on temporal fluency in favour of SA (Valls-Ferrer, 2012; Pérez-Vidal, 2012; Lara, 2014), it would be highly likely to obtain similar results and thus provide little novel information regarding the effects of SA in this respect.

To sum up, although improvement in some of the aforementioned aspects will have to be left for further research, we believe that our contribution has underscored the relevance of studying vocabulary acquisition across contexts and modalities. The most referenced works on SA and vocabulary acquisition have often looked at the knowledge of form-meaning connections through specific vocabulary tests (e.g., Ife, Vives Boix, & Meara, 2000; Milton & Meara, 1995). However, perhaps in addition to testing learners' ability to understand word meanings or exploring their lexical organisation, knowing what types

of words they use or how repetitive and accurate these word choices are may give us some further indications of what to expect from an SA period in an L2 productive lexical proficiency. That is, if learners are expected to speak and write in more appropriate, native-like ways after an SA experience, we should then investigate if this is actually the case, and how (and why) their post-SA production may differ from the output they produce in FI contexts. An account of what happens in learners' written and oral productive vocabulary after a 3-month SA period is what we have sought to cover in this study with samples of real language use.

Additionally, while completing the present study, we were also able to add to our understanding of a number of quantitative and qualitative data-driven measures testing different facets of lexical proficiency. We believe that our multi-measures methodology triangulating results from “more than one approach in order to achieve more robust findings” (Schmitt, 2010: 149) may have an additional value and should prove useful for future studies undertaking similar investigations.

Future research

Earlier, in Chapter 2, we discussed how formulaic language has acquired growing acceptance within the SLA community. We have seen that being immersed in an L2 environment that is abundant in rich native input may develop more efficiently networked lexicon and greater idiomaticity (Siyanova & Schmitt, 2008; Foster, 2009). Given this line of reasoning, it would be interesting to go beyond the

descriptive account of learners' word choices, as we have done in this study, and explore L2 formulaic language quantitatively. Computing the formulas by checking them against available corpora (e.g., the 100 million word sample of the British National Corpus, BNC), with Wray and Namba's (2003) criteria as a guide, would be an interesting step to take in the future.

In the development of this study, we noted that some measures were not particularly informative and did not give us a fully satisfying account of the changes occurred. For example, it came to our notice that lexical sophistication measures according to LFP, which follows a purely frequency-based approach, were not as sensitive as expected. Given that a few recent studies have succeeded in fine-tuning these measures by recategorizing low-frequency vocabulary based on teachers' perceptions (see Bardel et al., 2012; Daller et al., 2003), an approach of this type could also provide a more fine-grained account of progress in our data and should not be dismissed. Similarly, we are aware of the new measures and tools that are being developed for researching vocabulary (Meara & Miralpeix, in press); this might also be another potentially interesting and practical issue to be addressed at a later date.

Finally, considering that there is a plethora of factors that may dictate the ultimate success of the SA period in furthering language learning, future research might also combine quantitative and qualitative methods. Most programme variables and individual factors are notoriously difficult to control for in immersion contexts, as no matter how many questionnaires are administered to the informants, little can

be captured via self-reports alone. Furthermore, the role of new technologies on language learning since the turn of the present century should not be underestimated either, as the presence of both L1 and L2, not to mention other languages, can be a bit of a mixed blessing. That is, just as learners can easily immerse themselves in the target culture and absorb the L2 through podcasts or YouTube without leaving their homes, they can likewise access online press, Skype or Facebook if they want to keep close ties with their L1 community when studying abroad. In this sense, DeKeyser's (2014) proposal for mixed-method research, which combines quantitative and qualitative data, including in-depth documentation of students' activities, or the quantity and quality of their interaction, could lay the groundwork for future work in order to further elucidate previous findings.

References

- Agustín Llach, M.P. (2006). Lexical errors in young EFL learners: How do they relate to proficiency measures? *Interlingüística*, 17, 63-73.
- Agustín Llach, M.P. (2011). *Lexical errors and accuracy in foreign language writing*. Bristol: Multilingual Matters.
- Alderson, J.C. (2005). *Diagnosing foreign language proficiency: The interface between learning and assessment*. London: Continuum.
- Anderson, J.R. (1993). *Rules of the mind*. Hillsdale: Erlbaum.
- Anderson, R.C., & Freebody, P. (1981). Vocabulary knowledge. In J.T. Guthrie (Ed.), *Comprehension and Teaching: Research Reviews* (pp. 77-110). Newark: International Reading Association.
- Anderson, S.L. (2014). Individual differences & study abroad: Four profiles of oral proficiency gain. *Journal of Language Teaching and Research*, 5(3), 477-486.
- Arnaud, P. (1984). The lexical richness of L2 written productions and the validity of vocabulary tests. In T. Culhane, C. Klein-Braley & D.K. Stevenson (Eds.), *Practice and problems in language testing* (pp. 14-28). Colchester: University of Essex.
- Arnaud, P. (1992). Objective lexical and grammatical characteristics of L2 written compositions and the validity of separate-component tests. In P. Arnaud & H. Béjoint (Eds.), *Vocabulary and applied linguistics* (pp. 133-145). London: Macmillan.
- Avello, P. (2013). *L2 phonological development in speech production during study abroad*. (Doctoral dissertation, Universitat Pompeu Fabra, Spain). Retrieved from <http://www.tdx.cat/handle/10803/128624>.
- Bardel, C., Gudmundson, A., & Lindqvist, C. (2012). Aspects of lexical sophistication in advanced learners' oral production: Vocabulary acquisition and use in L2 French and Italian. *Studies in Second Language Acquisition*, 34(2), 269-290.

References

- Barquin, E. (2012). *Writing development in a study abroad context*. (Doctoral dissertation, Universitat Pompeu Fabra, Spain). Retrieved from <http://www.tdx.cat/handle/10803/96196>.
- Barron, A. (2006). Learning to say 'you' in German: The acquisition of sociolinguistic competence in a study abroad context. In M.A. DuFon & E. Churchill (Eds.), *Language learners in study abroad contexts* (pp. 59-90). Clevedon: Multilingual Matters.
- Batstone, R. (2002). Contexts of engagement: A discourse perspective on intake and pushed output. *System*, 30(1), 1-14.
- Beattie, J. (2008). Evaluating the acquisition of English listening comprehension skills in the framework of the SALA Project. *Monografías de Aula Abierta*, 32, 301-310. Oviedo: Universidad de Oviedo.
- Beattie, J. (2014). The 'ins and outs' of a study abroad programme: The SALA exchange programme. In C. Pérez-Vidal (Ed.), *Language acquisition in study abroad and formal instruction contexts* (pp. 59-84). Amsterdam: John Benjamins.
- Berman, R.A., & Nir, B. (2010). The lexicon in writing–speech differentiation: Developmental perspectives. *Written Languages & Literacy*, 13(2), 183-205.
- Berman, R.A., & Verhoeven, L. (2002). Developing text-production abilities across languages, genre, and modality. *Written Languages & Literacy*, 5(1), 1-44.
- Biber, D. (1988). *Variation across speech and writing*. New York: Cambridge University Press.
- Biber, D. (2006). *University language: A corpus-based study of spoken and written registers*. Philadelphia: John Benjamins.
- Biber, D. (2009). Are there linguistic consequences of literacy? Comparing the potentials of language use in speech and writing. In D.R. Olson & N. Torrance (Eds.), *The Cambridge handbook of literacy* (pp. 75-92). Cambridge: Cambridge University Press.

References

- Brecht, R., Davidson, D., & Ginsberg, R. (1990). The empirical study of proficiency gain in study abroad environments of American students of Russian. In D. Davidson (Ed.), *American contributions to the VII International Congress of MAPRLAL*. Washington, DC: American Council of Teachers of Russian.
- Broeder, P., Extra, G., & van Hout, R. (1993). Richness and variety in the developing lexicon. In C. Perdue (Ed.), *Adult language acquisition: Cross-linguistic perspectives* (pp. 145-163). Cambridge: Cambridge University Press.
- Broeder, P., Extra, G., van Hout, R., Strömqvist, S., & Voinmaa, K. (Eds.). (1988). *Processes in the developing lexicon: Final report to the European Science Foundation*. Strasbourg, Tilburg and Göteborg: European Science Foundation.
- Bulté, B., Housen, A., Pierrard, M., & Van Daele, S. (2008). Investigating lexical proficiency development over time: The case of Dutch-speaking learners of French in Brussels. *Journal of French Language Studies*, 18, 277-298.
- Carroll, J. (1967). Foreign language proficiency levels attained by language majors near graduation from college. *Foreign Language Annals*, 1(1), 131-151.
- Carter, R. (2012). *Vocabulary: Applied linguistic perspectives* (3rd ed.). London: Routledge.
- Chapelle, C.A. (1994). Are C-tests valid measures for L2 vocabulary research? *Second Language Research*, 10(2), 157-187.
- Churchill, E., & DuFon, M.A. (2006). Evolving threads in study abroad research. In M.A. DuFon & E. Churchill (Eds.), *Language learners in study abroad contexts* (pp. 1-27). Clevedon: Multilingual Matters.
- Clark, E. (1993). *The lexicon in acquisition*. Cambridge: Cambridge University Press.

References

- Cobb, T. (2003). Analyzing late interlanguage with learner corpora: Québec replications of three European studies. *The Canadian Modern Language Review*, 59(3), 393-423.
- Cobb, T. (2016). Web Vocabprofile [Online programme]. Retrieved from <http://www.lex tutor.ca/vp/>.
- Coleman, J.A. (1996). *Studying languages: A survey of British and European students: the proficiency, background, attitudes and motivations of students of foreign languages in the United Kingdom and Europe*. London: CILT.
- Coleman, J.A. (1997). Residence abroad within language study. *Language Teaching*, 30, 1-20.
- Coleman, J.A. (1998). Language learning and study abroad: the European perspective. *Frontiers*, 4, 167-203.
- Coleman, J.A. (2005). Residence abroad. In J.A. Coleman, & J. Klapper (Eds.), *Effective learning and teaching in modern languages* (pp. 126-132). London: Routledge.
- Coleman, J.A. (2013). Researching whole people and whole lives. In C. Kinginger (Ed.), *Social and cultural aspects of language learning in study abroad* (pp. 17-46). Amsterdam: John Benjamins.
- Collentine, J. (2004). The effects of learning context on morphosyntactic and lexical development. *Studies in Second Language Acquisition*, 26(2), 227-248.
- 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). New York: Wiley.
- Collentine, J., & Freed, B. (2004). Learning context and its effects on second language acquisition. *Studies in Second Language Acquisition*, 26(2), 153-171.
- Cook, V. (1999). Going beyond the native speaker in language teaching. *TESOL Quarterly*, 33(2), 185-210.

References

- Crossley, S.A., & McNamara, D.S. (2009). Computational assessment of lexical differences in L1 and L2 writing. *Journal of Second Language Writing*, 18, 119-135.
- Cubillos, J., Chieffo, L., & Fan, C. (2008). The impact of short-term study abroad programs on L2 listening comprehension skills. *Foreign Language Annals*, 41(1), 157-185.
- Daller, H., Milton, J., & Treffers-Daller, J. (Eds.). (2007). *Modelling and assessing vocabulary knowledge*. Cambridge: Cambridge University Press.
- Daller, H., van Hout, R., & Treffers-Daller, J. (2003). Lexical richness in the spontaneous speech of bilinguals. *Applied Linguistics*, 24(2), 197-222.
- Daller, H., & Xue, J. (2007). Lexical richness and the oral proficiency of Chinese EFL students. In H. Daller, J. Milton & J. Treffers-Daller (Eds.), *Modelling and assessing vocabulary knowledge* (pp. 150-164). Cambridge: Cambridge University Press.
- David, A., Myles, F., Rogers, V., & Rule, S. (2009). Lexical development in instructed L2 learners of French: Is there a relationship with morphosyntactic development? In B. Richards, D. Malvern, P.M. Meara, J. Milton & J. Treffers-Daller (Eds.), *Vocabulary studies in first and second language acquisition* (pp. 147-163). Basingstoke: Palgrave Macmillan.
- Davie, J. (1996). Language skills, course development and the year abroad. *The Language Learning Journal*, 13(1), 73-76.
- DeKeyser, R. (1990). From learning to acquisition? Monitoring in the classroom and abroad. *Hispania*, 73(1), 238-247.
- DeKeyser, R. (1991). Foreign language development during a semester abroad. In B. Freed (Ed.), *Foreign language acquisition research and the classroom* (pp. 104-119). Lexington: D.C. Heath.
- DeKeyser, R. (2007). Study abroad as foreign language practice. In R. DeKeyser (Ed.), *Practice in a second language: Perspectives from applied*

References

- linguistics and cognitive psychology* (pp. 208-226). Cambridge: Cambridge University Press.
- DeKeyser, R. (2010). Monitoring processes in Spanish as a second language during a study abroad program. *Foreign Language Annals*, 43(1), 80-92.
- DeKeyser, R. (2014). Research on language development during study abroad: Methodological considerations and future perspectives. In C. Pérez-Vidal (Ed.), *Language acquisition in study abroad and formal instruction contexts* (pp. 313-325). Amsterdam: John Benjamins.
- Dewaele, J.M. (2004). Vous or tu? Native and non-native speakers of French on sociolinguistic tightrope. *IRAL*, 42(4), 383-402.
- Dewaele, J.M., & Pavlenko, A. (2003). Productivity and lexical diversity in native and non-native speech: A study of cross-cultural effects. In V. Cook (Ed.), *Effects of the second language on the first* (pp. 120-141). Clevedon: Multilingual Matters.
- Dewey, D. (2004). A comparison of reading development by learners of Japanese in intensive domestic immersion and study abroad contexts. *Studies in Second Language Acquisition*, 26(2), 303-327.
- Dewey, D. (2008). Japanese vocabulary acquisition by learners in three contexts. *Frontiers*, 15, 127-148.
- Díaz-Campos, M. (2004). Context of learning in the acquisition of Spanish second language phonology. *Studies in Second Language Acquisition*, 26(2), 249-273.
- Dóczy, B., & Kormos, J. (2015). *Longitudinal developments in vocabulary knowledge and lexical organization*. New York: Oxford University Press.
- Dörnyei, Z., Durow, V., & Zahran, K. (2004). Individual differences and their effects on formulaic sequence acquisition. In N. Schmitt (Ed.), *Formulaic sequences* (pp. 87-106). Amsterdam: John Benjamins.

References

- Dörnyei, Z., & Skehan, P. (2003). Individual differences in second language learning. In C.J. Doughty & M.H. Long (Eds.), *The handbook of second language acquisition* (pp. 589-630). Oxford: Blackwell.
- DuFon, M.A., & Churchill, E. (Ed). (2006). *Language learners in study abroad contexts*. Clevedon: Multilingual Matters.
- Ellis, R. (1994). *The study of second language acquisition*. Oxford: Oxford University Press.
- Ellis, R. (2012). *Language teaching research and language pedagogy*. Malden: Wiley-Blackwell.
- Engber, C.A. (1995). The relationship of lexical proficiency to the quality of ESL compositions. *Journal of Second Language Writing*, 4, 139-155.
- Erman, B., & Warren, B. (2000). The idiom principle and the open choice principle. *Text*, 20(1), 29-62.
- Fitzpatrick, T. (2012). Tracking the changes: Vocabulary acquisition in the study abroad context. *The Language Learning Journal*, 40(1), 81-98.
- Fitzpatrick, T., & Barfield, A. (2009). *Lexical processing in second language learners*. Clevedon: Multilingual Matters.
- Foster, P. (2001). Rules and routines: A consideration of their role in the task-based language production of native and non-native speakers. In M. Bygate, P. Skehan & M. Swain (Eds.), *Researching pedagogic tasks: Second language learning, teaching, and testing* (pp. 75-94). Harlow: Longman.
- Foster, P. (2009). Lexical diversity and native-like selection: The bonus of studying abroad. In B. Richards, H. Daller, D. Malvern, P.M. Meara, J. Milton & J. Treffers-Daller (Eds.), *Vocabulary studies in first and second language acquisition* (pp. 91-106). Hampshire: Palgrave Macmillan.

References

- Foster, P., Bolibaug, C., & Kotula, A. (2014). Knowledge of nativelike selections in a L2: The influence of exposure, memory, age of onset, and motivation in foreign language and immersion settings. *Studies in Second Language Acquisition*, 36(1), 101-132.
- Foster, P., & Skehan, P. (1996). The influence of planning on performance in task-based learning. *Studies in Second Language Acquisition*, 18(3), 299-324.
- Foster, P., & Tavakoli, P. (2009). Native speakers and task performance: Comparing task effects on complexity, fluency and lexical diversity. *Language Learning*, 59(4), 866-896.
- Fotos, S. (1991). The cloze test as an integrative measure of EFL proficiency: A substitute for essays on college entrance examinations? *Language Learning*, 41(3), 313-336.
- Freed, B. (Ed). (1995a). *Second language acquisition in a study abroad context*. Amsterdam: John Benjamins.
- Freed, B. (1995b). What makes us think that students who study abroad become fluent? In B. Freed (Ed.), *Second language acquisition in a study abroad context* (pp.123-148). Amsterdam: John Benjamins.
- Freed, B., Dewey, D., Segalowitz, N., & Halter, R. (2004a). The language contact profile. *Studies in Second Language Acquisition*, 26(2), 349-356.
- Freed, B., Segalowitz, N., & Dewey, D. (2004b). 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(2), 275-301.
- Freed, B., So, S., & Lazar, N. (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.
- Garside, R., & Smith, N. (1997). A hybrid grammatical tagger: CLAWS4. In R. Garside, G. Leech & A. McEnery (Eds.), *Corpus*

References

- annotation: *Linguistic information from computer text corpora* (pp. 102-121). London: Longman.
- Gass, S., & Mackey, A. (2006). Input, interaction and output: An overview. *AILA Review*, 19, 3-17.
- Gass, S., & Mackey, A. (2007). Input, interaction and output in second language acquisition. In B. VanPatten & J. Williams (Eds.), *Theories in second language acquisition: An introduction* (pp. 175-200). Mahwah, NJ: Lawrence Erlbaum Associates.
- Gilbert, R. (2007). The simultaneous manipulation of task complexity along planning time and [+/- here-and-now]: Effects on L2 oral production. In M.P. García-Mayo (Ed.), *Investigating tasks in formal language learning* (pp. 44-68). Clevedon: Multilingual Matters.
- Ginsberg, R. (1992). *Language gains during study abroad: An analysis of the ACTR data*. Washington, DC: National Foreign Language Center.
- Golonka, E. (2006). Predictors revised: Linguistic knowledge and metalinguistic awareness in second language gain in Russian. *The Modern Language Journal*, 90(4), 496-505.
- Granger, S. (1993). Cognates: An aid or barrier to vocabulary development? *ITL Review of Applied Linguistics*, 99/100, 43-56.
- Granger, S., & Wynne, M. (1999). Optimising measures of lexical variation in EFL learner corpora. In J. Kirk (Ed.), *Corpora galore* (pp. 249-257). Amsterdam and Atlanta: Rodopi.
- Grey, S., Cox, J., Serafini, E., & Sanz, C. (2015). The role of individual differences in the study abroad context: Cognitive capacity and language development during short-term intensive language exposure. *The Modern Language Journal*, 99(1), 137-157.
- Hayden, J. (1998). The influence of a semester abroad on reading proficiency: A descriptive study. *Journal of the Chinese Language Teachers Association*, 33(3), 13-24.

References

- Henriksen, B. (1999). Three dimensions of vocabulary development. *Studies in Second Language Acquisition*, 21(2), 303-317.
- Hinkel, E. (2003). Simplicity without elegance: Features of sentences in L1 and L2 academic texts. *TESOL Quarterly*, 37(2), 275-302.
- Hokanson, S. (2000). Foreign language immersion homestays: Maximizing the accommodation of cognitive styles. *Applied Language Learning*, 11(2), 239-264.
- Hopper, P. (2001). Grammatical constructions and their discourse origins: Prototype or family resemblance? In M. Putz, S. Niemeier & R. Dirven (Eds.), *Applied cognitive linguistics: Theory and language acquisition* (pp. 109-130). Berlin: Mouton de Gruyter.
- Horst, M., & Collins, L. (2006). From “faible” to strong: How does their vocabulary grow? *The Canadian Modern Language Review*, 63(1), 83-106.
- Housen, A., & Kuiken, F. (2009). Complexity, accuracy and fluency in second language acquisition. *Applied Linguistics*, 30(4), 461-473.
- Howard, M. (2001). The effects of study abroad on L2 learner's structural skills: Evidence from advanced learners of French. In S.H. Foster-Cohen & A. Nizgorodcew (Eds.), *EUROSLA Yearbook 1*, (pp. 123-141). Amsterdam: John Benjamins.
- Howard, M. (2005). Second language acquisition in a study abroad context: A comparative investigation of the effects of study abroad and foreign language instruction on the L2 learner's grammatical development. In A. Housen & M. Pierrard (Eds.), *Investigations in instructed second language acquisition* (pp. 495-530). Berlin: Mouton de Gruyter.
- Howard, M. (2006). The expression of number and person through verb morphology in advanced French interlanguage. *International Review of Applied Linguistics in Language Teaching*, 44(1), 1-22.
- Huang, S., Willson, V., & Eslami, Z. (2012). The effects of task involvement load on L2 incidental vocabulary learning: A meta-analytic study. *The Modern Language Journal*, 96(4), 544-557.

References

- Huebner, T. (1998). Methodological considerations in data collection for language learning in a study abroad context. *Frontiers*, 4, 1-30.
- Hughes, A. (1989). *Testing for language teachers*. Cambridge: Cambridge University Press.
- Hulstijn, J. (2006). Incidental and intentional learning. In M.H. Long & C.J. Doughty (Eds.), *The handbook of second language acquisition* (pp. 349-382). Oxford: Blackwell.
- Hyltenstam, K. (1988). Lexical characteristics of near-native second-language learners of Swedish. *Journal of Multilingual and Multicultural Development*, 9, 67-84.
- Ife, A., Vives Boix, G., & Meara, P.M. (2000). The impact of study abroad on the vocabulary development of different proficiency groups. *Spanish Applied Linguistics*, 4(1), 55-84.
- Isabelli, C., & Nishida, C. (2005). Development of the Spanish subjunctive in a nine-month study-abroad setting. In D. Eddington (Ed.), *Selected proceedings of the 6th conference on the acquisition of Spanish and Portuguese as first and second Languages* (pp. 78-91). Somerville: Cascadilla Press.
- Isabelli-García, C.L. (2006). Study abroad social networks, motivation, and attitudes: Implications for SLA. In M.A. DuFon & E. Churchill (Eds.), *Language learners in study abroad contexts* (pp. 231-258). Clevedon: Multilingual Matters.
- Jackson, J. (2008). *Language, identity, and study abroad: Sociocultural perspectives*. London: Equinox.
- James, C. (1998). *Errors in language learning and use: Exploring error analysis*. London: Longman.
- Jarvis, S. (2002). Short texts, best-fitting curves and new measures of lexical diversity. *Language Testing*, 19(1), 57-84.
- Jarvis, S. (2013). Capturing the diversity in lexical diversity. *Language Learning*, 63(1), 87-106.

References

- Jiménez Jiménez, A.F. (2010). A comparative study on second language vocabulary development: Study abroad vs classroom settings. *Frontiers*, 19, 105-123.
- Johansson, V. (2008). Lexical diversity and lexical density in speech and writing: A developmental perspective. *Lund Working Papers in Linguistics*, 53, 61-79.
- Juan-Garau, M. (2012). Context matters: Variability across three SLA learning settings. In E. Piechurska-Kuciel & L. Piasecka (Eds.), *Variability and stability in foreign and second language learning contexts* (pp. 220-242). Newcastle upon Tyne: Cambridge Scholars Publishing.
- Juan-Garau, M. (2014). Oral accuracy growth after formal instruction and study abroad: Onset level, contact factors and long-term effects. In C. Pérez-Vidal (Ed.), *Language acquisition in study abroad and formal instruction contexts* (pp. 87-110). Amsterdam: John Benjamins.
- Juan-Garau, M., Salazar-Noguera, J., & Prieto-Arranz, J.I. (2014). English learners' lexico-grammatical and motivational development at home and abroad. In C. Pérez-Vidal (Ed.), *Language acquisition in study abroad and formal instruction contexts* (pp. 235-258). Amsterdam: John Benjamins.
- Kelly, P. (1991). Lexical ignorance: The main obstacle to listening comprehension with advanced foreign language learners. *IRAL*, 29(2), 135-149.
- Kinginger, C. (2008). Language learning in study abroad: Case studies of Americans in France. *The Modern Language Journal*, 92(1), 1-124.
- Kinginger, C. (2009). *Language learning and study abroad: A critical reading of research*. New York: Palgrave Macmillan.
- Kinginger, C. (Ed.). (2013). *Social and cultural aspects of language learning in study abroad*. Amsterdam: John Benjamins.

References

- Kirsner, K. (1994). Second language vocabulary learning: The role of implicit processes. In N.C. Ellis (Ed.), *Implicit and explicit learning of languages* (pp. 283-311). London: Academic Press.
- Krashen, S.D. (1976). Formal and informal linguistic environments in language learning and language acquisition. *TESOL Quarterly*, 10(2), 157-168.
- Krashen, S.D. (1985). *The input hypothesis: Issues and Implications*. London: Longman.
- Kroll, J.F., & Stewart, E. (1994). Category interference in translation and picture naming: Evidence for asymmetric connections between bilingual memory representations. *Journal of Memory and Language*, 33, 149-174.
- Lafford, B., & Ryan, B. (1995). The acquisition of lexical meaning in a study abroad context: The Spanish prepositions POR and PARA. *Hispania*, 78(3), 528-547.
- Lapkin, S., Hart, D., & Swain, M. (1995). A Canadian interprovincial exchange: Evaluating the linguistic impact of a three-month stay in Quebec. In B. Freed (Ed.), *Second language acquisition in a study abroad context* (pp. 67-94). Amsterdam: John Benjamins.
- Lara, A.R. (2014). *Complexity, accuracy and fluency development through study abroad programmes varying in duration*. (Doctoral dissertation, Universitat Pompeu Fabra, Spain). Retrieved from <http://www.tdx.cat/handle/10803/284229>.
- Laufer, B. (1991). The development of L2 lexis in the expression of the advanced language learner. *The Modern Language Journal*, 75(4), 440-448.
- Laufer, B. (1992). How much lexis is necessary for reading comprehension? In P. Arnaud & H. Béjoint (Eds.), *Vocabulary and applied linguistics* (pp. 126-132). London: MacMillan.
- Laufer, B. (1994). The lexical profile of second language writing: Does it change over time? *RELC Journal*, 25(2), 21-33.

References

- Laufer, B. (1995). Beyond 2000: A measure of productive lexicon in second language. In L. Eubank & M. Sharwood-Smith (Eds.), *The current state of interlanguage* (pp. 265-272). Philadelphia: John Benjamins.
- Laufer, B. (1998). The development of passive and active vocabulary in a second language: same or different? *Applied Linguistics*, 19(2), 255-271.
- Laufer, B. (2005). Lexical frequency profiles: From Monte Carlo to the real world. A response to Meara (2005). *Applied Linguistics*, 26(4), 582-588.
- 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.
- Laufer, B., & Goldstein, Z. (2004). Testing vocabulary knowledge: Size, strength, and computer adaptiveness. *Language Learning*, 54(3), 399-436.
- Laufer, B., & Nation, I.S.P. (1995). Vocabulary size and use: Lexical richness in L2 written production. *Applied Linguistics*, 16(3), 307-323.
- Laufer, B., & Paribakht, T.S. (1998). The relationship between passive and active vocabularies: Effects of language learning context. *Language Learning*, 48(3), 365-391.
- Lee, S.H., & Muncie, J. (2006). From receptive to productive: Improving ESL learners' use of vocabulary in a postreading composition task. *TESOL Quarterly*, 40(2), 295-320.
- Lennon, P. (1990). The advanced learner at large in the L2 community: Developments in spoken performance. *IRAL*, 28(4), 309-324.
- Levelt, W. (1989). *Speaking: From intention to articulation*. Cambridge: MIT Press.

References

- Levenston, E.A. (1979). Second language vocabulary acquisition: Issues and problems. *Interlanguage Studies Bulletin*, 41, 147-160.
- Lewis, M. (1993). *The lexical approach: The state of ELT and a way forward*. Hove: Language Teaching Publications.
- Linnarud, M. (1986). *Lexis in composition: A performance analysis of Swedish learners' written English*. Sweden: Gleerup.
- Llanes, À., & Muñoz, C. (2009). A short stay abroad: Does it make a difference? *System*, 37(3), 353-365.
- Long, M.H. (1996). The role of the linguistic environment in second language acquisition. In W. C. Ritchie & T. K. Bhatia (Eds.), *Handbook of second language acquisition* (pp. 413-468). New York, NY: Academic Press.
- Lu, X. (2012). The relationship of lexical richness to the quality of ESL learners' oral narratives. *The Modern Language Journal*, 96(2), 190-208.
- MacWhinney, B. (2000). *The CHILDES project: Tools for analyzing talk* (3rd ed.). Mahwah: Lawrence Erlbaum Associates.
- Malvern, D., & Richards, B. (1997). A new measure of lexical diversity. In A. Ryan & A. Wray (Eds.), *Evolving models of language* (pp. 58-71). Clevedon: Multilingual Matters.
- Malvern, D., & Richards, B. (2000). Validation of a new measure of lexical diversity. In M. Beers, B. van den Bogaerde, G. Bol, J. de Jong & C. Rooijmans (Eds.), *From sound to sentence: Studies on first language acquisition* (pp. 81-96). Gronigen: Centre for Language and Cognition.
- Malvern, D., & Richards, B. (2002). Investigating accommodation in language proficiency interviews using a new measure of lexical diversity. *Language Testing*, 19(1), 85-104.
- Malvern, D., Richards, B., Chipere, N., & Durán, P. (2004). *Lexical diversity and language development: Quantification and assessment*. New York: Palgrave Macmillan.

References

- Marriot, H. (1995). Acquisition of politeness patterns by exchange students in Japan. In B. Freed (Ed.), *Second language acquisition in study abroad context* (pp. 197-224). Amsterdam: John Benjamins.
- Marsden, E., & David, A. (2008). Vocabulary use during conversation: A cross-sectional study of development from year 9 to year 13 amongst learners of Spanish and French. *The Language Learning Journal*, 36(2), 181-198.
- McCarthy, M. (1994). It, this, and that. In M. Coultbard (Ed.), *Advances in written text analysis* (pp. 266-275). New York: Routledge.
- McCarthy, P.M., & Jarvis, S. (2007). Vocd: A theoretical and empirical evaluation. *Language Testing*, 24(4), 459-488.
- McCarthy, P.M., & Jarvis, S. (2013). From intrinsic to extrinsic issues of lexical diversity assessment. In S. Jarvis & H. Daller (Eds.), *Vocabulary knowledge: Human ratings and automated measures* (pp. 45-78). Amsterdam: John Benjamins.
- Meara, P.M. (1980). Vocabulary acquisition: A neglected aspect of language learning. *Language Teaching and Linguistics*, 14, 221-246.
- Meara, P.M. (1988). Learning words in an L1 and an L2. *Polyglot*, 9(3), 1-19.
- Meara, P.M. (1990). A note on passive vocabulary. *Second Language Research*, 6(2), 150-154.
- Meara, P.M. (1994). The year abroad and its effects. *The Language Learning Journal*, 10(1), 32-38.
- Meara, P.M. (1996a). The dimensions of lexical competence. In G. Brown, K. Malmkjaer & J. Williams (Eds.), *Performance & competence in second language acquisition* (pp. 33-53). Cambridge: Cambridge University Press.
- Meara, P.M. (1996b). *The vocabulary knowledge framework*. Retrieved from <http://www.lognostics.co.uk/vlibrary/meara1996c.pdf>.

References

- Meara, P.M. (1996c). The classical research in vocabulary acquisition. In G. Anderman & M. Rogers (Eds.), *Words, words, words* (pp. 27-40) Clevedon: Multilingual Matters.
- Meara, P.M. (2005). Lexical frequency profiles: A Monte Carlo analysis. *Applied Linguistics*, 26(1), 32-47.
- Meara, P.M. (2009). *Connected words: Word associations and second language vocabulary acquisition*. Amsterdam: John Benjamins.
- Meara, P.M., & Bell, H. (2001). P_Lex: A simple and effective way of describing the lexical characteristics of short L2 texts. *Prospect*, 16(3), 5-19.
- Meara, P.M., Lightbown, P.M., & Halter, R.H. (1997). Classrooms as lexical environments. *Language Teaching Research*, 1(1), 28-47.
- Meara, P.M., & Miralpeix, I. (2007). D_Tools (Version 2.0) [Computer software]. Swansea: Lognostics. Retrieved from <http://www.lognostics.co.uk/tools/>.
- Meara, P.M., & Miralpeix, I. (2008). *D_Tools: The Manual*. Retrieved from <http://www.lognostics.co.uk/tools/>.
- Meara, P.M., & Miralpeix, I. (in press). *Tools for researching vocabulary*. Clevedon: Multilingual Matters.
- Melka Teichroew, F. (1997). Receptive vs. productive aspects of vocabulary. In N. Schmitt & M. McCarthy (Eds.), *Vocabulary: Description, acquisition and pedagogy* (pp. 84-102). Cambridge: Cambridge University Press.
- Merino, E., & Avello, P. (2014). Contrasting intercultural awareness at home and abroad. In C. Pérez-Vidal (Ed.), *Language acquisition in study abroad and formal instruction contexts*. Amsterdam: John Benjamins.
- Milton, J. (2009). *Measuring second language vocabulary acquisition*. Clevedon: Multilingual Matters.

References

- Milton, J., & Fitzpatrick, T. (Eds.). (2014). *Dimensions of vocabulary knowledge*. Basingstoke: Palgrave Macmillan.
- Milton, J., & Hales, T. (1997). Applying a lexical profiling system to technical English. In A. Ryan & A. Wray (Eds.), *Evolving models of language* (pp. 72-82). Clevedon: Multilingual Matters.
- Milton, J., & Meara, P.M. (1995). How periods abroad affect vocabulary growth in a foreign language. *ITL Review of Applied Linguistics*, 107/108, 17-34.
- Miralpeix, I. (2008). *The influence of age on vocabulary acquisition in English as a foreign language*. (Doctoral dissertation, Universitat de Barcelona, Spain). Retrieved from <http://tdx.cat/handle/10803/1670>.
- Mitchell, R., Tracy-Ventura, N., & McManus, K. (Eds.). (2015). *Social interaction, identity and language learning during residence abroad*. Eurosla Monographs Series 4: EuroSLA.
- Mollet, E., Wray, A., Fitzpatrick, T., Wray, N.R., & Wright, M.J. (2010). Choosing the best tools for comparative analyses of texts. *International Journal of Corpus Linguistics*, 15(4), 429-473.
- Montgomery, D. (1997). *Spelling: Remedial strategies*. London: Cassell.
- Mora, J.C. (2008). Learning context effects on the acquisition of a second language phonology. In C. Pérez-Vidal, M. Juan-Garau & A. Bel (Eds.), *A portrait of the young in the new multilingual Spain* (pp. 241-263). Clevedon: Multilingual Matters.
- Mora, J.C. (2014). The role of onset level on L2 perceptual phonological development after formal instruction and study abroad. In C. Pérez-Vidal (Ed.), *Language acquisition in study abroad and formal instruction contexts* (pp. 167-194). Amsterdam: John Benjamins.
- Morris, L., & Cobb, T. (2004). Vocabulary profiles as predictors of the academic performance of teaching English as a second language trainees. *System*, 32(1), 75-87.

References

- Morris, S.B. (2008). Estimating effect sizes from pretest-posttest-control group designs. *Organizational Research Methods, 11*, 364-386.
- Muncie, J. (2002). Process writing and vocabulary development: Comparing lexical frequency profiles across drafts. *System, 30*(2), 225-235.
- Muñoz, C. (2006). *Age and the rate of foreign language learning*. Clevedon: Multilingual Matters.
- Muñoz, C. (2007). CLIL: Some thoughts on its psycholinguistic principles. In F. Lorenzo, S. Casal, V. de Aba & P. Moore (Eds.), *Models and Practice in CLIL. Revista Española de Lingüística Aplicada, 20*, 17-26.
- Murphy, L., & Roca de Larios, J. (2010). Searching for words: One strategic use of the mother tongue by advanced Spanish EFL writers. *Journal of Second Language Writing, 19*, 61-81.
- Nagle, C., Morales-Front, A., Moorman, C., & Sanz, C. (in press). Improvements in pronunciation during a short study abroad program. In D. Velliaris & D. Coleman-George (Eds.), *Handbook of research on study abroad programs and outbound mobility*. Hershey: IGI Global.
- Nation, I.S.P. (1990). *Teaching and learning vocabulary*. New York: Newbury House.
- Nation, I.S.P. (2001). *Learning vocabulary in another language*. Cambridge: Cambridge University Press.
- Nation, I.S.P., & Heatley, A. (1994). Range: A program for the analysis of vocabulary in texts [Computer software]. Retrieved from <http://www.victoria.ac.nz/lals/staff/paul-nation.aspx>.
- Nation, I.S.P., & Webb, S. (2011). *Researching and analyzing vocabulary*. Boston: Heinle.
- Nattinger, J.R., & DeCarrico, J.S. (1992). *Lexical phrases and language teaching*. Oxford: Oxford University Press.

References

- Navés, T. (2006). *The long-term effects of an early start on foreign language writing*. (Unpublished doctoral dissertation). Universitat de Barcelona, Spain.
- Newton, J. (2013). Incidental vocabulary learning in classroom communication tasks. *Language Teaching Research*, 17(2), 164-187.
- Nippold, M. (2006). *Later language development: School-age children, adolescents, and young adults* (3rd ed.). Austin: Pro-Ed.
- Olson, D.R. (2006). Continuing the discourse on literacy. *Research in Teaching of English*, 41, 175-179.
- Ortega, L. (2014). Ways forward for a bi/multilingual turn in SLA. In S. May (Ed.), *The Multilingual turn: Implications for SLA, TESOL and bilingual education* (pp. 32-53). New York: Routledge.
- Ovtcharov, V., Cobb, T., & Halter, R.H. (2006). La richesse lexicale des productions orales: Mesure fiable du niveau de compétence langagière. *The Canadian Modern Language Review*, 63(1), 107-125.
- Paige, M., Cohen, A., Kappler, B., Chi, J., & Lassegard, J. (2002). *Maximizing study abroad. A students' guide to strategies for language and culture learning and use*. Minneapolis MN: University of Minnesota Center for Advanced Research on Language Acquisition.
- 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. 191-225). London: Longman.
- Pérez-Vidal, C. (Ed.). (2014a). *Language acquisition in study abroad and formal instruction contexts*. Amsterdam: John Benjamins.
- Pérez-Vidal, C. (2014b). Study abroad and formal instruction contrasted: The SALA project. In C. Pérez-Vidal (Ed.), *Language acquisition in study abroad and formal instruction contexts* (pp. 17-58). Amsterdam: John Benjamins.

References

- Pérez-Vidal, C. (2015). Practice makes best: Contrasting learning contexts, comparing learner progress. *International Journal of Multilingualism*, 12(4), 453-470.
- Pérez-Vidal, C., & Juan-Garau, M. (2011). The effect of context and input conditions on oral and written development: A study abroad perspective. *International Review of Applied Linguistics in Language Teaching*, 49(2), 157-185.
- Pérez-Vidal, C., Juan-Garau, M., Mora, J.C., & Valls-Ferrer, M. (2012). Oral and written development in formal instruction and study abroad: Differential effects of learning context. In C. Muñoz (Ed.), *Intensive exposure experiences in second language learning* (pp. 213-233). Bristol: Multilingual Matters.
- Peters, A. (1983). *The units of language acquisition*. Cambridge: Cambridge University Press.
- Polinsky, M. (2005). Word class distinctions in an incomplete grammar. In D. Ravid & H. Bat-Zeev Shyldkrot (Eds.), *Perspectives on language and language development* (pp. 419-436). Dordrecht: Kluwer.
- Qian, D.D. (1999). Assessing the roles of depth and breadth of vocabulary knowledge in ESL reading comprehension. *The Canadian Modern Language Review*, 56(2), 282-308.
- Raupach, M. (1984). Formulae in second language speech production. In H. Dechert, D. Moehle & M. Raupach (Eds.), *Second language production* (pp. 114-137). Tübingen: Gunter Narr.
- Read, J. (1988). Measuring the vocabulary knowledge of second language learners. *RELJ Journal*, 19(2), 12-25.
- Read, J. (2000). *Assessing vocabulary*. Cambridge: Cambridge University Press.
- Read, J. (2005). Applying lexical statistics to the IELTS speaking test. *Research Notes*, 20, 12-16.

References

- Read, J., & Nation, I.S.P. (2006). An investigation of the lexical dimension of the IELTS speaking test. *IELTS Research Reports*, 6, 207-231. Canberra: IELTS Australia.
- Regan, V. (1995). The acquisition of sociolinguistic native speech norms: Effects of a year abroad on second language learners of French. In B. Freed (Ed.), *Second language acquisition in a study abroad context* (pp. 245-268). Amsterdam: John Benjamins.
- Regan, V. (1998). Sociolinguistics and language learning in a study abroad context. *Frontiers*, 4, 61-90.
- Regan, V., Howard, M., & Lemée, I. (2009). *The acquisition of sociolinguistic competence in a study abroad context*. Buffalo: Multilingual Matters.
- Richards, B. (1987). Type/Token ratios: What do they really tell us? *Journal of Child Language*, 14(2), 201-209.
- Richards, B., Daller, H., Malvern, D., Meara, P.M., Milton, J., & Treffers-Daller, J. (Eds.). (2009). *Vocabulary studies in first and second language acquisition: The interface between theory and application*. New York: Palgrave.
- Richards, B., & Malvern, D. (2000, September). *Measuring vocabulary diversity in teenage learners of French*. Paper presented at the British Educational Research Association Conference, Cardiff. Retrieved from <http://www.leeds.ac.uk/educol/documents/0001541.htm>.
- Richards, J.C. (1976). The role of vocabulary teaching. *TESOL Quarterly*, 10(1), 77-89.
- Robinson, P. (Ed.). (2002). *Individual differences and instructed language learning*. Amsterdam: John Benjamins.
- Robinson, P., & Gilabert, R. (2007). Task complexity, the cognition hypothesis and second language learning and performance. *IRAL*, 45(3), 161-176.

References

- Ruiz de Zarobe, Y. (2007). CLIL in a bilingual community: Similarities and differences with the learning of English as a Foreign Language. *VIEWS*, 16(3), 47-52.
- Ryan, J., & Lafford, B. (1992). Acquisition of lexical meaning in a study abroad environment: SER and ESTAR and the Granada experience. *Hispania*, 75(3), 714-722.
- Sanz, C. (2014). Contributions of study abroad research to our understanding of SLA processes and outcomes: The SALA Project, an appraisal. In C. Pérez-Vidal (Ed.), *Language acquisition in study abroad and formal instruction contexts* (pp. 1-13). Amsterdam: John Benjamins.
- Sasaki, M. (2007). Effects of study-abroad experiences on EFL writers: A multiple-data analysis. *The Modern Language Journal*, 91(4), 602-620.
- Sasaki, M. (2009). Changes in English as a foreign language students' writing over 3.5 years: A sociocognitive account. In R.M. Manchón (Ed.), *Writing in foreign language contexts: Learning, teaching, and research* (pp. 49-76). Clevedon: Multilingual Matters.
- Schmidt, R. (1994). Deconstructing consciousness in search of useful definitions for Applied Linguistics. *AILA Review*, 11, 112-116.
- Schmitt, N. (2010). *Researching vocabulary: A vocabulary research manual*. Basingstoke: Palgrave Macmillan.
- Schmitt, N. (2014). Size and depth of vocabulary knowledge: What the research shows. *Language Learning*, 64(4), 913-951.
- Schmitt, N., & Carter, R. (2004). Formulaic sequences in action: An introduction. In N. Schmitt (Ed.), *Formulaic sequences* (pp. 1-22). Amsterdam: John Benjamins.
- Segalowitz, N. (2011, September). *Cognitive mechanisms underlying fluency*. Paper presented at the 2nd Barcelona Summer School on Bilingualism and Multilingualism, Barcelona.

References

- Segalowitz, N., & Freed, B. (2004). Context, contact and cognition in oral fluency acquisition: Learning Spanish in at home and study abroad contexts. *Studies in Second Language Acquisition*, 26(2), 173-199.
- Serrano, R., Tragant, E., & Llanes, À. (2012). A longitudinal analysis of the effects of one year abroad. *The Canadian Modern Language Review*, 68(2), 138-163.
- Shaw, P., & Liu, E.T.K. (1998). What develops in the development of second-language writing? *Applied linguistics*, 19(2), 225-254.
- Siegal, M. (1995). Individual differences and study abroad: Women learning Japanese in Japan. In B. Freed (Ed.), *Second language acquisition in a study abroad context* (pp. 225-244). Amsterdam: John Benjamins.
- Silva, T. (1993). Toward an understanding of the distinct nature of L2 writing: The ESL research and its implications. *TESOL Quarterly*, 27(4), 657-677.
- Sinclair, J. (1991). *Corpus, concordance, collocation*. Oxford: Oxford University Press.
- Sinclair, J. (2004). New evidence, new priorities, new attitudes. In J. Sinclair (Ed.), *How to use corpora in language teaching* (pp. 271-299). Amsterdam: John Benjamins.
- Singleton, D. (1999). *Exploring the second language mental lexicon*. Cambridge: Cambridge University Press.
- Siyanova, A., & Schmitt, N. (2008). L2 learner production and processing of collocation: A multi-study perspective. *The Canadian Modern Language Review*, 64(3), 429-458.
- Skahan, P., & Foster, P. (1997). Task type and task processing conditions as influences on foreign language performance. *Language Teaching Research*, 1(3), 185-211.

References

- Sperling, M. (1996). Revisiting the writing-speaking connection: Challenges for research on writing and writing instruction. *Review of Educational Research*, 66(1), 53-86.
- Sperling, M., & Freedman, S.W. (2001). Research on writing. In V. Richardson (Ed.), *Handbook of research on teaching* (4th ed., pp. 370-389). Washington, DC: American Educational Research Association.
- Strömqvist, S., Johansson, V., Kriz, S., Ragnarsdóttir, H., & Ravid, D. (2002). Towards a cross-linguistic comparison of lexical quanta in speech and writing. *Written Language & Literacy*, 5(1), 45-68.
- Swain, M. (1985). Communicative competence: Some roles of comprehensible input and comprehensible output in its development. In S. Gass & C. Madden (Eds.), *Input in second language acquisition* (pp. 235-256). Rowley, MA: Newbury House.
- Swain, M. (1993). The output hypothesis: Just speaking and writing aren't enough. *The Canadian Modern Language Review*, 50(1), 158-164.
- Tanaka, K., & Ellis, R. (2003). Study-abroad, language proficiency, and learner beliefs about language learning. *JALT Journal*, 25(1), 63-85.
- Tavakoli, P., & Foster, P. (2008). Task design and second language performance: The effect of narrative type on learner output. *Language Learning*, 58(2), 439-473.
- Towell, R. (2002). Relative degrees of fluency: A comparative case study of advanced learners of French. *IRAL*, 40(2), 117-150.
- Towell, R., Hawkins, R., & Bazergui, N. (1996). The development of fluency in advanced learners of French. *Applied Linguistics*, 17(1), 84-119.
- Tragant, E., Marsol, A., Serrano, R., & Llanes, À. (2015). Vocabulary learning at primary school: A comparison of EFL and CLIL. *International Journal of Bilingual Education and Bilingualism*, 1-13.

References

- Treffers-Daller, J. (2013). Measuring lexical diversity among L2 learners of French: An exploration of the validity of D, MTLD and HD-D as measures of language ability. In S. Jarvis & H. Daller (Eds.), *Vocabulary knowledge: Human ratings and automated measures* (pp. 79-104). Amsterdam: John Benjamins.
- Trenchs-Parera, M. (2009). Effects of formal instruction and a stay abroad on the acquisition of native-like oral fluency. *The Canadian Modern Language Review*, 65(3), 365-393.
- Ure, J. (1971). Lexical density and register differentiation. In G.E. Perren & J.L.M. Trim (Eds.), *Applications of linguistics: Selected papers of the second international congress of applied linguistics* (pp. 443-452). Cambridge: Cambridge University Press.
- Valls-Ferrer, M. (2011). *The development of oral fluency during a study abroad period*. (Doctoral dissertation, Universitat Pompeu Fabra, Spain). Retrieved from <http://www.tdx.cat/handle/10803/52064>.
- Valls-Ferrer, M., & Mora, J.C. (2014). L2 fluency development in formal instruction and study abroad: The role of initial fluency level and language contact. In C. Pérez-Vidal (Ed.), *Language acquisition in study abroad and formal instruction contexts* (pp.111-136). Amsterdam: John Benjamins.
- Van Hell, J., & De Groot, A.M.B. (1998). Conceptual representation in bilingual memory: Effects of concreteness and cognate status in word association. *Bilingualism: Language and Cognition*, 1, 193-211.
- Van Hout, R., & Vermeer, A. (2007). Comparing measures of lexical richness. In H. Daller, J. Milton & J. Treffers-Daller (Eds.), *Modelling and assessing vocabulary* (pp. 93-115). Cambridge: Cambridge University Press.
- Vanniarajan, S. (1997). An interactive model of vocabulary acquisition. *Applied Language Learning*, 8(2), 183-216.
- Vermeer, A. (2000). Coming to grips with lexical richness. *Language Testing*, 17(1), 65-83.

References

- Vermeer, A. (2004). The relation between lexical richness and vocabulary size in Dutch L1 and L2 children. In P. Bogaards & B. Laufer (Eds.), *Vocabulary in a second language* (pp. 173-189). Amsterdam: John Benjamins.
- Webb, S. (2005). Receptive and productive vocabulary learning: The effects of reading and writing on word knowledge. *Studies in Second Language Acquisition*, 27(1), 33-52.
- Wesche, M.B., & Paribakht, T.S. (1996). Assessing second language vocabulary knowledge: Depth versus breadth. *The Canadian Modern Language Review*, 53(1), 13-40.
- Williams, J. (2012). The potential role(s) of writing in second language development. *Journal of Second Language Writing*, 21, 321-331.
- Willis, F., Doble, G., Sankarayya, U., & Smithers, A. (1977). *Residence abroad and the student of modern languages. A preliminary study*. Bradford: University of Bradford Modern Language Centre.
- Wilkins, D. (1972). *Second language learning and teaching*. London: Edward Arnold.
- Wolfe-Quintero, K., Inagaki, S., & Kim, H.Y. (1998). *Second language development in writing: Measures of fluency, accuracy, & complexity*. Manoa: University of Hawai'i Press.
- Wray, A. (2002). *Formulaic language and the lexicon*. Cambridge: Cambridge University Press.
- Yu, G. (2009). Lexical diversity in writing and speaking task performances. *Applied Linguistics*, 31(2), 236-259.
- Zareva, A., Schwanenflugel, P., & Nikolova, Y. (2005). Relationship between lexical competence and language proficiency: Variable sensitivity. *Studies in Second Language Acquisition*, 27(4), 567-595.
- Zimmerman, C.B. (1997). Historical trends in second language vocabulary instruction. In J. Coady & T. Huckin (Eds.), *Second language vocabulary acquisition: A rationale for pedagogy* (pp. 5-19). Cambridge: Cambridge University Press.

Appendices

Appendix 1. Complete listing of SALA instruments.....	246
Appendix 2. Oral interview questions.....	247
Appendix 3. Conventions to edit data (I).....	248
Appendix 4. Conventions to edit data (II).....	249
Appendix 5. List of spelling errors and non-words.....	250
Appendix 6. Output of VocabProfile for a text.....	253
Appendix 7. Error classification scheme.....	254
Appendix 8. Word Class taxonomy (CLAWS 5 tagset).....	255
Appendix 9. Off-list words (OffL) used in the T2 oral interviews..	256

Appendix 1. Complete listing of SALA instruments

Description of the test		
Written tests	Grammar 1*	Sentence-rephrasing task
	Grammar 2	Gap-filling exercise on tense and aspect
	Cloze*	Gap-filling story entitled <i>“The Lady Who Liked Adventure”</i>
	Composition*	Argumentative essay in response to the statement: <i>“Someone who moves to a foreign country should always adopt the customs and way of life of his/ her new country”</i>
Listening comprehension	Phonological perception	10-minute audio recording with multiple-choice and fill-in-the-gap questions on the content
		Auditory task for discrimination ability for the vowel quality and consonant voicing contrasts
Oral tests	Reading aloud	Reading out loud a passage from the story entitled <i>“The North Wind and the Sun”</i>
	Interview* (in pairs)	Semi-guided informal interview about students’ everyday experiences at university
	Role-play (in pairs)	Two-way, problem-solving style scenario in which one of the students acted as a house-owner and the other one as a decorator
Questionnaires	Linguistic profile*	Background questionnaire designed to gather participants’ biographical and linguistic information
	Attitude	Questionnaire on participants’ attitudes, beliefs and motivation towards language learning
	SAC*	Questionnaire on participants’ Stay Abroad conditions during the observation period
	SA diary	Personal diaries that students completed on a voluntary basis while abroad

(Adapted from Lara, 2014)

*tests used in the development of this study.

Appendix 2. Oral interview questions

SALA / INTERVIEW / SU code: _____ Date: _____ T1

Student A asks student B:

1. Why did you choose [*this university*⁴⁹] as the university where you wanted to study?
2. How do you like your classes so far and why?
3. Which classes do you enjoy the most and why?
4. Has it been easy to make friends here?
5. Tell us about either a very good experience or a very bad experience that you have had at the University.
6. How similar and how different is your University life from what you expected it to be like?
7. How similar and how different are your study habits now from what they were like in high school?

Student B asks student A:

1. Have you made any good friends at the university yet? If yes, what are they like? If no, why do you think you haven't?
2. How different is your university life from your high school life?
3. How do you like the library? How often do you go there? What resources do you usually use?
4. What advantages and disadvantages do you see in the location of the [*this university*] building?
5. What do you normally do for lunch during school days?
6. What do your friends and family think of your future profession as a translator and interpreter?
7. What do you think of the requirement at [*this university*] of having to study abroad?

⁴⁹ Here the name of the university, which was specified in the original interview question, has been substituted with *this university*.

Appendix 3. Conventions to edit data (I)

Set#	Changes made	Analysis	Format	Software
1	None	<u>Qualitative:</u> NLSs	.doc	n/a
2	Pruned data Spelling errors corrected Non-words eliminated No repetitions	<u>Quantitative:</u> Fluency (word count) Density (word class analysis) Diversity (GI & D) Sophistication (LFP & GLI)	.txt	VocabPr. D_Tools CLAWS
3	Formatting and punctuation modified to CHAT conventions	Accuracy	.cha	CLAN CHILDES

1) Raw data (.doc)

When someone moves to a foreign country, it's impossible that he/she can't adopt the customs and way of life of his/her new country.

In my opinion, one reason to go abroad is knowing different customs and cultures. When someone go abroad, he / she tries to show to their new friends that he/she is interested in their way of life and that he/she want to know more and more about it. (...)

2) Plane text (.txt)

when someone moves to a foreign country it is impossible that he she cannot adopt the customs and way of life of his her new country in my opinion one reason to go abroad is knowing different customs and cultures when someone go abroad he she tries to show to their new friends that he she is interested in their way of life and that he she want to know more and more about it

3) CLAN file (.cha)

@Begin

@Languages: eng

@Participants: STU Student

@ID: eng|RIRA02|STU||||Student||

*STU: When someone moves to a foreign country, it's impossible that he/she can't adopt the customs and way of life of his/her new country.

*STU: In my opinion, one reason to go abroad is knowing different customs and cultures.

*STU: When someone go abroad, he/she tries to show to their new friends that he/she is interested in their way of life and that he/she want to know more and more about it.

Appendix 4. Conventions to edit data (II)

The following set of conventions was applied to control the amount of proper nouns and make the data anonymous (Adapted from Navés, 2006; Miralpeix, 2008).

Countries	Thisland Euroland Otherland
Cities	Eurocity Europolis Newpolis
Continents	Xcontinent Ycontinent
Places (general)	Xplace Yplace Zplace
Foreign people	Xlander Ylander Zlander Otherlander
Foreign languages	Xlanguage Ylanguage Zlanguage Otherlanguage
Religion	Xreligist Yreligist Zreligist
Universities	Nobel Keynes
University programmes	Erasmus Xplan
University courses	Xcourse Ycourse Zcourse
Organisations	Xorganisation Yorganisation
Names (boys)	David Daniel
Names (girls)	Laura Esther

Appendix 5. List of spelling errors and non-words

1) Spelling errors

T1	Spelling error	T2	Spelling error	T3	Spelling error
CLFA	unconsciously	GAPE	extrem; reallity; cappa- cities; beggining; aban- done	CLFA	appropriate
GAPE	conciuous	GIRO	coloquial; bizarr; classe	GAPE	possition; tryies; possitive; enfacize; extrems
GIRO	you (your); remem- berance; yourselfe	GUMA	unconfortable; addapt; stablish; behabe	GIRO	loosing (losing)
GORE	handicappes	MOLL	lifes; mantain; forgeting	GUMA	pleasent; chose (choose); foreing lifes
PAAR	underdevelopped; simbol; inhabitats government (x2); chouvinists; filo- sofy; Morrocoo	PAAR	believes (beliefs)	PAAR	lifes
PARE	arguemet; mettdod; specielly; phisical; routin	RIRA	caracteristic	PARE	hollidays
VIGO	acquired; costumes	VIGO	loose (lose); life (live)	RIRA	complet
ALCA	missunderstandings	ALCA	lifes	VIGO	oportunity
ALLA	obiously; disapoint; the (them)	ALLA	loose (lose); essencial	ALCA	sample (simple); peope estable; loose (lose); were (where)
BOAD	enrichement	ARVI	of (off); the (them); the (they); lifes	ALLA	righ; aisolated (x2); latter (later); the (they); afronted; themsel
BRVA	Afganistan; think (thing) (x2); you're (your)	BOAD	communication; they (their)	ARVI	tipical (x2); chose (choose); exemple; foreing (x3); lifes; purpouse; aswell argumments; defi- netely
CAPU	practise (practice); themself (x2); they (them); les	CAMO	imprive	BOAD	alwas
CARA	dilema; main- tainance	FEJO	dilema; conciuous; loosing (losing)	CAMO	like (life); thing (think)
FEJO	nacionalist	FOTR	county (country); luck (lack); lat	CAPU	nevertheless; inmoral
FOTR	live (n)	JIFO	belives; althought	CARA	carless; one (on)
JIFO	rythm; af (at)	JUPL	problematique; life (live)	FOTR	althought oneshelf; life (live) goverment
JUPL	confortable sympathethic	MOM0	recieving; appart; loosing (losing)	JIFO	

Appendices

MUAL	sustitute	MUAL	who	MOM0	fourtunately; word (world); forcing
NICA	neighbor; inmerse; adquire	NICA	refering	NICA	fenomena; isollated believes (beliefs); neighbor
ROCR	live (life); comming	ROCR	government; disavantages	ROCR	always; humanbeing; te (the); adaptting; anothe; lifes (x2); belive
SAMA	confortable (x2)	SAMA	live (life) (x2)	SAMA	diferent

2) Non-words

T1	Composition	T2	Interview
GIRO01	conscient	GIRO01	esforç
MOLL01	typic; junkie	GUMA01	syntaxis
VAMO01	incondicional; conflictive	PAAR01	touristic
JIFO01	poblation	JUPL01	cooker
JUPL01	musulman; recomain; jude; ordre	NICA01	grammatic
MOMO01	unrespectful	SAMA01	centric
T2	Composition	T2	Interview
TOSA02	restringed	GIRO02	sympatish
BRVA02	living-day	GORE02	ambient; inconvenient
FEBA02	cohabitare	GUMA02	obligue
JUPL02	musulman; respetous	CAPU02	practic; centric
MOMO02	beneficiate	FOTR02	modals
		JIFO02	organisate
T3	Composition	T3	Interview
BRVA03	unpolite	GAPE03	subground
CAMO03	strengthness	PARE03	centric
		VAMO03	centric
		ARVI03	normatives
		JIFO03	incomfortable
		JUPL03	grammatic
		NICA03	centric
		SAMA03	centric

3) Irregular usage that was standardised (not counted as errors)

T1	Original form	Change made
MOLL01	open minded	open-minded
PAAR01	multi-cultural	multicultural
ARVI01	easy going	easy-going
BRVA01	burkha	burqa
CAMO01	life style	lifestyle
JUPL01	lunch time	lunchtime

Appendices

T2	Original form	Change made
GORE02	can not	cannot
GUMA02	can not	cannot
TOSA02	some times	sometimes
CAMO02	may be	maybe
FEBA02	may be	maybe
FEJO02	eversince	ever since
FOTR02	new comers	newcomers
JIFO02	burka	burqa
MUAL 02	may be; viceversa	maybe; vice versa
T3	Original form	Change made
GORE03	can not	cannot
GUMA03	open minded	open-minded
MOLL03	open minded	open-minded
ARVI03	new-comers	newcomers
FEJO03	new comers	newcomers
MUAL03	can not	cannot

Appendix 6. Output of VocabProfile for a text

Hems > VocabProfilers > English (All arrow left to preserve settings) > Output

EDIT-TO-A-PROFILE SPACE

WEB_VP_OUTPUT FOR FILE: Untitled (0.83 kb)

Words recategorized by user as 1k items (proper nouns etc): NONE (total 0 tokens)

	Families	Types	Tokens	Percent
K1 Words (1-1000):	69	85	160	92.49%
Function:	—	—	(96)	(56.48%)
Content:	—	—	(64)	(36.99%)
> Anglo-Sax	—	—	(43)	(24.80%)
<small>(incl. Greco-Latin)</small>	—	—	—	—
K2 Words (1001-2000):	5	5	10	5.78%
> Anglo-Sax	—	—	(4)	(2.31%)
1k+2k				98.27%
AWL Words (academic):	2	2	3	1.73%
> Anglo-Sax	—	—	(1)	(0.58%)
ORP-List Words:	2	2	2	0.80%
70+?	92	173	173	100%

Current profile	
%	Cumul.
92.49	92.49
5.78	98.27
1.73	100.00
0.80	100.00

Words in text (tokens):	173
Different words (types):	92
Type-token ratio:	0.53
Tokens per type:	1.88
Lex density (content words/total):	0.45
<i>Pertaining to orlist only</i>	
Tokens:	173
Types:	92
Families:	76
Tokens per family:	2.28
Types per family:	1.21
Anglo-Sax Index:	%
<small>(Anglo-Sax tokens + function) / (total tokens)</small>	
Greco-Latin/Cognate Index: <small>(inverse of above)</small>	%

For easy editing as MS Word or Excel table - [Edit/print flexibly table](#)

Page menu: [Tokens](#) (With shorter texts; includes AWL, subs) [Types](#) [Families](#) (With VP negative + Anglo-Saxons) [Colour-Coded Text](#)

Note: In the output text, punctuation is eliminated, all figures (1, 20, etc) are replaced by the word number, contractions are replaced by constituent words (won't => will not), type-token ratio is calculated using said constituents, and in the 1k sub-analysis content + function words may sum to less than total (depending on user treatment of proper nouns and program decision to class numbers as 1k although not contained in 1k list); single letters are eliminated except for 'l' and '1'.

Edit-to-a-Profile

Original text [\[Lengthen\]](#) [\[Shorten\]](#) [\[Narrow\]](#) [\[Widen\]](#) [\[Space\]](#) Profiled text Edit, Check with [RE-VP](#), and [SAVE](#)

RCCKT3 3020E

when someone moves to a foreign country it is impossible that he she cannot adopt the customs and way of life of his her new country in my opinion one reason to go abroad is knowing different customs and cultures when someone go abroad he she tries to show to their new friends that he she is interested in their way of life and that he she want to know more and more about it last summer i went to and i learned lots of things from them now i have a good relationship with all the team i went for a work camp and they are trying to learn the language and our customs because this is one characteristic of going abroad meet new ways of life it is true that you have to hold on your own customs but if you try to mix the two or more different cultures you will be able to have a new point of view of the foreign country and your own country too

Current profile	
%	Cumul.
92.49	92.49
5.78	98.27
1.73	100.00
0.80	100.00

when someone moves to a foreign country it is impossible that he she cannot adopt the **customs** and way of life of his her new country in my opinion one reason to go **abroad** is knowing different **customs** and **cultures** when someone go **abroad** he she tries to show to their new friends that he she is interested in their way of life and that he she want to know more and more about it last summer i went to and i learned **lots** of things from them now i have a good relationship with all the **team** i went for a work **camp** and they are trying to learn the language and and our **customs** because this is one characteristic of going **abroad** meet new ways of life it is true that you have to hold on your own **customs** but if you try to **mix** the two or more different **cultures** you will be able to have a new point of view of the foreign country and your own country too

<<< [\[Lengthen\]](#) [\[Shorten\]](#) [\[Narrow\]](#) [\[Widen\]](#) [\[Editing Space\]](#)

Appendix 7. Error classification scheme

Lexical	Description
L1 transfer	Literal translation or direct borrowings of L1 words; false friends
Word choice	Wrong or inappropriate lexical choices (not due to transfer); mistakes with commonly confused words (e.g., <i>make/do</i>)
Non-words	Non-existent words based on L1 forms (e.g., <i>conscient</i>) or resulting from erroneous morphology (e.g., <i>unrespectful</i>)
Phrasal verbs	Wrong use of phrasal verbs; problems with adverbial particles
Fixed expressions	Problems with formulaic language and idioms
All errors	
We computed an overall error count (E/w) by combining total lexical errors with concomitant grammar errors. The concomitant errors consisted of wrong word formation and morphology affecting verbs, nouns, adjectives and adverbs (e.g., <i>broad/broaden</i>).	

Appendix 8. Word Class taxonomy

	Category	CLAWS tag	Description	Examples	
CONTENT WORDS	Adjective	AJ0	Adjective (unmarked)	good	
		AJC	Adjective (comparative)	better	
		AJS	Adjective (superlative)	best	
	Adverb	AV0	Adverb (unmarked)	often	
	Noun	NN0	Noun (neutral)	people	
		NN1	Noun (singular)	country	
		NN2	Noun (plural)	countries	
	Verb	VVB	Lexical verb (base form)	take	
		VVD	Lexical verb (past tense)	took	
		VVG	Lexical verb (gerund)	taking	
		VVI	Lexical verb (infinitive)	take	
		VVN	Lexical verb (participle)	taken	
VVZ		Lexical verb (-s form)	takes		
FUNCTION WORDS	Conjunction	CJC	Conjunction (coordination)	and, or	
		CJS	Conjunction (subordination)	although, when	
		CJT	Conjunction THAT	that	
	Preposition	PRF	Preposition OF	of	
		PRP	Preposition	for, above, to	
	Pronoun	PNI	Pronoun (indefinite)	anybody, none	
		PNP	Pronoun (personal)	you, them	
		PNQ	Wh-pronoun	who, whoever	
		PNX	Pronoun (reflexive)	itself, ourselves	
	Article*	AT0	Article	a, an, the	
	Determiner*	DPS	Determiner (possessive)	your, their	
		DT0	Determiner (general)	these, some	
		DTQ	Wh-determiner	whose, which	
	Modal verb*	VM0	Modal verbs	can, should	
	Auxiliary verbs*	BE	VBB	Verb "BE" (base form)	am, are
			VBD	Verb "BE" (past tense)	was, were
			VBG	Verb "BE" (gerund)	being
			VBI	Verb "BE" (infinitive)	be
			VBN	Verb "BE" (participle)	been
			VBZ	Verb "BE" (-s form)	is
DO		VDB	Verb "DO" (base form)	do	
		VDD	Verb "DO" (past tense)	did	
		VDG	Verb "DO" (gerund)	doing	
		VDI	Verb "DO" (infinitive)	do	
		VDN	Verb "DO" (participle)	done	
		VDZ	Verb "DO" (-s form)	does	
HAVE		VHB	Verb "HAVE" (base form)	have	
		VHD	Verb "HAVE" (past tense)	had	
		VHG	Verb "HAVE" (gerund)	having	
		VHI	Verb "HAVE" (infinitive)	have	
		VHN	Verb "HAVE" (participle)	had	
	VHZ	Verb "HAVE" (-s form)	has		

*Articles, determiners, modals and auxiliary verbs were not calculated as separate lexical categories, yet were included in the global category pertaining to function words. (Adapted from Claws C5 tagset: <http://ucrel.lancs.ac.uk/claws5tags.html>).

Appendix 9. Off-list words (OffL) used in the T2 oral interviews

Education	City	Adjectives
high-school (x27)*	surroundings	boring (x5)
homework (x14)	downtown (x2)	fantastic (x2)
exam(s) (x13)	atmosphere (x5)	talkative
translator(s) (x7)	countryside	compulsory
interpreter (x2)	spotlight	spectacular
professors	outskirts	enthusiastic
timetable (x2)		sociable (x2)
career (x5)	Objects	embarrassing
seminars	television	impersonal
semester	photocopy (x2)	shy
vocabulary	printer	easy-going
(class)mates (x2)	bookstores	conservative
procrastinator	certificate	Unclassified
sessions		prestige
pronunciation	Food	impression
curriculum	salad	flatmate
specializing		hobbies (x3)
oral	Transport	banging
schooldays	subway	placement (x2)
physics	metro	celebrated
linguistics (x2)		

*(x#) indicates the number of times these words occurred in the T2 samples.

