

Accepted version for publication (postprint)

Zaytseva, V., Miralpeix, I., & Pérez-Vidal, C. (2021). Because words matter: Investigating vocabulary development across contexts and modalities. *Language Teaching Research*, 25(2), 162–184.

Because words matter: Investigating vocabulary development across contexts and modalities

Victoria Zaytseva

Universitat Pompeu Fabra, Spain

Imma Miralpeix

Universitat de Barcelona, Spain

Carmen Pérez-Vidal

Universitat Pompeu Fabra, Spain

Abstract

The present study investigates the effects of two subsequent learning contexts, formal instruction (FI) at home and a 3-month stay abroad (SA), on vocabulary acquisition in English as a foreign language (EFL) writing and speaking. Data were collected from 30 Catalan/Spanish learners of English before and after each learning period using a written composition and an oral interview. These samples were examined in terms of quantitative measures of lexical proficiency in the domains of diversity, sophistication, density and accuracy. Native-speaker baseline data were also obtained through the same tasks 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 diversity. FI, however, shows a modest effect on the improvement of oral productive vocabulary and affects namely lexical sophistication.

Keywords

ESL, lexical proficiency, oral-written connection, production modality, speaking, study abroad, vocabulary acquisition, writing

Introduction

Acquiring a second language (L2) in a naturalistic context associated with a study abroad (SA) experience, unlike learning it in a traditional formal instruction (FI) classroom, has often been regarded by the academic community as an important means of speeding up the learning process and boosting a learner's language skills. The growing body of research in SA contexts has provided compelling evidence for this claim and has reported superior benefits of SA over FI settings in speaking skills, particularly, in oral fluency (Lennon, 1990; Llanes & Muñoz, 2009; Mora & Valls-Ferrer, 2012; Serrano et al., 2012; Towell, 2002). The superiority of SA over FI, however, is less clear for written skills, as research has produced mixed results (Barquin, 2012; Freed, So & Lazar, 2003; Sasaki, 2007). Studies assessing both written and oral skills in different types of learning contexts are extremely rare and, apart from fluency, very little is known on the development of other components of written and oral productions, such as the use of vocabulary when one context adds to the other.

The present study intends to fill this void by considering vocabulary development in both writing and speech and seeks to shed new light on the impact of a 3-month SA period experienced after a previous period of FI on lexical acquisition in oral and written production modes. Given that this area of research has received little coverage in the field of SA and SLA to date, our aim is twofold: (1) to track English as a foreign language (EFL) vocabulary development in learners' written and oral production separately and in contrast with native-speaker performance, over time and after two learning environments, FI and SA, experienced one after the other; and (2) to examine the amount of lexical gains achieved as a result of context (FI vs. SA) and modality (Writing vs. Speech). To that end, a combination of quantitative lexical metrics is employed targeting different aspects of vocabulary knowledge: lexical diversity, lexical sophistication, lexical density, and lexical accuracy. To the best of our knowledge, this is the first study to systematically address the effects of FI and SA experiences, one subsequent to the other, on vocabulary in both writing and speech samples produced by the same participants, thus contributing considerably to the limited research in the area. In the following section, we delve into the relationship between writing and speech and examine how written and oral production can be measured from a lexical perspective. We then review the few studies that have compared both modalities across FI and SA learning contexts to provide relevant vocabulary-related data in light of the growing recognition of the importance of vocabulary in language learning.

1 The relationship between writing and speech

The study of language in the written compared with the oral modality has a long research tradition. Early work in the field contrasting the two modes has often demonstrated that writing differs from speaking at the textual level (Sperling & Freedman, 2001). Written language is syntactically more complex and requires more concision, elaboration, and accuracy than spoken language (Sperling & Freedman, 2001; Williams, 2012). Writing is also lexically denser than speaking: it has a higher degree of information packaging (Ure, 1971), which is presented in a more concentrated way than is typical of speech. At

the word class level, the oral dimension generally contains a high proportion of verbs, adverbs, and 1st and 2nd person pronouns, whereas the literate dimension is associated with a greater frequency of noun and prepositional phrases (Biber, 1988). Writing reflects a style of expression which is grammatically more complex and compact than speech, and 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 (Berman & Nir, 2010; Berman & Verhoeven, 2002: 183; Johansson, 2008; Strömquist et al., 2002) across the population of schoolchildren, adolescents, and adults. That is, not only writers construct their texts with greater compactness, they also draw on a more enlarged vocabulary as compared to when they speak. On the other hand, written language also differs from spoken language in terms of the cognitive demands and communicative conditions associated with each modality. While speech typically allows feedback and is processed under strict online constraints (Strömquist et al., 2002), these constraints can be lifted in written production. In other words, the inherent benefit for the written medium in terms of time, planning, and revision (Williams, 2012), which allow for greater lexical precision and complexity, contrasts with the pressure associated with the spoken medium, where the primary concern is to maintain the flow of conversation.

In light of such differences in communicative circumstances and information-processing constraints, it is not surprising that research on vocabulary has predominantly focused on the analysis of written rather than spoken language (because among other reasons, spoken language is more difficult to analyse) and that there is more published work on the assessment of writing than on the assessment of speaking (Read & Nation, 2006). Lexical proficiency, however, can be analysed in both writing and speech using a number of comprehensive measures, i.e. measures that take account of the whole vocabulary content of the test-taker's response in writing and speaking tasks. These are different from discrete measures, which aim at assessing just specific vocabulary items (e.g. in a multiple-choice test). When analysing student's oral/written productions from a lexical perspective, four domains are usually assessed in the literature, which are diversity, sophistication, density and accuracy (Read, 2000). These are well-known components of lexical richness as they show whether learners vary the words they use (diversity), the sort of words they use in terms of frequency (sophistication) and word-class (density) and if students make errors in these words (accuracy).

2 Previous research on vocabulary across learning contexts and modalities

It has long been acknowledged in the field of both L1 and L2 learning that the “‘nitty-gritty” of the lexicon’ (Singleton, 1999: 4) rather than syntax is the major challenge of language acquisition. The importance of lexical knowledge in learning a language has been consistently corroborated from different angles. First, vocabulary knowledge plays an integral role in spoken production theoretically (Kormos, 2006; Levelt, 1989) and empirically (De Jong et al., 2012; Koizumi & In’nami, 2013; Uchihara & Saito, 2019); second, it is a key indicator of writing quality (Alderson, 2005; Engber, 1995; Laufer & Nation, 1995), general linguistic proficiency (Zareva et al., 2005), and academic achievement (Daller et al., 2003); and third, lexical errors are graded as the most severe ones in both written (Agustín Llach, 2011; Santos, 1998) and oral (Ellis, 1995) communication.

The acquisition of vocabulary has also been examined in studies dealing with different contexts of learning. Many authors coincide in finding positive effects of SA programmes on receptive vocabulary knowledge – understanding the meaning of words with a help of a stimulus (generally in reading and listening) – and report that SA is highly beneficial for growth in receptive vocabulary in L2 English (Milton & Meara, 1995; Ife, Vives Boix & Meara, 2000); L2 Japanese (Dewey, 2008); and L2 Spanish (Jiménez Jiménez, 2010) (for further review of these studies see Zaytseva, Pérez-Vidal & Miralpeix, 2018). However, very little is known about how SA affects productive vocabulary knowledge – ability to speak or write the word in the appropriate context (Nation, 1990), and even less research exists on the impact of SA on productive vocabulary comprising both written and oral production modalities, as noted by Zaytseva, Pérez-Vidal & Miralpeix (2018).

One of the first studies to consider the effects of SA on both written and oral production modes is the study by Freed et al. (2003). Although not focused exclusively on vocabulary knowledge, Freed's team examined gains in written and spoken fluency in American L2 undergraduates learning French: 15 of them at home (AH) and another 15 abroad. The authors found that SA was more beneficial to the development of learners' oral fluency than to their written fluency; yet, the SA group's post-test compositions were slightly denser in lexical use (though not statistically significant) than their pre-test compositions, unlike the post-compositions written by the AH group, which did not show this change. Another study pointing to SA as the context that favours oral skills more than written skills is the one by Llanes et al. (2011). Focusing on the development of English from the perspective of complexity, accuracy, and fluency (CAF), Llanes and colleagues (2011) found that SA had a significant impact on the improvement of oral lexical diversity when compared to written in 24 adult Spanish learners of English spending a semester in the UK. Similarly, Serrano et al. (2012), examining 14 Spanish adult learners studying English in the UK over a course of a year, found that while one semester abroad was already enough for significant progress to occur in oral lexical diversity, improvement in the same domain for written production did not become visible until the second semester.

Somewhat different results can be found in Pérez-Vidal & Juan-Garau (2011) with data from the SALA project. With vocabulary acquisition not being the object of study, the authors aimed to analyse the differential effects of FI and SA on CAF oral and written development. In their analysis of writing and speech samples of Spanish/Catalan adult learners of English through sequential FI and SA learning experiences, they reported significant improvement in written lexical diversity as a result of the SA period, while no significant changes were detected in lexical diversity of student speech productions. The same findings were reported in another SALA-based study a year later (Pérez-Vidal, Juan-Garau, Mora & Valls-Ferrer, 2012), where 29 Spanish/Catalan participants showed improvement in written lexical diversity but not in oral lexical diversity after the SA experience in the English-speaking country. A summary of the existing studies and results can be consulted in Table 1.

As can be seen from the aforementioned literature review, previous research in the area of both oral and written vocabulary knowledge and SA has produced inconclusive findings. One of the reasons for such mixed results is that there is a great variability

Table 1. Studies analysing the effects of stay abroad (SA) on vocabulary in both written and oral modalities.

	Freed et al. (2003)	Llanes et al. (2011)	Pérez-Vidal & Juan-Garau (2011)	Pérez-Vidal et al. (2012)	Serrano et al. (2012)
Design	Cross-sectional	Pre-test/post-test	Longitudinal	Longitudinal	Pre-test/post-test
Participants	15	24	20 (spoken data); 35 (written data)	29	14
L1	English	Spanish	Spanish/Catalan	Spanish/Catalan	Spanish
L2	French	English	English	English	English
Duration	4 months	3 months	3 months	3 months	3 months to 8 months
Instrument (speaking)	Oral proficiency interview	Cartoon prompt Picnic	Role-play task	Semi-guided interview	Cartoon prompt Picnic
Instrument (writing)	Narration in Past Tense	Descriptive essay	Argumentative essay	Argumentative essay	Descriptive essay
Measure(s)	Lexical density (in writing)	Lexical diversity GI	Lexical diversity GI	Lexical diversity GI	Lexical diversity GI
Results (speaking)	Not applicable	Pre-SA < Post-SA	AH = SA	AH = SA	Pre-SA < Post-SA
Results (writing)	AH = SA	Pre-SA = Post-SA	AH < SA	AH < SA	Pre-SA < Post-SA*

Note. *Significant improvement in written lexical diversity in Serrano et al. (2012) did not occur after the first semester abroad (3 months) but only after the second semester abroad (8 months). AH = at home. GI = Guiraud's index. SA = study abroad.

across studies in terms of design, programme duration, the number of participants, and the instruments used. The existing research is particularly scanty and limited from the vocabulary perspective, as the only lexical measure that the studies have so far consistently analysed is the Guiraud's index (GI). However, several studies have repeatedly emphasized the need to assess different aspects of lexical performance to gauge development (e.g. Bulté & Housen, 2014; Wolfe-Quintero et al. 1998). The paucity of SA research on vocabulary encompassing both modes of production highlights the contribution that our study makes to the subfield in question, as, to the best of our knowledge, no study to date has offered a comprehensive analysis of lexical features in both writing and speech samples produced by the same participants in different learning circumstances.

The study

The study presented here belongs to a larger project, the SALA project, which has analysed in depth the effects of two learning contexts, FI and SA – the latter within the European Community Action Scheme for the Mobility of University Students (ERASMUS) –, on the progress made by advanced EFL learners across a full range of language skills and individual factors (for a detailed description of the project, see Pérez-Vidal, 2014). The present study aims to investigate the corresponding effects on EFL learners' lexical ability in both production modes: writing and speaking. First of all, we seek to measure learners' oral and written lexical development, longitudinally, after both FI and SA, and in comparison with native-speaker baseline performance. We assess EFL oral and written samples in terms of quantitative measures of lexical diversity, sophistication, density and accuracy. Second, we aim to compare the amount of gains obtained after each learning context (FI vs. SA) and in each production modality (Writing vs. Speaking), in order to determine which context leads to the largest gains, on the one hand, and which production mode is associated with these gains, on the other. In order to accomplish these objectives, we addressed the following research questions.

1 Research questions

1. Does learners' lexical proficiency in written and oral production improve significantly over time, and after either the FI or SA learning contexts to approximate native-like norms at any point over the observation period?
2. In which context (FI vs. SA) and modality (writing vs. speaking) do learners accrue greater lexical gains?

2 Participants

The participants in this study were 30 randomly selected Catalan–Spanish learners of English (NNSs), of which 24 were female and 6 male. They were undergraduate students, majoring in a language specialization degree in Barcelona and had an advanced level of English (a B2.1 level with the Common European Framework of Reference for Languages), which they had to accredit upon enrolment in the programme. Students'

average age at the beginning of the study was 17.93 years and the range 17–21, with the majority (93.3%) having enrolled in the studies at the age of 17–18.

During the first two quarters of their first academic year, students had to attend two English language university courses, corresponding to the period of FI. These courses consisted of lectures and seminar sessions with English as the medium of instruction at all times. 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. By the end of FI, students had received a total of 80 hours of in-class instruction with no further FI in the third trimester.

As a curriculum requirement, during the second academic year of students' degree, they were required to go on a compulsory three-month ERASMUS exchange to an English-speaking country. The vast majority 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. 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.

Baseline data from 27 native speakers (NSs) of English were also collected for comparison purposes. The NSs were undergraduate students (77.8% female) participating in an exchange programme at different Spanish universities. They were originally from the US (15), the UK (6) and Ireland (4). NS participants afforded highly similar data to that of the NNS sample in terms of age ($M = 20$ years) and education, and were used in the absence of a better means of comparison with advanced-level learners but never as the idealized golden yardstick for language knowledge (Ortega, 2014) nor as the measure of ultimate attainment in the L2 (Cook, 1999).

It should be mentioned that 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 contribution to the study.

3 Data collection

In order to evaluate the linguistic benefits of two subsequent learning conditions (FI followed by SA), we followed a repeated measures within-participants design, in which learners' 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 overcomes the problem of between-participant variability (Milton & Meara, 1995) and bypasses the question of whether or not groups (experimental and control) are validly comparable. The data collection was conducted longitudinally over a 15-month

Table 2. Data collection times (adapted from Pérez-Vidal, 2014).

	Year 1			Year 2		
Trimester	I	2	3	I	2	3
Context	FI		—*	SA		
Testing times	(T1)	(T2)		(T3)		

Note. *There was no formal language instruction during the third quarter of the first academic year. FI = formal instruction. SA = study abroad.

period, comprising two academic years. The first data collection took place at time 1 (T1) when students entered the programme; then at time 2 (T2), after six months (80 hours) of FI in their home university, and before SA (with the third trimester of no formal English language instruction and the summer holidays in between); and finally at time 3 (T3), after students’ return from their 3-month SA exchange in English-speaking countries, that is 15 months after T1. Following this within-participants design, T1 served to determine participants’ initial level, while T2 was both a post-test for T1 and a pre-test for T3, and T3 was a post-test for T2. Table 2 displays the research design involving the three data collection times (T1, T2, and T3).

4 Task and data analysis

The instruments used to gather the participants’ writing and speech samples were an argumentative essay and a semi-guided interview. Students were required to write a composition in response to the prompt: ‘Someone who moves to a foreign country should always adopt the customs and way of life of his/her new country’. To that end, they were given a ruled, double-sided sheet of paper and were allowed 30 minutes to complete the task by hand without the use of a dictionary or any outside materials. Although text length was not specified in the instructions, all participants finished writing their essays within the allotted time period without exceeding a single double-sided sheet. The decision to examine productive vocabulary through free written expression was made in light of previous literature finding written compositions as highly reliable for assessment in higher education and sufficiently complex to discriminate between different language profiles (Laufer, 1998). They also provide valuable data when examining L2 vocabulary knowledge and lexical errors (Agustín Llach, 2011; Santos, 1988), since writing quality is closely linked to effective vocabulary use (Engber, 1995; Laufer & Nation, 1995). We decided to maintain the same prompt at each testing time so as to avoid possible variation in vocabulary use since changes in topic have been reported to influence L2 written narrative and result in uncontrolled variation in lexical selection (Laufer, 1998).

The oral task consisted of a semi-guided interview, where students were required to act as both interviewers and interviewees (Appendix 1). They were assigned ‘Student A’ and ‘Student B’ roles randomly and were given the interview questions on the topic: *University Life*. First, student A asked student B the questions one at a time, and then they switched roles. The oral interviews were recorded digitally in a quiet room and were approximately 5 minutes long. A research assistant, though present in order to ensure the

quality of the digital recordings, did not intervene unless it was necessary to inform the test-takers of the approaching time limit. Participants were instructed to perform the task as if the researcher was not there, favouring a relaxed atmosphere and a spontaneous interaction. The reason for focusing the conversation around students' life at university was that of topic familiarity. Talking in an L2 about familiar information (i.e. one's personal experience) typically entails a more fluent and accurate production and prevents speakers from running out of ideas (Foster & Skehan, 1996; Skehan & Foster, 1997). Since our goal was to elicit as natural and close to real-life speech as possible, having learners engage in a cognitively undemanding task was a priority. Only the answers to the interview questions were included in the analysis.

In light of previous SA research, which has approached lexical development in a superficial and fragmented way (e.g. most studies examining L2 oral and written samples have limited their analysis to one measure only, generally as part of the domain of lexical complexity within the CAF triad, as summarized in Table 1, our study considers vocabulary as a multifaceted construct. In our analysis, we adopt Read's (2000) list of lexical features which should be used to assess L2 production and that would reflect learners' lexical skills:

- lexical variation or diversity, defined as the variety of words used in oral or written production;
- lexical sophistication, defined as the proportion of low-frequency words to high-frequency, everyday vocabulary;
- lexical density, described as the proportion of content words (nouns, verbs, adjectives, and adverbs) to the total number of words; and
- lexical accuracy, that is, occurrence of errors in vocabulary use.

In our study, participants' written and oral samples were analysed quantitatively in these four domains (see Table 3), following Read (2000), and resulted in a total of 117 written compositions (90 by NNSs and 27 by NSs) and 115 oral interviews (90 by NNSs and 25 by NSs). A full corpus of 27,774 words for compositions and 30,370 words for interviews was transcribed following CHAT transcription conventions of the CLAN software (MacWhinney, 2000; for examples, see Appendix 2) and then computed for each domain in question:

- Lexical diversity: To calculate lexical diversity, our study included two measures: GI and D. GI is calculated by using the square root for relating types and tokens, and D uses a curve-fitting approach to take into account how the type-token ratio changes in a text (for details on how D operates, see Malvern & Richards, 1997, 2000; see Meara & Miralpeix, 2017). In our study, type and token counts were calculated using CLAN and then the formula $\text{types}/\sqrt{\text{tokens}}$ was computed to arrive at GI. D values were obtained using the D_Tools software (Meara & Miralpeix, 2017).
- Lexical sophistication: To calculate lexical sophistication, Laufer and Nation's (1995) Lexical Frequency Profile (LFP) was used. LFP classifies learners' vocabulary into four frequency levels, from the most common 1,000 words (1k) to less

Table 3. Measures used per domain.

Domain	Measure	Abbreviation
Lexical diversity	Guiraud's Index	GI
	D	D
Lexical sophistication	First 1,000 words	1k
	Second 1,000 words	2k
	Academic word list	AWL
	Off-list	+2k
Lexical density	Content density	CW/w
	Function density	FW/w
	Adjective density	Adj/w
	Adverb density	Adv/w
	Noun density	N/w
	Verb density	V/w
Lexical accuracy	Vocabulary errors per word	VE/w
	Spelling errors per word	SpE/w*

Note. *Spelling errors were calculated in written production, only.

frequent vocabulary (i.e. 1k, 2k, Academic Word List, +2k). This was performed using the VocabProfile programme (Cobb, 2002) and the results of each frequency level were recorded in percentages.

- Lexical density: Lexical density was calculated as total number of content words divided by total number of words (CW/w) and minor variants of these content words were also calculated: that is, the total number of adjectives (Adj/w), adverbs (Adv/w), nouns (N/w) and verbs (V/w) calculated separately in relation to the total number of words. The density of function words was also calculated: total number of function words (consisting of articles, pronouns, prepositions, conjunctions, and determiners) were divided by total number of words (FW/w). The computation of different parts of speech was done using the CLAWS part-of-speech tagger (see Garside & Smith, 1997).
- Lexical accuracy: To calculate lexical accuracy, vocabulary errors were marked in the transcripts and then its total number was divided by total number of words (VE/w), following the error classification scheme from previous SALA-based publications (Barquin, 2012; Lara, 2014; Pérez-Vidal et al., 2012). Vocabulary errors consisted of problems with wrong lexical choice, non-words, and L1 transfer (for details, see Zaytseva, 2016). Errors were counted using CLAN and then entered into SPSS for analysis.

It must be noted that out of the large array of measures that can be assessed in each lexical dimension, only the most appropriate ones were considered for analysis in light of the context of learning and the students' level (for instance, lexical diversity measures that could overcome text-length dependency shortcomings, or lexical frequency profiles capturing the development of lexical sophistication). In the following section, we present the results addressing each of the two research questions in turn.

Table 4. Descriptive statistics for written productive vocabulary: means obtained at T1, T2 and T3 for learners ($n = 30$) and NS ($n = 27$).

Writing		T1	T2	T3	NSs
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)
Sophistication	1k	91.30 (2.36)	91.76 (2.93)	91.26 (2.48)	87.23 (3.61)
	2k	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)
	+2k	1.06 (0.81)	0.88 (0.85)	1.11 (1.09)	2.85 (1.78)
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)
Accuracy	FW/w	53.16 (3.07)	52.32 (3.82)	52.31 (3.45)	48.35 (3.04)
	LexE/w	2.28 (1.50)	1.88 (0.85)	1.40 (0.98)	0.07 (0.15)
	SpE/w	0.76 (0.63)	0.87 (0.79)	0.69 (0.88)	0.34 (0.42)

Results

1 Vocabulary development in oral and written production

The descriptive statistics of the measures used in the analysis of the written and oral productive vocabulary are presented in Table 4 (for writing) and Table 5 (for speaking). Participants' mean values obtained at T1, T2, and T3 for learners and native-speaker baseline data (NSs) are presented together with standard deviation (in parentheses).

To determine whether learners demonstrated statistically significant growth in oral and written productive vocabulary over time, after FI and SA, and in comparison with NS performance, repeated measures ANOVAs were performed with Bonferroni post-hoc tests (see Results in Table 6). To further examine whether these changes took learners nearer to native-speaker norms, a series of independent *t*-tests were conducted on NS data and learner data at all three data collection times for each production modality. The alpha level for all tests was set at .05.

As can be seen in Table 6, the results of the main ANOVAs for written production revealed a significant effect of time on measures of lexical diversity (GI), adverb and noun density (Adv/w, N/w), and lexical accuracy (LexE/w). The largest main effect was seen for lexical diversity (GI) ($p = .000$), with partial eta squared (η^2p) of .51. We did not find statistically significant differences for any of the variables associated with lexical sophistication. Follow-up Bonferroni pairwise comparisons further revealed that changes in these measures occurred gradually (from T1 to T3) but were not attributable to any context in particular, except for the Guiraud's index of lexical diversity, which improved significantly between T2 and T3. That is, learners used a significantly greater variety of words in their compositions and this improvement occurred as a result of their SA experience, following FI. In the domain of lexical density, a gradual growth was found in the

Table 5. Descriptive statistics for oral productive vocabulary: means obtained at T1, T2 and T3 for learners (*n* = 30) and NS (*n* = 25).

Speaking		T1	T2	T3	NSs
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	1k	94.49 (1.87)	93.35 (1.56)	94.10 (1.71)	91.68 (2.02)
	2k	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)
	+2k	1.25 (0.97)	1.81 (1.01)	1.28 (0.68)	2.49 (0.99)
Density	CW/w	34.74 (3.30)	34.99 (2.48)	35.60 (2.62)	39.37 (2.70)
	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)
Accuracy	LexE/w	1.72 (1.00)	1.30 (0.80)	0.86 (0.71)	0.06 (0.17)

Table 6. RM-ANOVAs: Longitudinal changes with context-specific (FI or SA) differences in written and oral productive vocabulary (all *F*s = [2, 28]).

	Writing					Speaking				
	F	<i>p</i>	η^2p	FI	SA	F	<i>p</i>	η^2p	FI	SA
GI	14.583	.000	.51	.062	.000	.166	.848	.01	—	—
D	2.478	.102	.15	—	—	1.795	.185	.11	—	—
1k	.510	.606	.04	—	—	4.090	.028	.23	.034	.131
2k	1.376	.269	.09	—	—	1.345	.277	.09	—	—
AWL	2.137	.137	.13	—	—	.626	.542	.04	—	—
+2k	1.141	.334	.08	—	—	3.880	.033	.22	.050	.061
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	—	—
LexE/w	4.819	.016	.26	.390	.096	10.482	.000	.43	.052	.056
SpE/w	.367	.696	.03	—	—	N/A	N/A	N/A	N/A	N/A

measure of adverb density (Adv/w), reaching statistical significance (*p* = .028), although not in favour of any specific learning context, indicating that learners produced significantly more adverbs per word in their compositions and that both FI and SA proved beneficial to the improvement in this area. Significant changes were also found in the noun density measure (*p* = .024), showing a continued decrease in the proportion of nouns in the text. Finally, in the domain of lexical accuracy, there was a linear

longitudinal decrease in the mean number of lexical errors per word (LexE/w), which reached significance ($p = .016$) but again was not specific to any context in particular, suggesting that both contexts favoured greater accuracy in vocabulary use. We did not find any significant changes in learners' spelling accuracy.

ANOVA results for oral production showed a significant effect of time on two measures of lexical sophistication (1k and +2k), adverb density (Adv/w), and lexical accuracy (LexE/w). The largest effect size was observed for accuracy ($p = .000$), with η^2p of .43. We did not find statistically significant differences in the domain of lexical diversity, unlike in writing. Subsequent pairwise comparisons showed that while changes in adverb density (Adv/w) reached significance after SA, changes in the 1k and +2k frequency levels were significant after FI. In other words, learners were producing a significantly larger proportion of adverbs after SA than after FI, but the sophistication in learner vocabulary use improved significantly after FI and not after SA, this time. As for accuracy, there was a gradual decrease in the number of lexical errors in learner speech, yet this improvement did not correspond to any learning context in particular. This finding suggested that although learners substantially improved in lexical accuracy, both learning conditions were equally beneficial for this area.

To determine whether learners' lexical development approached NS norm, we ran a series of independent t -tests on learner scores obtained at each data collection time and NS data for each production modality (participants' mean scores and standard deviation are presented in Tables 4 and 5 above). In written production, the results of the t -tests revealed that despite scoring below NS patterns on average, learners improved over time and converged with NSs by the end of the study on a number of measures: the GI of lexical diversity [$t(55) = -.637, p = .527$], the AWL of lexical sophistication [$t(55) = -1.684, p = .098$], and accuracy in spelling (SpE/w) [$t(42.6) = 1.961, p = .056$]. This convergence with native speaker values occurred at T3, that is, after the SA experience and suggested that having differed significantly at T1 and T2, learners coalesced with NSs in these measures at T3 to the degree that they were no longer statistically distinguishable. A trend of this kind indicated that at least some aspects of written vocabulary use, such as lexical diversity, sophistication in academic vocabulary or spelling accuracy were being mastered in more appropriate, target-like ways after the SA period. In oral production, t -tests results showed a less prominent approximation to the target-like norms. The only growth toward NS usage was registered in adverb density (Adv/w), which approached NS values at T3 to the point at which learner productions no longer varied significantly from those of NSs [$t(53) = -.707, p = .483$]. In other words, the proportion of adverbs in learner speech was the only measure that represented target-like behaviour after the SA exchange.

2 Vocabulary gains across learning contexts and production modalities

To assess participants' vocabulary gains across contexts and modalities, we calculated the amount of gains obtained during FI (FI gains) by subtracting T1 scores from T2 scores ($T2 - T1$), and the amount of gains made during SA (SA gains) by subtracting T3 scores from T2 scores ($T3 - T2$). We first conducted a set of paired-samples t -tests to compare FI gains to SA gains for each modality separately (FI gains vs. SA in gains in writing; FI gains vs. SA in gains in speaking) and then ran another series of t -tests to

compare written vocabulary gains to oral vocabulary gains for each learning context in turn (Writing gains vs. Speaking gains in FI; Writing gains vs. Speaking gains in SA).

Regarding learners' vocabulary gains after SA following FI, the results of the *t*-tests showed that greater gains were made after SA than after FI in written lexical diversity indexes, as there was a statistically significant difference in GI between FI gains ($M = -.33$, $SD = .75$) and SA gains [$M = .76$, $SD = .76$; $t(29) = -4.825$, $p = .000$, $r = .67$] and in D between FI gains ($M = -3.84$, $SD = 14.17$) and SA gains [$M = 5.02$, $SD = 13.32$; $t(29) = -2.201$, $p = .036$, $r = .38$]. This suggested that learners' capacity to use more varied vocabulary was mainly due to the period abroad following the home classroom instruction. Conversely, larger gains were made after FI preceding SA in oral lexical sophistication and noun density, as there was a statistically significant difference in 1k between FI gains ($M = -1.15$, $SD = 2.32$) and SA gains [$M = .75$, $SD = 1.96$; $t(29) = -2.883$, $p = .007$, $r = .47$], in +2k between FI gains ($M = .57$, $SD = 1.24$) and SA gains [$M = -.54$, $SD = 1.20$; $t(29) = 2.835$, $p = .008$, $r = .47$], and in N/w between FI gains ($M = .71$, $SD = 3.04$) and SA gains [$M = 1.09$, $SD = 2.18$; $t(29) = 2.417$, $p = .022$, $r = .41$]. That is, as learners replaced the most frequent words (1k) in English with the least frequent (+2k) vocabulary, they also incorporated significantly more nouns into their speech, moving toward native-like usage. We did not register significant differences for the size of gains in the remaining measures. In light of these *t*-test results, we can consider that SA yielded greater benefits than FI for the development of written lexical diversity, as seen in the gains made in both GI and D indexes. FI, however, was more beneficial to learners' oral productive vocabulary than SA, as confirmed through the gains obtained in two measures of lexical sophistication and noun density.

As for the size of gains in relation to production modality, *t*-tests results showed that gains obtained in oral production were significantly superior to those obtained in written production in terms of lexical diversity and sophistication after the period of FI but not after SA. More specifically, there was a statistically significant difference in GI between written gains ($M = -.33$, $SD = .75$) and oral gains [$M = .05$, $SD = .51$; $t(29) = -2.270$, $p = .031$, $r = .39$], in 1k between written gains ($M = +.46$, $SD = 2.71$) and oral gains [$M = 1.15$, $SD = 2.32$; $t(29) = 2.341$, $p = .026$, $r = .40$], and +2k between written gains ($M = -.18$, $SD = 1.04$) and oral gains [$M = .57$, $SD = 1.24$; $t(29) = -2.490$, $p = .019$, $r = .42$]. Conversely, larger gains were achieved in written modality than in oral as a result of SA (and not after FI) in the GI measure of lexical diversity, with written gains ($M = .76$, $SD = .76$) and oral gains [$M = -.02$, $SD = .54$; $t(29) = 5.099$, $p = .000$, $r = .69$]. In sum, while learners seemed to gain more in oral productive vocabulary than in written during classroom FI (as represented by significant gains in sophisticated vocabulary and lexical diversity in their speech), the pattern was the reverse during SA, with learners making more gains in written productive vocabulary than in oral (based on the significant increase in their lexical diversity).

Discussion

The first aim of this study was to evaluate changes in oral and written productive vocabulary of Spanish–Catalan learners of English before and after FI and SA learning periods and examine whether these changes bring learners closer to native-like competence. We considered learners' productive vocabulary in free writing and speech and analysed it in

terms of quantitative measures in the domains of lexical diversity, sophistication, density, and accuracy.

We found that learners' productive oral and written vocabulary abilities in two consecutive learning environments do not develop in the same ways. In written production, we found a significant main effect of time on the Guiraud's index of lexical diversity (GI), adverb and noun density (Adv/w, N/w), and lexical accuracy (LexE/w). Results also revealed that while both FI and SA learning contexts contributed to increased adverb density and lexical accuracy, integrating a SA experience into the institutional curriculum at the students' university is particularly beneficial for the development of lexical diversity. We did not find any significant changes in the domain of lexical sophistication. Further analyses revealed that despite scoring below NS patterns on average, learners improved over time and converged with NSs on a number of measures: lexical diversity (GI), lexical sophistication (AWL), and lexical accuracy (SpE/w) and this convergence occurred at T3, that is, after the SA experience. The results for oral production, in contrast, presented a rather different picture. There were significant changes over time for adverb density (Adv/w), lexical sophistication (1k, +2k), and lexical accuracy (LexE/w), with adverb density showing a significant improvement especially after the SA exchange. While lexical accuracy improved after both learning periods, the progress made in sophisticated vocabulary occurred as a result of classroom FI. We did not find any significant changes in the domain of oral lexical diversity. Follow-up analyses also revealed that learners converged with the native speakers only on one measure (adverb density) at T3, corresponding to the SA experience. Overall, we can say that SA had a less prominent role on the development of oral productive vocabulary unlike what was found for written production, and that FI proved especially advantageous for the sophistication of L2 lexical usage. Our results pointing to the positive impact of SA on written production support previous studies assessing writing in an analogous population (Barquin, 2012; Pérez-Vidal et al., 2012; Serrano et al., 2012), where learners considerably improved in lexical diversity after the experience abroad and became more target-like in other areas (sophistication, spelling accuracy). A caveat needs to be mentioned because significant improvement after the SA experience can be equally attributed to the cumulative effects of both learning contexts (first, FI and then SA), in which progress was already happening at T2 but did not exteriorize itself until T3, where it was found to be significant, potentially as a result of the succession of both learning contexts. This observation follows the view that FI and SA may not be in opposition (DeKeyser, 2007), but that rather learners must play on their abilities to transfer knowledge acquired in one context to the new learning contexts. However, it should be noted that the SALA project data collection was organized around the trimester system that the university follows, with English language instruction in quarter 1 and quarter 2, and with no further formal language study in quarter 3. Had the institutional context been different (for example, if half the students had been required to complete the two learning contexts in the reverse order -first SA, then FI-), we would have been able to evaluate the impact of each context in the opposite direction, as also noted by Barquin (2012).

On the other hand, the lack of progress after SA in oral productive vocabulary, with the exception of adverb density, corroborates previous SALA findings for the same interview task and a role-play task (Lara, 2014; Pérez-Vidal & Juan-Garau, 2011; Pérez-Vidal et al., 2012), in which SA did not do much to enhance learners' lexical knowledge. FI, in turn,

was shown to account for greater progress in the area of lexical sophistication. Specifically, learners seemed to replace the most frequent vocabulary (1k) by the least frequent (+2k) words in their speech after the FI period. One possible interpretation for this finding can be related to the fact that form-focused instruction with academic language and textbooks used within a university degree context propelled a more sophisticated vocabulary use than the naturalistic SA setting, which boosts rather a high-frequency everyday vocabulary. On closer inspection, the list of learners' sophisticated vocabulary at T2 reveals a great proportion of lexis related to education (*exam, seminar, professor, homework, semester, high-school, session, curriculum, translator, interpreter*), everyday objects (*printer, photocopy, television, bookstores, certificate*), city places (*countryside, downtown, outskirts, surroundings*), or personality adjectives (*talkative, easy-going, shy, sociable*). However, this finding must be regarded with caution, since most of these words are cognates and have a strikingly similar form in both Catalan and Spanish languages (e.g. *exam* – *exàmen/examen*, *seminar* – *seminari/seminario*, *professor* – *profesor/professor*, etc.), easily accessible to our learners. With regard to adverb density as the only measure that showed significant improvement after SA to converge with native-speaker values, our results are in alignment with Trenchs-Parera (2009), who reported significant post-SA increases in lexical fillers following an analogous group of SALA participants. The lexical fillers in her study were mostly represented by lexicalized phrases (e.g. 'you know', 'it's like', 'and stuff') and single words, generally adverbs (e.g. 'well', 'so', 'like'), adding fluency and richness to L2 speech. Trenchs-Parera's results concur with the findings of Mehnert (1998), showing that fluency was strongly associated with lexical density and that more fluent speakers tended to produce more content words.

The second aim of the study was to explore differences in vocabulary gains in one learning context followed by another (FI vs. SA) and in different production modes (Writing vs. Speaking). When results from the two subsequent learning contexts were examined, both FI and SA were shown to induce substantial gains on written and oral vocabulary, though with noticeable differences. It appears that the effect of adding a SA period was especially beneficial for the development of written productive vocabulary, as greater gains were obtained in lexical diversity (GI and D) at T3 than at T2. However, FI proved favourable to the advancement of oral productive vocabulary, as greater gains were found in lexical sophistication (1k and +2k) and noun density (N/w) at T2 than at T3.

When the two modalities (oral-written) were contrasted, a similar picture emerged. More gains were found for oral productive vocabulary than for written during the period of FI based on the evidence of improvement in lexical diversity and sophistication measures (GI, 1k, +2k). However, superior gains in written productive vocabulary over oral were reported during the SA experience, as seen through the progress in lexical diversity (GI). Several possible explanations can be offered in this respect. First of all, the positive effect of SA on written productive vocabulary may be explained on the basis of the writing practice while abroad since not only were learners attending lectures at their host university, they were also submitting academic assignments and keeping personal diaries at the request of their AH university. A practice of this kind together with ample opportunities for interaction in the new community might have enhanced their written lexical abilities already cultivated in the grammar-focused English classes. Second, the positive impact of FI, particularly for oral lexical sophistication, can be attributed to the role of

teachings materials. Although speaking may have not been the most practised skill in our participants' formal classroom setting AH, the teaching methods that prevailed at their university, with English as the medium of instruction at all times used in a predominantly academic register, were equally beneficial for their oral lexical competence. Our findings fit well with prior work conducted via the SALA project (Pérez-Vidal et al., 2012), indicating that progress did not take place in a parallel fashion across the different domains of linguistic proficiency, and that oral lexical complexity improved at home and written lexical complexity abroad, with both learning contexts (FI + SA) complementing each other. As already mentioned, the observation that classroom instruction and learning abroad are not necessarily in opposition was made by DeKeyser in 2007 and later refined in the *Combination and Complementary of Contexts Hypothesis* (Pérez-Vidal, 2014; Pérez-Vidal & Juan-Garau, 2011), suggesting that the declarative and procedural linguistic knowledge gained from classroom instruction may result in automatized language use with immersion abroad. That is, although progress became apparent abroad, both learning experiences may have complementary effects by equipping learners with a solid lexical foundation that would bear fruit later on.

On the other hand, we should not forget the modality-related differences that might have impacted our results as well. For example, the lack of improvement in lexical diversity in speaking unlike in writing can be explained in terms of the task planning factor, in the sense that the opportunity for planning, revision and editing required to achieve precision in vocabulary choice is only afforded in the written mode (Biber, 2009; Williams, 2012). As previously discussed, the different processing constraints and communicative conditions associated with each mode of language production have a marked effect on the lexicon used in texts differing in modalities. 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' (Strömquist et al., 2002: 47), in which the immediate priority is to maintain the flow of conversation rather than search for synonyms to avoid repetition. A wider range of vocabulary is much more accessible in the written mode, where the on-line constraints are relaxed. To finish, the advantage of FI over SA regarding sophisticated vocabulary can be interpreted as the product of exposure to classroom and degree-related vocabulary, such as the aforementioned 'high-school', 'homework', 'exam', 'translator', 'interpreter', 'professor', among others, propelling growth in this type of words. Although falling short of finding comparable studies to document changes in lexical sophistication across contexts and modalities, further research in this domain must not be misplaced.

Conclusions

The present study has explored the development of productive vocabulary in written and oral modalities by advanced EFL learners engaging in subsequent learning experiences (FI and SA). Specifically, the study contributes to filling the gap concerning written and oral lexical performance in the line of research that investigates the effects of SA and FI learning contexts. The results of the study have provided evidence for different developmental trends depending on the learning context and the production modality under scrutiny. In written production, we found that improvement in EFL vocabulary occurred predominantly after SA with a number of significant trends toward target-like values in

terms of lexical diversity, academic vocabulary and spelling accuracy. In oral production, overall development was less pronounced and affected lexical sophistication, which improved after FI, and adverb density, which improved after SA to equal target-like language. In light of these findings, it would seem advisable for pedagogic practice to pay special attention to EFL academic writing skills (even at advanced levels) and, in particular, to help students get a stronger grasp of what types of lexical features make high standard academic texts (e.g. how to avoid repetition or use more precise vocabulary). More specifically, it would be advisable to introduce adverbial expressions and discourse markers early in the classroom so that students can acknowledge the importance of a fluent speech and avoid empty pauses and hesitation when practicing speaking. Our findings also corroborate several claims: (1) both contexts may supplement each other, lending support to DeKeyser's (2007) and Pérez-Vidal's (2014) hypothesis of the complementarity and combination of learning contexts; and (2) the different teaching materials specific to each context may play a role in the acquisition process, potentially influencing the vocabulary outcomes (Pérez-Vidal, 2015). Ultimately, while it is fair to say that SA benefits written productive vocabulary to a greater extent than oral productive vocabulary and that sophistication in oral vocabulary use may be, conversely, enhanced with classroom teaching, these differences may be equally attributed to the artefact of the modality. That is, the lack of changes in oral vocabulary (except for lexical sophistication) may be also related to the lack of opportunities for planning and revision, which are abundant in writing but limited in speech.

The study has some important limitations that have to be acknowledged. One of them is related to the different tasks that we used to measure vocabulary in written and oral modes. 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. Nevertheless, it should be noted that the use of different tasks for oral and written production modes was justified on the basis of previous studies suggesting that task repetition can affect L2 performance in a number of ways (Gilabert, 2007; Robinson & Gilabert, 2007; Tavakoli & Foster, 2008). Repeating the same composition and interview twice as many times as originally planned would have been of little scientific value (if we had used the same topic for composition and interview, students would have been required to write and talk about the same thing six times). Assessing productive vocabulary with other specific vocabulary tests can also be considered in further research (e.g. Fitzpatrick & Clenton, 2010).

To conclude, while adding a 3-month ERASMUS exchange programme proves to be particularly beneficial for the development of written EFL vocabulary, traditional classroom instruction is also a necessary condition for the development of oral EFL vocabulary, especially in the sophistication of learners' lexical ability. In light of the educational implications of ERASMUS programmes and the lack of consistent empirical evidence on the development of productive oral and written vocabulary in an SA experience, further investigation in this domain is certainly necessary.

Acknowledgements: We would like to gratefully acknowledge research grants HUM2004-05442-c02-01, HUM2007-66053-C02-01/02, FFI2010-21483-C02-01/02 and SGR2005-01086/2009-140 and the participants' collaboration.



Note

1. An inter-rater reliability analysis was conducted on 10% of randomly chosen texts using the Intraclass Correlation Coefficient (ICC), with a two-way fixed effects model and the confidence interval set at .95, and proved to be acceptable, at .809, showing a strong agreement between the two raters.

References

- 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.
- Barquin, E. (2012). *Writing development in a study abroad context*. Unpublished PhD thesis, Universitat Pompeu Fabra, Barcelona, Spain.
- Beattie, J. (2014). The 'ins and outs' of a study abroad programme: The SALA exchange programme. In: Pérez-Vidal C. (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, 183–205.
- Berman, R.A., & Verhoeven, L. (2002). Developing text-production abilities across languages, genre, and modality. *Written Languages & Literacy*, 5, 1–44.
- Biber, D. (1988). *Variation across speech and writing*. New York: Cambridge University Press.
- Biber, D. (2009). Are there linguistic consequences of literacy? Comparing the potentials of language use in speech and writing. In: Olson, D.R., & N. Torrance (Eds.), *The Cambridge handbook of literacy* (pp. 75–92). Cambridge: Cambridge University Press.
- Bulté, B., & Housen, A. (2014). Conceptualizing and measuring short-term changes in L2 writing complexity. *Journal of Second Language Writing*, 26, 42–65.
- Cobb, T. (2002). *VocabProfilers* [computer program]. Available at: <https://www.lex tutor.ca/vp/> (accessed May 2019).
- Cook, V. (1999). Going beyond the native speaker in language teaching. *TESOL Quarterly*, 33, 185–210.
- Daller, H.R., Van Hout, R., & Treffers-Daller, J. (2003). Lexical richness in the spontaneous speech of bilinguals. *Applied Linguistics*, 24, 197–222.
- De Jong, N.H., Steinel, M.P., Florijn, A.F., Schoonen, R., & Hulstijn, J.H. (2012). Facets of speaking proficiency. *Studies in Second Language Acquisition*, 34, 5–34.
- DeKeyser, R. (2007). Study abroad as foreign language practice. In: DeKeyser, R. (Ed.), *Practice in a second language: Perspectives from applied linguistics and cognitive psychology* (pp. 208–226). Cambridge: Cambridge University Press.
- Dewey, D. (2008). Japanese vocabulary acquisition by learners in three contexts. *Frontiers*, 15, 127–148.

- Ellis, R. (1995). Modified oral input and the acquisition of word meanings. *Applied Linguistics*, 16, 409–435.
- Engber, C.A. (1995). The relationship of lexical proficiency to the quality of ESL compositions. *Journal of Second Language Writing*, 4, 139–155.
- Fitzpatrick, T., & Clenton, J. (2010). The challenge of validation: Assessing the performance of a test of productive vocabulary. *Language Testing*, 27, 537–554.
- Foster, P., & Skehan, P. (1996). The influence of planning on performance in task-based learning. *Studies in Second Language Acquisition*, 18, 299–324.
- 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, 34–40.
- Garside, R., & Smith, N. (1997). A hybrid grammatical tagger: CLAWS4. In: Garside, R., Leech, G., & A. McEnery (Eds.), *Corpus annotation: Linguistic information from computer text corpora* (pp. 102–121). London: Longman.
- Gilabert, R. (2007). The simultaneous manipulation of task complexity along planning time and [+/- here-and-now]: Effects on L2 oral production. In: García-Mayo, M.P. (Ed.), *Investigating tasks in formal language learning* (pp. 44–68). Clevedon: Multilingual Matters.
- 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, 55–84.
- 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.
- Koizumi, R., & In'nami, Y. (2013). Vocabulary knowledge and speaking proficiency among second language learners from novice to intermediate levels. *Journal of Language Teaching and Research*, 4, 900–913.
- Kormos, J. (2006). *Speech production and second language acquisition*. Mahwah, NJ: Lawrence Erlbaum.
- Lara, R. (2014). Complexity, accuracy and fluency development through study abroad programmes varying in duration. Unpublished PhD thesis, Universitat Pompeu Fabra, Barcelona, Spain.
- Laufer, B. (1998). The development of passive and active vocabulary in a second language: Same or different? *Applied Linguistics*, 19, 255–271.
- Laufer, B., & Nation, I.S.P. (1995). Vocabulary size and use: Lexical richness in L2 written production. *Applied Linguistics*, 16, 307–323.
- Lennon, P. (1990). Investigating fluency in EFL: A quantitative approach. *Language Learning*, 3, 387–417.
- Levelt, W. (1989). *Speaking: From intention to articulation*. Cambridge: MIT Press.
- Llanes, À., & Muñoz, C. (2009). A short stay abroad: Does it make a difference? *System*, 37(3), 353–365.
- Llanes, À., Tragant, E., & Serrano, R. (2011). The role of individual differences in a study abroad experience: The case of ERASMUS students. *International Journal of Multilingualism*, 9, 318–342.
- MacWhinney, B. (2000). *The CHILDES project: Tools for analyzing talk*. 3rd edition. Mahwah, NJ: Lawrence Erlbaum.
- Malvern, D., & Richards, B. (1997). A new measure of lexical diversity. In: Ryan, A., & 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: Beers, M., van den Bogaerde, B., Bol, G., de Jong, J., & C. Rooijmans (Eds.), *From sound to sentence: Studies on first language acquisition* (pp. 81–96). Gronigen: Centre for Language and Cognition.
- Meara, P.M., & Miralpeix, I. (2017). *Tools for researching vocabulary*. Bristol: Multilingual Matters.

- Mehnert, U. (1998). The effects of different lengths of time for planning on second language performance. *Studies in Second Language Acquisition*, 20, 83–108.
- 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.
- Mora, J.C., & Valls-Ferrer, M. (2012). Oral fluency, accuracy, and complexity in formal instruction and study abroad learning contexts. *TESOL Quarterly*, 46, 610–641.
- Nation, I.S.P. (1990). *Teaching and learning vocabulary*. New York: Newbury House.
- Ortega, L. (2014). Ways forward for a bi/multilingual turn in SLA. In: May, S. (Ed.), *The multilingual turn: Implications for SLA, TESOL and bilingual education* (pp. 32–53). New York: Routledge.
- Pérez-Vidal, C. (2014). Study abroad and formal instruction contrasted: The SALA project. In: Pérez-Vidal, C. (Ed.), *Language acquisition in study abroad and formal instruction contexts* (pp. 17–58). Amsterdam: John Benjamins.
- Pérez-Vidal, C. (2015). Practice makes best: Contrasting learning contexts, comparing learner progress. *International Journal of Multilingualism*, 12, 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, 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: Muñoz, C. (Ed.), *Intensive exposure experiences in second language learning* (pp. 213–233). Bristol: Multilingual Matters.
- Read, J. (2000). *Assessing vocabulary*. Cambridge: Cambridge University Press.
- Read, J., & Nation, I.S.P. (2006). An investigation of the lexical dimension of the IELTS speaking test. *IELTS Research Reports*, 6, 207–231.
- Robinson, P., & Gilabert, R. (2007). Task complexity, the cognition hypothesis and second language learning and performance. *IRAL*, 45(3), 161–176.
- Santos, T. (1988). Professors' reactions to the academic writing or nonnative-speaker students. *TESOL Quarterly*, 22, 69–90.
- Sasaki, M. (2007). Effects of study-abroad experiences on EFL writers: A multiple-data analysis. *Modern Language Journal*, 91, 602–620.
- Serrano, R., Tragant, E., & Llanes, À. (2012). A longitudinal analysis of the effects of one year abroad. *The Canadian Modern Language Review*, 68, 138–163.
- Singleton, D. (1999). *Exploring the second language mental lexicon*. Cambridge: Cambridge University Press.
- Skehan, P., & Foster, P. (1997). Task type and task processing conditions as influences on foreign language performance. *Language Teaching Research*, 1, 185–211.
- Sperling, M., & Freedman, S.W. (2001). Research on writing. In: Richardson, V. (Ed.), *Handbook of research on teaching* (pp. 370–389). Washington, DC: American Educational Research Association.
- Strömquist, 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, 45–68.
- 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. *International Review of Applied Linguistics*, 40, 117–150.
- 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, 365–393.
- Uchihara, T., & Saito, K. (2019). Exploring the relationship between productive vocabulary knowledge and second language oral ability. *The Language Learning Journal*, 47, 64–75.

- Ure, J. (1971). Lexical density and register differentiation. In: Perren, G.E., & J.L.M. Trim (Eds.), *Applications of linguistics: Selected papers of the second international congress of applied linguistics* (pp. 443–452). London: Cambridge University Press.
- Williams, J. (2012). The potential role(s) of writing in second language development. *Journal of Second Language Writing*, 21, 321–331.
- Wolfe-Quintero, K., Inagaki, S., & Kim, H.J. (1998). *Second language development in writing: Measures of fluency, accuracy and complexity*. Honolulu, HI: University of Hawaii at Manoa.
- Zareva, A., Schwanenflugel, P., & Nikolova, Y. (2005). Relationship between lexical competence and language proficiency: Variable sensitivity. *Studies in Second Language Acquisition*, 27, 567–595.
- Zaytseva, V. (2016). Vocabulary acquisition in study abroad and formal instruction: An investigation on oral and written lexical development. Unpublished PhD thesis, Universitat Pompeu Fabra, Barcelona, Spain.
- Zaytseva, V., Pérez-Vidal, C., & Miralpeix, I. (2018). Vocabulary acquisition during study abroad: A comprehensive review of the research and some methodological considerations. In: Sanz, C., & A. Morales-Front (Eds.), *The Routledge handbook of study abroad research and practice* (pp. 210–224). New York: Routledge.

Appendix 1 Oral interview questions.

Student A asks student B

1. Why did you choose [*this university*¹] 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?

¹ Here the name of the university, which was specified in the original interview question, has been substituted with *this university*.

7. What do you think of the requirement at *[this university]* of having to study abroad?

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. Plain 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: engIRIRAO2ISTUIIIIIStudentII

*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 2. Conventions to edit data.

Set number	Changes made	Analysis	Format	Software
1	None	Raw data	.doc	n/a
2	Pruned data Spelling errors corrected Non-words eliminated No repetitions	Sophistication (LFP) Lexical diversity (D) Lexical density	.txt	VocabProfile D_Tools CLAWS
3	Formatting and punctuation modified to CHAT conventions	Accuracy	.cha	CLAN CHILDES