ACCEPTED VERSION FOR PUBLICATION (POSTPRINT)
Zaytseva, V., Miralpeix, I., & Pérez, C. (2021). The effects of contexts on the acquisition of oral lexical ability in English as a foreign language. *The Language Learning Journal, 49*(5), 597–613. https://www.tandfonline.com/doi/full/10.1080/09571736.2019.1656766

The effects of contexts on the acquisition of oral lexical ability in English as a foreign language

Victoria Zaytseva[a], Imma Miralpeix[b] and Carmen Pérez-Vidal[a]

[a]Universitat Pompeu Fabra, Barcelona, Spain; [b]Universitat de Barcelona, Barcelona, Spain

ABSTRACT

While there is ample evidence that study abroad (SA) enhances oral fluency in a foreign language, the effects of different types of learning context on other aspects of oral skills, such as vocabulary use, have not received much attention in academic research and are less clear. The present study tries to fill this void by investigating lexical development in oral production by advanced learners of English experiencing two different contexts of acquisition: formal instruction (FI) at home, followed by a 3-month stay abroad (SA). Speech samples were elicited from a group of 30 Catalan/Spanish undergraduates before and after each learning context by means of an oral interview and were later analyzed in terms of lexical diversity, sophistication, density and accuracy. Additionally, we examined baseline data from 25 native speakers of English, elicited through the same task. Our results reveal that both contexts enhance significant development in lexical accuracy and that SA proves especially beneficial for growth in adverb density, which moves towards target-like norm and adds fluidity to learners' speech. FI, in contrast, causes a significant impact on lexical sophistication, as learners seem to acquire more specific vocabulary during classroom instruction.

KEYWORDS: study abroad, formal instruction, vocabulary acquisition, English as a foreign language, second language acquisition.

Introduction

The long trajectory of Study Abroad (SA) and language acquisition research has often pointed to SA as a highly beneficial context for the advancement of foreign language speaking skills. Many studies have provided compelling evidence of gains in oral fluency as a result of SA, whether measured as syllables or words per second or minute, (D'Amico 2010; Lennon 1990; Llanes and Muñoz 2009; Mora and Valls-Ferrer 2012; O'Brien et al. 2007; Serrano et al. 2012; Towell 2002). Post-SA gains have also been reported for the domain of oral accuracy (Llanes and Muñoz 2009; Serrano et al. 2012) and rhythm (Valls-Ferrer 2011), but have been inconclusive for syntactic complexity (Mora and Valls-Ferrer 2012; Serrano et al. 2012) and pausing (Trenchs-Parera 2009). Little research, however, has been conducted on SA and its effects on lexical development (lexical diversity, in particular), finding no significant impact in Mora and Valls-Ferrer (2012) and Serrano et al. (2012). Even fewer studies have looked at other aspects of lexical proficiency, such as lexical sophistication, or lexical density.

As far as we are aware, the present study is one of the first to examine several domains of English as a Foreign Language (EFL) oral lexical proficiency and observe changes after a 3-month SA period preceded by a 6-month formal instruction (FI) period at home. Our aim is threefold: (a) to evaluate development in oral lexical proficiency by advanced Catalan/Spanish learners of English over time,

and after each learning context under observation, FI and SA, as they are experienced one after the other; (b) to contrast learners' oral production with baseline data from native speakers, (c) to compare gains made during FI to those made during SA. By considering changes in lexical proficiency in two different learning environments, we can gain new insight into what type of changes occur in EFL lexis when learners are immersed in the target language, in contrast with classroom instruction, and how each context of acquisition can add to their vocabulary building. In the following section, we discuss the relevance of lexical competence in speaking skills and examine how vocabulary can be measured in free oral production; we then review the studies that have to some degree analysed oral lexical development in EFL during study abroad.

Lexical knowledge and speaking proficiency

Vocabulary knowledge contributes substantially to target language speaking proficiency. Several studies have supported this claim after studying this relationship from different angles and reporting that: vocabulary knowledge and correct sentence intonation are the best predictors of speaking skills (Laufer and Nation 1999); lexical knowledge is strongly associated with speaking performance through IELTS (Read 2005); vocabulary size can predict oral skills comparably with written skills (Milton 2013), and it plays a determinant role in spoken communication where misinterpretation may result from lexical errors (Ellis 1995). Furthermore, it has also been found that vocabulary knowledge not only significantly improves speaking proficiency in low and up-to-intermediate levels (Koizumi 2005; Koizumi and In'nami 2013), but it does so even at advanced levels (De Jong et al. 2012), which is the case of the learners in our study: university students with an already advanced level of English trying to improve their proficiency in two successive learning contexts.

In addition to the empirical relationship between speaking and vocabulary, there is a general assumption that vocabulary knowledge can be mastered either receptively or productively (Read 2000) and that receptive vocabulary is larger than productive vocabulary (Melka 1997). Receptive vocabulary involves understanding the meaning of words (generally in reading and listening), while productive vocabulary implies producing words in writing and speech. Traditionally, vocabulary research has focused on receptive vocabulary (Schmitt 2010), although receptive vocabulary does not affect speaking ability as much as productive vocabulary does (Koizumi and In'nami 2013). While little agreement exists as to how vocabulary knowledge can be best measured, productive vocabulary is usually assessed using *controlled* tasks (as in cloze and translation tests) or through *free* writing and speech, sometimes without any specific prompts (Laufer 1998). According to Ellis (1994), vocabulary acquisition very much takes place through oral input; yet, research on productive vocabulary has predominantly been conducted in relation to the written modality rather than to the oral (Read and Nation 2006) and there is more published work on measures of writing than on measures of speaking (Wolfe-Quintero et al. 1998). Nonetheless, vocabulary in spoken production can be assessed using the same quantitative measures as the ones used in studies analyzing writing (Read 2000). These measures pertain to the following domains: (1) lexical diversity, (2) lexical sophistication, (3) lexical density, and (4) the number of lexical errors (if any).

(1) *Lexical diversity*, also labelled lexical variation (Read 2000) or lexical range refers to the variety of words used in a spoken or written text. Speakers with rich vocabularies generally make use of synonyms to avoid repetition in their discourse and, as learner linguistic proficiency increases, their ability to produce more lexically diverse language also increases. Several formulas have been proposed to quantify lexical diversity in second language acquisition (SLA) research, most of which are the transformations of the type-token ratios (TTR) that rectify its dependence on text length.[1] Guiraud's index (GI) is an example of such transformation: it is calculated by dividing the number of word types by the square root of tokens. Although GI has proven its validity in numerous studies (Broeder et al. 1993; Vermeer 2000), other TTR variants have been put into practice in SLA research. An example of these is the D measure, which follows a curve-fitting

approach to detect how the TTR changes in a text (see Malvern et al. 2004 and Meara and Miralpeix 2017; for details on how D operates). While D has been found to be theoretically more valid than its predecessors (Malvern et al. 2004), looking at lexical diversity through these indexes might be insufficient if one aims at providing a complete picture of all the lexical resources of L2 speakers and capture vocabulary knowledge in all its richness. This takes us to consider another important vocabulary-related domain: lexical sophistication.

(2) *Lexical sophistication* can be defined as the proportion of low-frequency or 'rare' words to general, everyday vocabulary. The use of specific words or terms enables speakers to get their message across more effectively adding precision to their language. One well known index of lexical sophistication is Laufer and Nation's (1995) Lexical Frequency Profile (LFP), which classifies the words in a text according to their frequency: the most frequent 1000 words in English, the second most frequent 1000 words, the most frequent 'academic' words, and all remaining vocabulary not found in the previous frequency levels (known as 'off-list' vocabulary). The LFP is considered a valid measure of productive L2 vocabulary knowledge and an objective diagnostic test of L2 lexical proficiency; nevertheless, some authors find that it may not be particularly sensitive in discriminating between profiles with minor vocabulary size differentials (Meara 2005; Meara and Bell 2001; but see Laufer 2005, for rebuttal).

(3) *Lexical density* is usually described as the proportion of content words (nouns, verbs, adjectives, and adverbs) in the text and is used to analyze the density of texts or information packaging (Johansson 2008; Read and Nation 2006; Ure 1971). That is, since content words basically convey information, the higher the proportion of content words in a text, the denser the information it contains. While it remains unclear what developmental patterns are associated with different content words (Marsden and David's 2008) and in different contexts of learning (Freed et al. 2003), other minor variants of lexical density (i.e. the proportion of nouns, adjective or adverbs in the text) have been recently proposed (Johansson 2008). At the same time, some authors question the reliability of the lexical density measures in foreign language acquisition research, as a high index of content words in a text may be associated with learners' inability to construct a coherent discourse rather than generating an informationally dense one, which seems to occur particularly at low levels (Hyltenstam 1988).

(4) *Lexical accuracy* or the number of errors (if any) in vocabulary use is the last component that can be measured in free production following Read (2000). Given that foreign language speech generally contains vocabulary errors of various kinds, the expectation is that as learners improve in their overall language ability and advance in speaking skills, they will also improve in their lexical accuracy and, thus, make fewer errors. The relationship between lexical accuracy and overall foreign language proficiency has been corroborated in many studies (Agustín Llach 2006; Engber 1995; Zareva et al. 2005), although the pitfalls in all accuracy analysis (i.e. subjectivity when classifying and interpreting the error) have also been acknowledged (Polio 1997).

Previous research on oral vocabulary across learning contexts

Although SA research has been quite prolific in the area of speaking skills, most studies drawing on L2 oral samples have measured vocabulary (see Table 1) in a vague and fragmented way, as Zaytseva, Miralpeix and Pérez-Vidal (2018) have pointed out. Early research on SA and spoken vocabulary use comes from North America and lies at the interface between lexis and grammar. Pioneering research-ers in the field (Lafford and Ryan 1995; Ryan and Lafford 1992) have approached the study of lexis through the copula choice of *ser/estar* and the use of prepositions *por/para* in English-speaking learners acquiring Spanish in Spain. The input received in the SA environment was found to account for the variation in the order of acquisition of these features across the group. A few years later, another study concerning L2 lexico-grammatical development (Collentine 2004) also presented interesting results. Following 46 American participants learning Spanish prior to and after a semester-long

Table 1. Previous studies analyzing vocabulary in oral production across contexts.

Study	Design; L1; L2;	Duration	Instrument	Measure(s) used	Results
Ryan and Lafford (1992)	Pre-test/post-test: L1: English (*n* = 16) L2: Spanish	4 months	OPI	Stages of *ser/estar*	post-SA>pre-SA*
Lafford and Ryan (1995)	Pre-test/post-test: L1: English (*n* = 9) L2: Spanish	4 months	OPI	Stages of *por/para*	post-SA>pre-SA
Collentine (2004)	Cross-sectional; L1: English AH (*n* = 20); L1: English SA (*n* = 26); L2: Spanish	4 months	OPI	Semantic density: nouns, adj., prep., multisyllabic words, and a high TTR	SA > AH**
Freed et al. (2003)	Cross-sectional; L1: English AH (*n* = 15); L1: English SA (*n* = 15); L2: French	4 months	OPI	Lexical density: proportion of content words in the text	SA = AH
Segalowitz and Freed (2004)	Cross-sectional; L1: English AH (*n* = 18); L1: English SA (*n* = 22); L2: Spanish	4 months	OPI; various cognitive	Lexical speed: different fluency measures	SA > AH SA = AH
Foster (2009)	Cross-sectional; L1: Farsi AH (*n* = 60); L1:Different languages SA (*n* = 40); L2: English	minimum 1 year	4 cartoon prompts: *Picnic, Journey, Walkman, Football*	Lexical diversity: D	SA > AH
Llanes and Muñoz (2009)	Pre-test/post-test; L1:Spanish/Catalan (*n* = 24); L2: English	1 month	Cartoon prompt: *Picnic*	Lexical accuracy: vocabulary errors per word	post-SA>pre-SA
Llanes et al. (2012)	Pre-test/post-test: L1: Spanish (*n* = 24); L2: English	3 months	Cartoon prompt: *Picnic*	Lexical diversity: GI	post-SA>pre-SA
Pérez-Vidal and Juan-Garau (2011)	Longitudinal; L1:Spanish/Catalan (*n* = 20); L2: English	3 months	Role-play task	Lexical diversity: GI	SA = AH
Lara (2014)	Longitudinal; L1:Spanish/Catalan (*n* = 47); L2: English	3 months to 6 months	Role-play task	Lexical diversity: GI	SA = AH
Pérez-Vidal et al. (2012)	Longitudinal; L1:Spanish/Catalani(*n* = 29); L2: English	3 months	Semi-guided interview	Lexical diversity: GI	SA = AH
Serrano et al. (2012)	Pre-test/post-test; L1: Spanish (*n* = 14); L2: English	3 months to 8 months	Cartoon prompt: *Picnic*	Lexical diversity: GI	post-SA>pre-SA
Leonard and Shea (2017)	Pre-test/post-test; L1: English (*n* = 39) L2: Spanish	3 months	3 tasks Cartoon prompt: *Picnic*	Lexical diversity: D Lexical sophistication: Guiraud Adv. Index	post-SA=pre-SA post-SA>pre-SA

*Studies following a *pre-test/post-test* design have no control group and, therefore, the effects of SA cannot be compared to the effects of the AH context in these studies. We only report post-SA gains in relation to pre-SA gains.
**SA was significantly superior to FI AH for the development of the corresponding domain.

treatment in either FI at home (AH) or SA, Collentine assessed L2 oral production, as elicited via OPI (Oral Proficiency Interview), and included measures of the semantic density of information (features associated with informational richness, following Biber's[2] definition in 1988). Collentine found that the FI group outperformed the SA group in the ability to generate unique word types (nouns and adjectives), but overall the SA group's speech was more semantically dense by virtue of speaking more fluently and thus producing more words in the given time frame. Collentine's findings are consistent with those of Freed and colleagues (2003), drawing on data from an analogous sample of American university students learning French 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. One last study to consider the effects of SA and AH contexts, as based on OPI, was that of Segalowitz and Freed (2004), investigating oral lexical access in American undergraduates studying Spanish in Spain. The authors found that SA participants improved significantly from the pre-test to the post-test in terms of oral fluency, as measured by temporal and hesitation phenomena. While North-American SA research remains influential, a growing interest in SA also began to emerge in Europe (see Kinginger 2009, for a comprehensive history of SA research). European-based studies though varying in design, instrument and length of SA programmes have shed new light on the acquisition of oral skills and vocabulary. Thus, the comprehensive SALA project bore fruit to a number of publications investigating English language acquisition by Catalan/Spanish students in a translation degree in Barcelona. Pérez-Vidal and Juan-Garau (2011) analyzed oral skills using a role-play task in 20 subjects participating in a 3-month SA programme following FI AH. They did not find any significant improvement in learners' Guiraud's index of lexical diversity. A year later, drawing from the same SALA corpus, Pérez-Vidal and the SALA team (2012) looked at the advancement of oral skills but employed a different instrument: a semi-guided interview. The authors came up with the same results: they found no significant improvement for lexical diversity, as measured by GI, after the SA experience. The lack of improvement in lexical variety of L2 speech productions was also reported in another study by Lara (2014). Lara examined the impact of SA programmes varying in length (3-month stay vs. 6-month stay) on speaking development in a similar sample of Spanish/Catalan participants learning English. She was unable to detect any significant changes in this domain based on the GI measure in the case of the 3-month programme group. 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 from native speaker's (NS) scores.

As part of the Barcelona Age Factor (BAF) project, another string of studies emerged focusing on length of residence, individual differences and the age factor in the European SA context. Following Spanish learners acquiring English in the UK, the studies (Llanes et al. 2012 ; Serrano et al. 2012) coincided in finding a significant improvement in lexical diversity (based on GI) after 3 months spent abroad. The instrument used to gather oral speech data in both studies was a picture-elicited story task (i.e. cartoon prompts from 'The Picnic Story', Heaton 1966). Using the same task, Llanes and Muñoz (2009) also found that SA significantly improved the oral accuracy of young-adult learners' (from 13 to 22 years of age), and that 3–4 weeks were enough to decrease the number of lexical errors in their narratives. Llanes and Muñoz's findings parallel those of Foster (2009), although differing notoriously in informants' demographics and the measures used. Foster contrasted two comparable groups of English learners in two different contexts: a group of L1 speakers of Farsi study-ing English AH in Tehran and an SA group learning English in the UK. Foster used D to measure lear-ners' lexical diversity and included baseline data from native English speakers as a yardstick for comparison. In addition to the 'The Picnic Story' by Heaton, she used other cartoon prompts (see Table 1 for details) to elicit learner and native speech productions. The results showed that there was a significant effect for group and that not only were the SA group's productions significantly more lexically diverse than those of the AH group, as measured by D, but their vocabulary was also no less diverse than that of the NSs.

In a more recent study, Leonard and Shea (2017) set out to investigate oral development in English-speaking learners of Spanish after a 3-month stay abroad. The authors relied on two measures to analyze spoken vocabulary use: the D measure for lexical diversity and the Guiraud advanced index for lexical sophistication. The results showed that learners progressed to a significant degree in lexical sophistication, as they incorporated a greater number of low-frequency words into

their active vocabulary; but this improvement did not lead to a significant growth in their lexical diversity, as assessed by D. The researchers concluded that by using more than one measure of vocabulary knowledge, they were able to find some sizeable gains in lexis during SA, which would not have come out had they been limited to a single measure.

As we can see from the aforementioned literature, results in the area of oral lexical proficiency and SA are mixed and inconclusive. While SA can lead to considerable gains in lexis, it does not affect all the domains equally and its impact is highly dependent on the duration of the programmes, the data-elicitation instrument and the measures used (see Table 1 for summary). This lack of consistency and the growing recognition of the crucial role of vocabulary in SLA gives ample justification to delve deeper into the differential effects of learning contexts on spoken vocabulary use and further analyze oral EFL development from a lexical perspective, which is what we aim at in the study presented here.

The study

Based on data from the previously mentioned SALA project, which has investigated in depth the impact of two learning contexts, FI and SA, the study presented here aims to examine the corresponding effects on learners' lexical ability in oral production. More specifically, we seek to measure learners' lexical development in speaking, longitudinally and after FI and SA, respectively, as they are experienced one after the other. We assess EFL oral samples elicited through a structured interview in terms of quantitative measures of lexical knowledge pertaining to the domains of diversity, sophistication, density and accuracy. Furthermore, we also consider baseline data from native speakers of English, as elicited through the same task, so as to ascertain whether learners' oral lexis approximated that of native speakers in any significant way after either FI or SA. Finally, we aim to compare the amount of gains accrued in each learning context (FI vs. SA), so as to uncover which context leads to more gains. In order to accomplish these objectives, we addressed the following research questions.

RQ1. What progress is made in the acquisition of EFL oral productive vocabulary over time in two subsequent learning contexts, FI and SA?

RQ2. To what extent does learner lexical performance approximate native-like norms over the observation period?

RQ3. Which context (FI vs. SA) leads to greater vocabulary gains?

Methodology

Participants

The participants in the study were 30 learners of English (NNSs). They were Catalan/Spanish EFL-major undergraduates (mainly composed by females: 80%) studying a language specialisation degree at a Catalan university. Their ages ranged between 17 and 21 (mean age = 17.93) at the beginning of the study, with the majority (93.3%) having enrolled in the programme at the age of 17-18. Upon entrance to the degree, students had to accredit an Advanced Level of English: equivalent to a B2.1 within the Common European Framework of Reference (CEFR).

During the first two terms of their first academic year, students had to attend two English language university courses, corresponding to the FI period, with English as the medium of instruction at all times. These courses consisted of English language lectures, covering formal linguistic analysis and morpho-syntactic aspects of the language, and seminar sessions, which 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 courses. By the end of FI, students had received a total of 80 h of in-class instruction with no further FI in the third trimester.

During the second year of their degree, as part of their curriculum students were required to spend a compulsory three-month ERASMUS stay abroad in an English-speaking country. The vast

majority went to the UK (93.3%), although a small percentage of students completed their SA in the USA (6.7%). During their stays, students attended lectures mainly in Modern Languages and Humanities departments. Although no official across-the-board requirements were specified, they all took a minimum of 4 academic courses (Beattie 2014). Most students registered for modules focused on translation and their second foreign languages (German or French), alongside other elective courses, depending on their host institution's regulations. According to self-reports, students received an average of 9 h of classroom instruction per week.

In addition to NNSs, baseline data from 25 native speakers of English (NSs) were collected for comparison purposes. The NSs were undergraduate students following an exchange programme at different Spanish universities and were highly comparable to NNSs in terms of age and education. They came from the USA (15), the UK (6) and Ireland (4), 77.8% were female. They performed the same task as the NNSs, yet on only one occasion. The inclusion of NS data in our study was made in the absence of a better means of comparison with advanced-level learners and with the expectation that they will provide special insight into students' linguistic behaviour across time.

Data collection

Following a repeated measures within-subjects design, data collection was conducted longitudinally over a 15-month period. The first testing time (T1) took place upon students' entry to the degree; then at time 2 (T2), after six months (80 h) of their first year of FI in their home university, and prior to SA (with the third trimester of no formal English classes and the summer holidays in between); and finally at time 3 (T3), after the students returned from their 3-month SA exchange in an English-speaking country, that is 15 months after T1. It should be noted that although we acknowledge the possibility of task repetition effects when using the same instruments over the different treatment periods, the key advantage of a repeated measures design is that it overcomes the problem of between-subject variability (Milton and Meara 1995) and bypasses the question of whether or not groups (experimental and control) are validly comparable in the first place. Table 2 displays the three data collection times.

Task and data analysis

The instrument used to gather the participants' speech samples was a semi-guided interview, in which students were asked to act as both interviewers and interviewees. The oral interviews took place in a quiet room and were digitally recorded. The allotted time for the task was up to 5 min. In pairs, students were presented with seven questions written in the instructions and assigned 'Student A' and 'Student B' roles randomly. All the questions were different but covered the same topic: *university life*. First, student A asked student B the questions one at a time, and then they switched roles. It must be noted that even though the setting and research context may have shaped the kind of data we later obtained, an attempt was made to simulate 'natural conversation' and to make our participants as comfortable as possible. Although the researcher was there (Observer's Paradox; Labov 1972: 209), students were explicitly instructed to perform the interview as if she was not present, therefore favouring a more relaxed atmosphere and a spontaneous real-life interaction. The researcher was just an observer and only intervened to solve any technical problem or inform the test-takers of the approaching time limit. Only the answers to the interview questions were considered for analysis (see Appendix 1 for interview questions).

Participants' oral samples were analyzed quantitatively in several lexical domains, following Read (2000). Thus, we looked at lexical diversity, sophistication, density, and accuracy (See Table 3). A total

Table 2. Data collection times (adapted from SALA).

	Year 1			Year 2		
Term	1	2	3	1	2	3
Context	FI		–	SA		
Testing times	T1	T2		T3		

of 115 oral interview samples (90 by NNSs and 25 by NSs) were transcribed following CHAT transcription conventions of the CLAN software (MacWhinney 2000), resulting in the full corpus of 30,370 words. In what follows we present the measures for each lexical domain.

Lexical diversity

To operationalise the construct of lexical diversity, we included two measures: GI and D. First, type and token counts were computed using CLAN and then entered into an SPSS spreadsheet to calculate GI using the formula types/√tokens. D values were obtained using the D_Tools software (Meara and Miralpeix 2017).

Lexical sophistication

In order to calculate lexical sophistication, LFP was used, classifying learners' vocabulary into four frequency bands, as described before. This was carried out using the VocabProfile online tool (Cobb 2002), which classifies all words on the basis of four frequency bands: (a) the first 1000 most frequently used words (1k), (b) the second 1000 (2k), (c) the Academic Word List (AWL), and (d) the remainder not found on the other lists (+2k). The results of each frequency band in our study were recorded in percentages.

Lexical density

Lexical density was measured in terms of the proportion of content words to the total number of words *(CW/w)* and the proportion of function words to the total number of words *(FW/w)*. Function words included articles, prepositions, pronouns, conjunctions, and determiners. In order to have a more detailed picture of what developmental patterns were associated with content words, we also calculated separately the density of adjective *(Adj/w)*, adverbs *(Adv/w)*, nouns *(N/w)* and verbs *(V/w)* in relation to the the total number of words. All calculations were performed using CLAWS part-of-speech tagger (see Garside and Smith 1997).

Lexical accuracy

Finally, in order to assess lexical accuracy, we computed the number of vocabulary errors per word (*VE/w*), following the error classification system from previous SALA-based publications (Barquin 2012; Lara 2014; Pérez-Vidal et al. 2012), which is summarised in Zaytseva (2016: 254). These errors included wrong lexical choice, non-words, and L1 transfer errors (i.e. false cognates, confusion

Table 3. Measures used per domain.

Measures	
Lexical diversity	Guiraud's Index (GI)
	D (D)
Lexical sophistication	First 1,000 words (1k)
	Second 1,000 words (2k)
	Academic words (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)

between verb pairs 'make/do' or phrasal verbs, etc.). Errors were counted[3] using CLAN and then recorded in SPSS.

Results

RQ1: changes in EFL oral productive vocabulary over time, after FI and SA

Table 4 summarises mean scores for all lexical measures and contains information about the standard deviations in parentheses. Together with the learners' mean scores obtained at T1, T2, and T3, we include native speakers' baseline data, although the statistical comparisons with NSs are presented in the next section (*RQ2: NS-NNS comparisons*).

In order to address RQ1 exploring EFL oral vocabulary development over time, and after each learning context (FI and SA), we performed a series of one-way repeated measures ANOVAs with *Time* (T1, T2, T3) as the within-subjects factor. We further applied Bonferroni-adjusted pairwise comparisons to test whether changes occurred after either FI (between T1 and T2) or SA (between T2 and T3). Table 5 summarises the ANOVA results with the Bonferroni-adjusted pairwise comparisons for each measure. In addition to the significance value, we also report partial eta squared (η^2p) for effect sizes.

As can be seen in Table 5, the results of the main ANOVAs revealed a significant main effect of time on adverb density (Adv/W), lexical sophistication, as measured by 1k and + 2k, and accuracy (LexE/w). The largest effect size was observed for accuracy $\eta^2p = .53$. We did not find statistically significant differences in the measures of lexical diversity. Subsequent pairwise comparisons showed that while significant changes in adverb density occurred after the SA period, changes in 1k and + 2k bands reached significance after FI. That is, learners were using a significantly higher proportion of adverbs in their speech after SA and not after FI, but the sophistication of their vocabulary ability improved after the FI period and not after SA (the full list of the + 2k words used by the learners at T2 can be consulted in Appendix 2). Finally, with regard to accuracy, measures improved over time, although not in favour of any specific learning context. This indicated that learners gradu-ally reduced the number of errors in speech, finding comparable benefits from both learning conditions.

RQ2: NS-NNS comparisons

A series of independent samples *t*-tests between NS and NNS mean scores at each testing time were used to examine whether learner performance approached native-like norms at any point over the

Table 4. Descriptive statistics for oral productive vocabulary: means obtained at T1, T2 and T3 for NNSs (*n* = 30) and NSs (*n* = 25). (Standard deviations in parentheses).

		NSSs			NSs
	Measure	T1	T2	T3	
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 5. One-way Repeated Measures ANOVAs showing longitudinal changes with context-specific (FI or SA) differences for each measure (all Fs = [2, 28]).

	Measure	F	*p*	η^2p	FI	SA
Diversity	GI	.166	.848	.01	–	–
	D	1.795	.185	.11	–	–
Sophistication	1k	4.090	.028	.23	.034	.131
	2k	1.345	.277	.09	–	–
	AWL	.626	.542	.04	–	–
	+2k	3.880	.033	.22	.050	.061
Density	CW/w	1.161	.328	.08	–	–
	Adj/w	.211	.811	.02	–	–
	Adv/w	7.615	.002	.35	1.00	.002
	N/w	2.174	.133	.07	–	–
	V/w	.732	.490	.05	–	–
	FW/w	1.318	.284	.09	–	–
Accuracy	LexE/w	10.482	.000	.43	.052	.056

Note: All F values are significant at p <.001. Bold values indicate significance.

observation period and thus provide an answer to RQ2. The descriptive statistics have been previously reported in Table 4 and the inferential statistics are presented in Table 6.

As can be seen from the inferential statistics, *t*-tests results revealed little change in learner performance towards native-like standards across the contexts. Overall, significant differences were revealed between learners and NSs in all measures of lexical diversity, accuracy and sophistication at all times, except in the 2k frequency band (the second 1000 most frequent words in English), which remained unchanged throughout the study. Significant NS-NNS divergence was also found for most measures of lexical density, with Adj/w and V/w being the only ones 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. In other words, the proportion of adverbs in learner speech was the only measure that reflected target-like behaviour after the SA experience.

RQ3: differential vocabulary gains across learning contexts

Our RQ3 enquired into whether the differential gains in oral lexical proficiency were attributable to any context in particular (FI vs. SA). In order to address this question, we calculated the amount of gains made during FI, *FI gains*, by subtracting T1 scores from T2 scores (T2 – T1), and the gains obtained during SA, *SA gains*, by subtracting T3 scores from T2 scores (T3 – T2). We then ran a series of paired-samples *t*-tests to evaluate the resulting gains for each lexical measure (see Table 7 for results).

Table 6. Independent-samples *t*-tests comparing the NNS lexical mean scores to those of the NSs at T1, T2 and T3.

	T1			T2			T3		
Measure	*t*	*df*	*p*	*t*	*df*	*p*	*t*	*df*	*p*
GI	−3.709	53	<.001	−3.729	53	<.001	−3.913	53	<.001
D	−6.824	53	<.001	−7.942	53	<.001	−6.720	53	<.001
1k	5.358	53	<.001	3.462	53	.001	4.820	53	<.001
2k	−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
+2k	−4.668	53	<.001	−2.472	53	.017	−5.183	41.2	<.001
CW/w	−5.619	53	<.001	−6.279	53	<.001	−5.247	53	<.001
Adj/w	−.953	53	.345	−1.407	53	.165	−1.234	53	.223
Adv/w	−3.351	53	.001	−4.004	53	<.001	−.707	53	.483
N/w	−4.618	53	<.001	−4.101	53	<.001	−6.383	53	<.001
V/w	−1.503	53	.139	.645	53	.521	.954	53	.344
FW/w	4.463	53	<.001	5.711	53	<.001	3.645	53	.001
VE/w	8.918	31	<.001	8.307	32.2	<.001	5.961	33	<.001

Note: Bold values indicate significance.

Table 7. Paired-samples *t*-tests comparing FI vs. SA mean gains in lexical measures.

	Measure	FI gains	SA gains	*t*(29)	*p*
Diversity	GI	+.05 (.51)	−.02 (.54)	−.516	.610
	D	−1.33 (7.71)	+3.01 (8.57)	−1.719	.096
Sophistication	1k	−1.15 (2.32)	+.75 (1.96)	−2.883	.007
	2k	+.45 (1.59)	−.24 (1.50)	1.597	.121
	AWL	+.13 (.89)	+.02 (.91)	.359	.722
	+2k	+.57 (1.24)	−.54 (1.20)	2.835	.008
Density	CW/w	+.25 (4.03)	+.61 (3.03)	−.323	.749
	Adj/w	−.27 (2.21)	+.15 (2.19)	−.594	.557
	Adv/w	+.23 (2.62)	+1.37 (1.93)	−1.824	.079
	N/w	+.71 (3.04)	−1.09 (2.18)	2.417	.022
	V/w	−.43 (1.96)	+.19 (2.51)	−.852	.401
	FW/w	+.57 (3.41)	−1.20 (3.98)	1.510	.142
Accuracy	VE/w	−.42 (.90)	−.44 (.97)	.098	.923

Note: positive values (+) represent gains, negative values (−) indicate losses. Bold values indicate significance.

In view of these results, we did not register any significant differences for the amount of gains in lexical diversity and accuracy when comparing FI and SA. However, gains between the two learning contexts were significant in the domain of lexical sophistication, as measured by 1k and + 2k, with medium-large effect sizes, $r = .47$ in both cases. In other words, as learners decreased in their use of frequent vocabulary (first 1000 English words) during FI, the proportion of infrequent words increased and were attributable to the FI period. That is, learners obtained larger gains in sophisticated vocabulary during FI than during SA, becoming more target-like in English. We did find significantly larger gains in noun density resulting from the FI learning context, with the effect sizes between moderate and large. We did not register any significant differences for the size of gains in the remaining measures of lexical sophistication such as 2k, AWL, and accuracy. In light of the *t*-test results, we can consider that FI was more beneficial to learners' vocabulary acquisition than SA, as seen in the gains obtained in two measures of lexical sophistication and noun density.

Discussion

The main goal of this study was to investigate how oral lexical proficiency develops in a group of Spanish/Catalan EFL learners participating in a 3-month SA programme following FI AH. Specifically, we examined what progress is made in each of the four lexical domains – diversity, sophistication, density, and accuracy – over time, and after each particular learning context; how learners' behaviour approaches target-like norms; and whether lexical gains vary substantially from one learning context to another.

In response to RQ1 which addressed the issue of whether progress is made in EFL lexical development over time in two subsequent learning contexts experienced subsequently, FI and SA, the answer is a qualified 'yes'. Overall, our findings show that learners improved longitudinally, with the FI context yielding better results in lexical sophistication than the SA context, and SA proving more beneficial in the area of adverb density. The one domain which shows lexical development over time in the two subsequent contexts is that of lexical accuracy, as opposed to diversity, sophistication, and density.

Our failure to detect significant changes in *lexical diversity* over FI and SA indicated that learners were not using a more varied and less repetitive vocabulary and that there was no advantage for SA over and above FI in this area of lexical proficiency. The lack of progress in oral lexical diversity, as measured by GI, corroborated previous SALA findings (Lara 2014; Pérez-Vidal et al. 2012; Pérez-Vidal and Juan-Garau 2011), in which SA was not found to be particularly helpful in triggering advancement in this domain. Our findings, in contrast, do not fully coincide with other studies of similar Spanish-speaking demographic (Llanes et al., 2012 ; Serrano et al. 2012), where significant improvement was observed for oral lexical diversity based on GI, as a result of three months spent

abroad. It should be noted, however, that the instrument employed to elicit oral speech in these studies was different to ours, as it was an oral narrative task consisting of cartoon picture prompts from Heaton (1966). Regarding the lack of improvement in the D measure, our findings also differ from 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 our study and hers is again the instrument and the sample: while we compiled our data using a semi-guided informal interview, Foster elicited L2 speech productions from Farsi-speaking participants and speakers of other languages by means of again the same cartoon prompts. In this sense, a common observation in SLA research that task design and complexity can affect L2 performance in a number of ways (Robinson and Gilabert 2007; Tavakoli and Foster 2008) is very likely to account for the different results between the aforementioned studies. The lack of change in our data as far as oral lexical diversity could also be attributed to the inherent characteristics of the oral production modality. 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 so as to keep up with the conversation, it is more challenging for them to demonstrate the full range of their vocabularies when speaking.

Turning to the domain of *lexical sophistication*, improvement reached significance in the 1k and +2k measures and occurred after FI and not after SA. In other words, learners used a substantially lower percentage of the 1000 most frequent words in English (1k) and a significantly higher percentage of the least frequent (or sophisticated) vocabulary (+2k) in their post-FI speech samples. This substitution of the most frequent vocabulary by the most sophisticated vocabulary after classroom instruction AH echoes previous work by Laufer and Paribakht (1998). Although we lack comparative data from SA and oral vocabulary studies (Laufer and Paribakht examined written expression), their results are likely the product of a similar process, as gains in lexical sophistication (academic + offlist bands) by FI group in their study surpassed those of the SA group. One possible interpretation of these gains can be related to the fact that learners may sometimes 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.

Regarding the domain of oral *lexical density*, we did not find any significant changes except for adverb density (Adv/w). The mean scores for adverb density (Adv/w) showed a significant jump over time and this improvement occurred as a result of the SA experience and not after FI AH. Considering Biber's (1988) observation that adverbs are highly frequent in spoken and conversational discourse, a higher concentration of adverbs in learner speech can be interpreted as a sign of improvement given that it was also becoming more target-like by the end of the study. Bulté and colleagues (2008: 10) suggested that vocabulary growth may be related to growth in 'one or two content word classes', which in our case has been reflected in adverb word class. Although we are unaware of any other study where significant improvement in lexical density was seen through gains in adverbs, Marsden and David (2008) posited that the proportions of certain word classes change with learning and advancing in proficiency. According to the authors, as L2 proficiency increases, the number of verbs rather than nouns also increases in L2 speech, and as learners begin to use more verbs, they also incorporate more adjectives. While a single variable may not adequately depict the differential effects of FI and SA contexts on the acquisition of content words in our data, the presence of adverbs can also be associated with gains in fluency, in which lexical fillers in the form of adverbs (i.e. *well, so, really, actually*) can substitute silent pauses.

Finally, in the domain of *lexical accuracy*, significant improvement was found for all measures (LexE/w) from T1 to T3, being attributable to both FI and 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 which 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 the previous section.

Another goal of our study was to test to what extent learner production approximates target-like language in each lexical domain over the observation period. In response to RQ2, we see that learners converged with native speakers only in adverb density score (Adv/w), remaining significantly different from native-like standards across contexts for the rest of the domains (lexical diversity, sophistication, accuracy). This was unsurprising since previous research on quantitative L1-L2 comparisons in oral production has demonstrated that overall L2 speech contains less lexical variety and sophistication (Crossley and McNamara 2009). With adverb density as the only measure that showed NS-NNS convergence in our data at T3, it appears that SA was particularly beneficial for the adjustment of L2 speech to the NS pattern, at least in this word class. While research on lexical development in different contexts of acquisition involving L1-L2 comparisons is scarce, our finding may only be related to a few studies. Trenchs-Parera (2009), for instance, examining dysfluency phenomena in an analogous group of SALA participants, reported significant increases in lexical fillers after SA, making L2 speech sound more fluent and lexically richer. The lexical fillers in her study were of two types: lexicalised phrases (e.g. 'you know', 'it's like', 'and stuff') and single words (e.g. 'so', 'well', 'like'), the latter mostly represented by adverbs. The author argued that they were used as helpful crutches in oral discourse, as happens with formulaic speech in language (Nattinger and DeCarrico 1992), and that growth in lexical fillers could be related to growth in lexical repertoire, as reported in other SA studies (Milton and Meara 1995). Trenchs-Parera's results echo those of Mehnert (1998), who found that lexical density was strongly associated with fluency and that speakers who were more fluent produced a larger percentage of content words. The interpretation of increased adverb density in our data should be read along these lines as well.

Finally, in our third RQ we asked whether any context in particular, FI or SA, led to greater gains in oral productive vocabulary. Our findings have shown that superior gains were obtained mostly during the FI period, as seen in noun density (N/w) and lexical sophistication (1k and + 2k). Our results are consistent with previous work conducted via the SALA project, finding that FI leads to greater progress than SA for oral lexical complexity (Pérez-Vidal et al. 2012). Like our precursors, we may speculate that the positive impact of FI, particularly for oral lexical sophistication gains, can be attributed to the role of teaching materials. That is, English as the medium of instruction at students' home university 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. DeKeyser (2007) argued that SA does not always bring about the desired outcomes and that classroom learning may complement stay abroad in various ways (see also Pérez-Vidal (2014) for *Combination and Complementarity of Contexts Hypothesis*).

Conclusions

In the present study we investigated lexical changes in spoken production by EFL Spanish/Catalan university students undergoing two different contexts of acquisition: FI and SA. Our findings revealed that progress occurred and affected mainly lexical sophistication, which improved significantly after the FI period, and that the SA proved beneficial to the development of adverb density, which signifi-cantly increased after the stay to converge with native-like standards. These results suggest that although the SA context yields better ease and smoothness of speech, which in our case has taken the shape of adverb density, it does not necessarily enhance each and every facet of lexical proficiency. Accuracy, in contrast, has shown to yield sustained improvement encompassing both contexts.

Given that research on spoken vocabulary in different learning contexts is scarce and unsystematic, the present study represents a contribution to this field of enquiry with a longitudinal examination of advanced oral English from a lexical perspective. By analyzing lexical development in greater detail, we have attempted to go beyond 'a partial and relatively surface level representation of lexical knowledge' (Crossley et al. 2012: 2), which abounds especially in SA research. We believe that this study proves useful for later work in the area and that it allows for possible future replications.

Notes

1. TTR is considered to be highly sensitive to text length (i.e. total amount of words). As text length increases, TTR decreases complicating comparisons of samples in which text length varies considerably. Different alternatives have come into use to correct the text length factor, as mentioned in the text.
2. According to Biber, an information rich text contains nouns, attributive adjectives, prepositions, multisyllabic words, and a high TTR.
3. Error detection was performed with the help of a second coder to ensure reliability. An inter-rater reliability analysis was performed on 10% of randomly selected transcriptions using the Intraclass Correlation Coefficient (ICC), with a two-way fixed effects model and the confidence interval set at .95. Results indicated that inter-rater reliability was acceptable, at .809, showing a strong agreement between raters.

Acknowledgements

This research received financial support from the Spanish Ministry of Economy and Competitiveness [grant number FFI2013-48640-C2-1-P], the Agencia Universitària de Recerca (AGAUR) in Catalonia [grant number 2014 SGR 1563], and the three-year doctoral scholarship [grant number FI-DGR 2013]. The authors would like to acknowledge all the members of the SALA team and the students who have contributed to this multi-year project.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

This work was supported by Ministerio de Ciencia e Innovación: [Grant Number FFI2010-21483-C02-01/02,HUM2004-05442-c02-01,HUM2007-66053-C02-01/02]; Agència de Gestió d'Ajuts Universitaris i de Recerca: [Grant Number SGR2005-01086/2009-140].

References

Agustín Llach, M.P. 2006. Lexical errors in young EFL learners: How do they relate to proficiency measures? *Interlingüística* 17: 63–73.

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.

Beattie, J. 2014. The 'ins and outs' of a study abroad programme: the SALA exchange programme. In *Language Acquisition in Study Abroad and Formal Instruction Contexts*, ed. C. Pérez-Vidal, 59–84. Amsterdam: John Benjamins.

Biber, D. 1988. *Variation Across Speech and Writing*. New York: Cambridge University Press.

Broeder, P., G. Extra and R. van Hout. 1993. Richness and variety in the developing lexicon. In *Adult Language Acquisition: Cross-Linguistic Perspectives*, ed. C. Perdue, 145–163. Cambridge: Cambridge University Press.

Bulté, B., A. Housen, M. Pierrard and S. Van Daele. 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.

Cobb, T. 2002. *Web Vocabprofile* [accessed 10 January 2016 from http://www.lextutor.ca/vp/], an adaptation of Heatley, Nation & Coxhead's (2002) *Range*.

Collentine, J. 2004. The effects of learning context on morphosyntactic and lexical development. *Studies in Second Language Acquisition* 26, no. 2: 227–248.
Crossley, S.A., T. Salsbury, D.S. McNamara and S. Jarvis. 2012. Predicting lexical proficiency in language learner texts using computational indices. *Language Testing* 28, no. 4: 561–580.
Crossley, S.A. and D.S. McNamara. 2009. Computational assessment of lexical differences in L1 and L2 writing. *Journal of Second Language Writing* 18: 119–135.
D'Amico, M.L. 2010. *The effects of intensive study abroad and at home language programs on second language acquisition of Spanish*. (Unpublished doctoral dissertation, University of Florida, United States of America).
De Jong, N.H., M.P. Steinel, A.F. Florijn, R. Schoonen and J.H. Hulstijn. 2012. Facets of speaking proficiency. *Studies in Second Language Acquisition* 34: 5–34.
DeKeyser, R. 2007. Study abroad as foreign language practice. In *Practice in a Second Language: Perspectives From Applied Linguistics and Cognitive Psychology*, ed. R. DeKeyser, 208–226. Cambridge: Cambridge University Press.
Ellis, R. 1994. Factors in the incidental acquisition of vocabulary from oral input: a review essay. *Applied Language Learning* 5, no. 1: 1–32.
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.
Foster, P. 2009. Lexical diversity and native-like selection: The bonus of studying abroad. In *Vocabulary Studies in First and Second Language Acquisition*, ed. B. Richards, H. Daller, D. Malvern, P.M. Meara, J. Milton and J. Treffers-Daller, 91–106. Hampshire: Palgrave Macmillan.
Freed, B., S. So and N. Lazar. 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, no. 3: 34–40.
Garside, R. and N. Smith. 1997. A hybrid grammatical tagger: CLAWS4. In *Corpus Annotation: Linguistic Information From Computer Text Corpora*, ed. R. Garside, G. Leech and A. McEnery, 102–121. London: Longman.
Heaton, J.B. 1966. *Composition Through Pictures*. London: Longman.
Hyltenstam, K. 1988. Lexical characteristics of near-native second-language learners of Swedish. *Journal of Multilingual and Multicultural Development* 9: 67–84.
Johansson, V. 2008. Lexical diversity and lexical density in speech and writing: A developmental perspective. *Lund Working Papers in Linguistics* 53: 61–79.
Kinginger, C. 2009. *Language Learning and Study Abroad: A Critical Reading of Research*. New York: Palgrave Macmillan.
Koizumi, R. 2005. Predicting speaking ability from vocabulary knowledge. *JLTA Journal* 7, no. 1: 1–20.
Koizumi, R. and Y. In'nami. 2013. Vocabulary knowledge and speaking proficiency among second language learners from novice to intermediate levels. *Journal of Language Teaching and Research* 4: 900–913.
Labov, W. 1972. *Sociolinguistic Patterns*. Philadelphia: University of Pennsylvania.
Lafford, B. and B. Ryan. 1995. The acquisition of lexical meaning in a study abroad context: The Spanish prepositions POR and PARA. *Hispania (madrid, Spain* 78, no. 3: 528–547.
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. 1998. The development of passive and active vocabulary in second language: same or different? *Applied Linguistics* 19, no. 2: 255–271.
Laufer, B. 2005. Lexical frequency profiles: from Monte Carlo to the real world. A response to Meara (2005). *Applied Linguistics* 26, no. 4: 582–588.
Laufer, B. and I.S.P. Nation. 1995. Vocabulary size and use: lexical richness in L2 written production. *Applied Linguistics* 16, no. 3: 307–323.
Laufer, B. and I.S.P. Nation. 1999. A vocabulary-size test of controlled productive ability. *Language Testing* 16, no. 1: 33–51.
Laufer, B. and T.S. Paribakht. 1998. The relationship between passive and active vocabularies: effects of language learning context. *Language Learning* 48, no. 3: 365–391.
Lennon, P. 1990. Investigating fluency in EFL: A quantitative approach. *Language Learning* 3: 387–417.
Leonard, K.R. and C.E. Shea. 2017. L2 speaking development during study abroad: fluency, accuracy, complexity, and underlying cognitive factors. *The Modern Language Journal* 101, no. 1: 1–15.
Llanes, À. and C. Muñoz. 2009. A short stay abroad: does it make a difference? *System* 37, no. 3: 353–365.
Llanes, À., E. Tragant and R. Serrano. 2012. The role of individual differences in a study abroad experience: The case of ERASMUS students. *International Journal of Multilingualism* 9, no. 3: 318–342.
MacWhinney, B. 2000. *The CHILDES Project: Tools for Analyzing Talk*,3rd ed.). Mahwah: Lawrence Erlbaum Associates.
Malvern, D., B. Richards, N. Chipere and P. Durán. 2004. *Lexical Diversity and Language Development*. Basingstoke: Palgrave.
Marsden, E. and A. David. 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, no. 2: 181–198.
Meara, P.M. 2005. Lexical frequency profiles: A Monte Carlo analysis. *Applied Linguistics* 26, no. 1: 32–47.
Meara, P.M. and H. Bell. 2001. P_Lex: A simple and effective way of describing the lexical characteristics of short L2 texts. *Prospect* 16, no. 3: 5–19.

Meara, P.M. and I. Miralpeix. 2017. *Tools for Researching Vocabulary*. Bristol: Multilingual Matters.
Melka, F. 1997. Receptive vs. productive aspects of vocabulary. In *Vocabulary: Description, Acquisition and Pedagogy*, ed. N. Schmitt and M. McCarthy, 84–102. Cambridge: Cambridge University Press.
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. 2013. Measuring the contribution of vocabulary knowledge to proficiency in the four skills. In *L2 Vocabulary Acquisition, Knowledge and use: New Perspectives on Assessment and Corpus Analysis*, ed. C. Bardel, C. Lindqvist and B. Laufer, 57–78. Eurosla Monographs Series 2: EuroSLA.
Milton, J. and P.M. Meara. 1995. How periods abroad affect vocabulary growth in a foreign language. *ITL Review of Applied Linguistics* 107/108: 17–34.
Mora, J.C. and M. Valls-Ferrer. 2012. Oral fluency, accuracy, and complexity in formal instruction and study abroad learning contexts. *TESOL Quarterly 4*6: 610–641.
Nattinger, J.R. and J.S. DeCarrico. 1992. *Lexical Phrases and Language Teaching*. Oxford: Oxford University Press.
O'Brien, I., N. Segalowitz, B. Freed and J. Collentine. 2007. Phonological memory predicts second language oral fluency gains in adults. *Studies in Second Language Acquisition* 29: 557–581.
Pérez-Vidal, C. 2014. Study abroad and formal instruction contrasted: The SALA project. In *Language Acquisition in Study Abroad and Formal Instruction Contexts*, ed. C. Pérez-Vidal, 17–58. Amsterdam: John Benjamins.
Pérez-Vidal, C. 2015. Practice makes best: Contrasting learning contexts, comparing learner progress. *International Journal of Multilingualism* 12, no. 4: 453–470.
Pérez-Vidal, C. and M. Juan-Garau. 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, no. 2: 157–185.
Pérez-Vidal, C., M. Juan-Garau, J.C. Mora and M. Valls-Ferrer. 2012. Oral and written development in formal instruction and study abroad: differential effects of learning context. In *Intensive Exposure Experiences in Second Language Learning*, ed. C. Muñoz, 213–233. Bristol: Multilingual Matters.
Polio, C. 1997. Measures of linguistic accuracy in second language writing research. *Language Learning* 47, no. 1: 101–143.
Robinson, P. and R. Gilabert. 2007. Task complexity, the cognition hypothesis and second language learning and performance. *IRAL* 45, no. 3: 161–176.
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.
Read, J. and I.S.P. Nation. 2006. An investigation of the lexical dimension of the IELTS speaking test. *IELTS Research Reports*, *6*, 207-231. Canberra: IELTS Australia.
Robinson, P. and R. Gilabert. 2007. Task complexity, the cognition hypothesis and second language learning and performance. *IRAL* 45, no. 3: 161–176.
Ryan, J. and B. Lafford. 1992. Acquisition of lexical meaning in a study abroad environment: SER and ESTAR and the Granada experience. *Hispania (madrid, Spain* 75, no. 3: 714–722.
Schmitt, N. 2010. *Researching Vocabulary: A Vocabulary Research Manual*. Houndmills, Basingstoke: Palgrave.
Segalowitz, N. and B. Freed. 2004. Context, contact and cognition in oral fluency acquisition: learning Spanish in at home and study abroad contexts. *Studies in Second Language Acquisition* 26, no. 2: 173–199.
Serrano, R., E. Tragant and À. Llanes. 2012. A longitudinal analysis of the effects of one year abroad. *The Canadian Modern Language Review* 68, no. 2: 138–163.
Tavakoli, P. and P. Foster. 2008. Task design and second language performance: The effect of narrative type on learner output. *Language Learning* 58, no. 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, no. 3: 365–393.
Ure, J. 1971. Lexical density and register differentiation. In *Applications of Linguistics: Selected Papers of the Second International Congress of Applied Linguistics*, ed. G.E. Perren and J.L.M. Trim, 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.
Vermeer, A. 2000. Coming to grips with lexical richness. *Language Testing* 17, no. 1: 65–83.
Wolfe-Quintero, K., S. Inagaki and H.Y. Kim. 1998. *Second Language Development in Writing: Measures of Fluency, Accuracy, & Complexity*. Manoa: University of Hawai'i Press.
Zareva, A., P. Schwanenflugel and Y. Nikolova. 2005. Relationship between lexical competence and language proficiency: variable sensitivity. *Studies in Second Language Acquisition* 27, no. 4: 567–595.
Zaytseva, V. 2016. *Vocabulary Acquisition in Study Abroad and Formal Instruction: An Investigation on Oral and Written Lexical Development*. (Doctoral dissertation, Universitat Pompeu Fabra, Spain). Retrieved from http://www.tdx.cat/handle/10803/387120.
Zaytseva, V., I. Miralpeix and C. Pérez-Vidal. 2018. Vocabulary acquisition during study abroad: A comprehensive review of the research. In *The Routledge Handbook of Study Abroad Research and Practice*, ed. C. Sanz and A. Morales-Front, 210–224. New York: Routledge.

Appendices

Appendix 1. oral interview questions

Student A asks student B:

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

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

Appendix 2. the list of + 2k words used in the T2 oral interview

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)		conservative
procrastinator		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