



Lexical Knowledge in Instructed Language Learning: The Effects of Age and Exposure

INMACULADA MIRALPEIX*
Universitat de Barcelona

ABSTRACT

The aim of this study is to analyse the possible effects of Age of Onset (AO), Cognitive Maturity (Age at Testing-AT-) and Amount of Exposure (AE) on the productive vocabularies of learners of English as a Foreign Language (FL). Three groups of bilingual Catalan/Spanish students were tested towards the end of Secondary Education. The groups differed in AO (8 vs. 11 years), AT (16 vs.17) and AE (726 vs. 800 hours). They performed four different tasks: three oral (an interview, a storytelling and a roleplay) and one written (a composition). The tasks were analysed with measures extrinsic to the learners' production. Firstly, their Lexical Frequency Profiles were computed with *VocabProfile* (Nation, 1995). Secondly, *P_Lex* (Meara, 2001) was used to assess the lexical richness of the texts. Furthermore, Anglo-Saxon and Greco-Latin Cognate indices were obtained for each of the tasks. Results show that an early AO does not necessarily suppose an advantage for Early Starters (ES), as Late Starters' (LS) productive vocabularies are very similar to those of their younger peers.

KEYWORDS: age, exposure, foreign language, lexical frequency profile, school context, vocabulary.

* *Address for correspondence:* Inmaculada Miralpeix, Departamento de Filología Inglesa y Alemana, Universitat de Barcelona, Gran Via 585, 08007 Barcelona, Spain. E-mail: miralpeix@ub.edu.

I. INTRODUCTION

Age is one of the individual learner variables most thoroughly investigated in second language acquisition (SLA) research. Regarding the question of starting age (or Age of Onset -AO-), it is a popular belief that the earlier one starts learning a language, the better (for an up-to-date revision of age studies see, for instance, Nikolov & Djigunovic, 2006; Scovel, 2000). Also research with deaf and hearing individuals exposed to language in infancy has shown that they both perform comparably well in learning a new language later on in life, whereas deaf individuals with little language experience in early life perform poorly (Mayberry, Lock, & Kazmi, 2002: 38).

Several studies have shown the benefits of starting to learn a new language as early as possible. This has mostly been shown in naturalistic situations, where children normally have a slower rate of development in the target language and do not perform as well as older learners in the short term, but they quite often surpass older learners in the long run (Ekstrand, 1976; Snow, 1983; Snow & Hoefnagel-Höhle, 1978). Also Krashen, Scarcella and Long (1982:161) claim that “acquirers who begin a *natural* exposure to the L2 during childhood generally achieve higher L2 proficiency than those beginning as adults” (my italics).

Some recent studies on the age factor, however, have challenged the “consensus view” in formal contexts. That is, it has been pointed out that the belief “*the younger the better*” does not always hold when a language is learnt only at school by receiving minimal input (García Mayo & García Lecumberri, 2003; Griffin, 1993; Muñoz, 2006). In formal settings, older learners have been found to outperform younger ones in the short-run, but it is not evident that the Early Starters (ES) catch up with Late Starters (LS) in the long term (Burstall *et al.*, 1974; Singleton, 1999). In a similar vein, some neurolinguistic studies claim that proficiency and exposure to the language may be more important than age of acquisition for certain aspects (Abutalebi, Cappa, & Perani, 2001).

In FL learning, unlike in naturalistic settings, there is a limited temporal exposure to the language, as well as other limitations such as the poor quality and quantity of the input received. Exposure has been shown to be an important element in language acquisition, Muñoz (1997:21) insists that “exposure may be as crucial as the age at which initial exposure takes place, that is, the age at which pupils begin their instruction in the foreign language”. As Harley and Hart (1997) note, exposure is very much reduced when the medium of instruction in the class is not the FL but the mother tongue, as often happens in formal settings. It can also be decisive to explain the results of the studies in naturalistic and instructed settings, especially regarding the long term benefits (Singleton, 1995).

Several researchers have recommended different amounts of exposure to test long-term effects both in natural and formal settings. DeKeyser (2000) proposes a 10-year minimum period of residence in the country where the language is spoken for learners to reach ultimate attainment levels. Snow and Hoefnagel-Höhle (1978) indicate that the period of time needed by young children to catch up with older children could be of 12 months in a situation of natural immersion with unlimited exposure. Singleton (1995:3) advises that “more than 18 years would need to be spent in a formal instructional setting in order to obtain the same amount of second language input to be required for older learners’ initial advantage to begin to disappear”. As Torras and Celaya (2001) notice “the problem one comes across in formal contexts is that the advantage in ultimate attainment of younger learners that seems to exist in naturalistic contexts cannot always be tested empirically in instructional settings [...] so there is a need for studies measuring the long-term effects of an early introduction to a FL” (2001:105).

Apart from the lack of studies following up the effects of earlier and later L2 programmes over long periods of time, Singleton and Ryan (1999) also noted that in some age-based studies, children who start learning a language earlier in an instructional setting are at some point mixed with those students that started later (Oller & Nagato, 1974). Therefore, there is a blurring effect and a possible masking effect of the older learners’ initial superiority. It is also hard to find an advantage for ES in the long run, as there may be a levelling-off of their scores with those of LS. Similarly, it is also very difficult to separate AO from exposure, as learners who start early have usually received more hours of instruction. Consequently, results could be due to the amount of exposure to the language, to age or to an effect of these two combined variables. If we want to untangle the complex relationship between age, proficiency and exposure, it is very important to analyse the performance of groups of learners of different starting ages who have actually received the same amount of exposure.

Most of the studies concerned with age and FL learning have focused on phonology and syntax and very few on vocabulary. However, vocabulary cannot be neglected in age studies, it is an essential aspect in language acquisition as the following examples show. According to Mayberry and Eichen (1991), age of acquisition exerts multiple and discrete effects at each level of language structure. Specifically, they claim that age of acquisition “exerts one effect that reverberates throughout the processing of language structure” (1991:507) and this primary effect is basically lexical. Therefore, according to them, the multiple effects of age of acquisition may originate from one single source: difficulty in lexical access. Furthermore, “when we have little of the new language at our command” (as it

happens with young learners in formal settings), “it is the lexicon that is crucial [...], the words [...] will make basic communication possible” (Hatch, 1983:74). Finally, further evidence of the vocabulary neglect is found in the tests designed to check L2 ultimate attainment: they are grammaticality judgment tasks, or elicited imitation tasks to evaluate phonetics, as Singleton (1995) notes. According to Long (1990), other vocabulary tasks might be used to explore the lexical domain and ultimate attainment.

There are, however, some studies on rate of acquisition where lexis has been taken into account. Cummins (1979) predicts that older learners, with better developed cognitive skills, would acquire academic L2 skills more rapidly than younger learners, but that this would not necessarily happen in areas of L2 proficiency related to communicative skills. Nevertheless, he does not specifically state which aspects of L2 learning, apart from phonology, will be more efficiently acquired by young learners, as Harley (1986) notices. Swain (1981) and Cummins and Swain (1986) found that older learners acquired cognitively demanding aspects of L2 proficiency more rapidly than younger learners. In a school immersion context in Canada, which does not necessarily imply any contact with the language outside school, they show that older learners acquired more vocabulary in the same amount of time than did younger learners, as evaluated in a Picture Vocabulary Test.

Therefore it seems plausible that rate increases with age, because if the amount of exposure time is held constant, older learners learn faster than younger ones. McLaughlin, Osterhout, and Kim (2004) studied the rate of L2 vocabulary learning of adult learners during the first classes in a second language and they reached the conclusion that they learned different aspects of L2 words quite fast (initially about form and then about meaning). Adult L2 learning is not then “uniformly slow and laborious” as “some aspects of the language are acquired with remarkable speed” (2004:704). Also Ervin Tripp specifies that adults “tend to pay most attention to vocabulary” (1974:222). Nevertheless, some studies conducted on language learning beliefs (Torrás, Tragant, & García, 1997) point out that what parents and teachers commonly believe that children acquire in the first stages of learning a language is phonology and vocabulary.

Very few studies, however, analyse age and exposure in relation to lexical abilities in free production¹ tasks, especially in oral tasks. Some of these very few examples are Cenoz (2002), Griffin (1993) and Muñoz (2006). There is also Spadaro (1998), whose results support the existence of a sensitive period for lexical acquisition in a second language that seems to close around the age of six. These studies make use of very different measures to assess lexical knowledge: ETS French Achievement and Advanced Placement Examination Tests,

essays and stories evaluated holistically or different lexical measures and tests devised specifically for the purpose of the study.

Computational standard tools to describe and assess lexical gains are very much needed in studies of SLA studies, as researchers often want to see students' lexical improvements over a period of time in free production tasks and check how their productive lexicon expands. Meara and Bell (2001) make a distinction between intrinsic and extrinsic measures of vocabulary assessment. In the former, the assessment is carried out only in terms of the words that appear in the text (like lexical density -LD- or Type/Token Ratio -TTR-). In the latter, items are classified according to criteria external to the text itself. They claim that extrinsic measures help in making fairly strong inferences about the total lexical resources that are available to the learner. Other authors also affirm that the lexical diversity of a text is not fully self-contained and that the contribution that words make to the diversity of a text cannot be determined without considering the word's role in the language as a whole (Jarvis, 2003) or their frequencies in daily input (Vermeer, 2004).

Up to now, one of the best-known extrinsic measures of lexical richness is the *Lexical Frequency Profile* (LFP) developed by Nation (1995). This profile shows the percent of words from four different frequency levels and the calculation is done by the *VocabProfile* program. This program operates on the basis of four word lists (Nation, 1996): *Word List One* (1k) is formed by the 1,000 most frequent words in the language, *Word List Two* (2k) consists of the second 1,000 words, *Word List Three* (3k) is the University Word List -UWL- and, finally, the program classifies automatically as belonging to Level 4 (4k) all the words that do not belong to any of these lists. The program calculates the LFP on the basis of types, tokens and word families

The creation of this program entailed some advantages (Laufer & Nation, 1995): first, it provided a more detailed picture of the different types of words that learners used. Second, it made a distinction between subjects who used frequent and less frequent vocabulary and not just between those who were or were not able to vary their limited vocabularies. Moreover, LFP is claimed to be stable across administrations, to show positive correlations with other measures of lexical knowledge, and to work well with relatively short texts. Lately LFP has been used widely among researchers for different purposes: for evaluation of the vocabulary presented in language classrooms (Meara, Lightbown, & Halter, 1997) or textbooks (Milton & Hales, 1997), for analysis of writing development (Lenko-Szymanska 2002; Muncie, 2002), as a predictor of academic and pedagogic performance of TESL trainees (Morris & Cobb, 2004), to study the relationship between active and passive vocabulary knowledge

(Laufer, 1998) or to assess lexical richness of spoken productions (Ovtcharov, Cobb, & Halter, 2006). Although claimed to be able to discriminate between proficiency levels, Horst and Collins (2006) found out that, in some cases, LFP did not identify the expected increases in use of less frequent words; therefore they complemented their analysis with other indicators such as the Greco-Latin cognate index, which is also an extrinsic measure inasmuch as words are categorised in terms of their origin (that is, whether they are present in a list of cognates or not).

Some shortcomings of the LFP have already been pointed out (Coniam, 1999; Meara, 2005), two of them being that the data it produces is not easy to work with and the mathematics behind it are not enough sophisticated. Therefore, an alternative approach was proposed by Meara and Bell (2001): P_Lex. This is a computational tool that assesses the lexical richness of texts and gives information about how frequent the vocabulary that learners use is. It works on the assumption that difficult words are infrequent occurrences and thus it uses the *Poisson* Distribution² as its basis. Among the advantages of P_Lex two stand out: it works better than LFP with shorter texts (like compositions written by L2 learners) and the output it produces (λ values) is easier to work with, although the mathematical process it uses to arrive at a final score is more complex (Meara and Bell, 2001: 13-14). These authors have also found that: 1) P_Lex scores are reliably stable across administrations 2) There is an overall good correlation with other measures of productive vocabulary (such as the Levels Test) and scores for groups of different proficiency levels are reliably different 3) P_Lex can discriminate among proficiency levels even with short texts.

On account of the foregoing discussion, the aim of the present study is to analyse the possible effects of Age of Onset (AO), Cognitive Maturity (Age at Testing -AT-) and Amount of Exposure (AE) on the productive vocabularies of students of English as a Foreign Language (FL), as measured by LFP, P_Lex and Cognate Indices. It concentrates on the vocabulary used in FL production tasks and its design is an attempt to overcome the problem of mixing up ES and LS in the same class, at the same time that exposure is controlled for both ES and LS without interfering with AO, that is, an earlier AO does not entail longer exposure to the language for the younger group.

Therefore, the main purpose of this investigation is to provide an answer to these two research questions:

1. Are lexical gains favoured by an early or late AO when AE is kept constant?

2. Given the same AT but different AE, who will have richer productive vocabularies: ES with more exposure or LS with less exposure?

II. METHOD

II.1. Participants

Participants in the study are three groups (A1, A2, B1) of Catalan/Spanish bilingual students learning English as a FL in high-schools in a middle-class district in Barcelona. As it is displayed in Table 1, the groups differed with respect to AO, AT and AE.

Group A1 (N=36) started learning English when they were 8 (Grade 3 at Primary School), their average age at testing was 16.9 (when they were in Grade 11) and they had received 726 hours of formal exposure to the language, the same as group B1 (N=41). This group, however, had started instruction in English at 11 years of age (Grade 6 in Primary School) and was tested when the group average age was 17.9 (Grade 12). Therefore, group A1 had received 726 hours of exposure throughout 9 years and group B1 throughout 7. Group A2 (N=16)³ had the AO (8 years old) in common with A1, and very similar AT (17.7) with B1; they were also in the last year of high-school (Grade 12). However, students in this group had received 800 hours of curricular exposure to the English language.

Group	N	Grade	Age of Onset	Age at Testing	Amount of Exposure
A1	36	11th	8	16.9	726 h
B1	41	12th	11	17.9	726 h
A2	16	12th	8	17.7	800 h

Table 1. Participants in the study.

None of the participants in these groups had attended extracurricular English classes, nor had they had any contact with English outside school.

II.2. Procedure

Students were first asked to fill in a background questionnaire written in Catalan that elicited extensive biographical and linguistic information about the learners, with the purpose of controlling for their amount of L2 exposure, stays abroad and age/grade of first contact with English at school.

The participants then wrote a composition about themselves in a time-compressed condition (they were given 15 minutes to complete the task). Students were not allowed to consult any dictionaries or reference books and could not ask the teacher for help. Instructions were given in Catalan so as to avoid possible comprehension problems. The participants were also informed that their performance in the test would not affect their class marks.

After writing, learners were asked to perform two oral tasks in a face-to-face situation with a researcher (a semi-guided interview and a picture-elicited narrative) and an oral task with another student (a roleplay). In the semi-guided interview, they were asked about their daily routine, hobbies and families in an attempt to elicit as much information as possible and to create a situation as natural and interactive as possible. In the picture-elicited narrative - storytelling- the learners were asked to tell a story represented in six pictures showing a boy and a girl preparing food and going to a picnic with their little dog who, hidden in the basket, ended up eating their sandwiches. Finally, in the roleplay task (performed in pairs with students from the same class), one of the students was given the role of father/mother and the other that of a son/daughter who had to ask for permission from the parent to have a party with friends at home. They were both asked to negotiate the time, activities, settings, etc. The researcher was present while the task was performed but only intervened in case the learners had to be reminded of the topics to discuss or to bring the task to an end.

II.3. Data Analysis

Interviews, storytellings and roleplays were recorded and transcribed using Chiles conventions (MacWhinney, 1995). Compositions were typed and saved in *txt* files. The transcripts were adapted and coded for analysis, with special attention paid to the oral data: immediate self-repetitions, fillers and false starts were not included. Non-complete words, which can occur in the shortened form in the compositions, were turned into their complete forms (*'cos-because*), lexical inventions and L1 words were removed, and words with more than one spelling were consistently changed into one (*dining room/ dining-room*).

All the tasks were analysed using two programs: LFP (Nation, 1995) and P_Lex (Meara, 2001). The first gave us the vocabulary profile (percentages of words in each

frequency list) for each learner in each task. With the second, a lambda value for each task that each learner performed was obtained, this value shows the proportion of infrequent words in a text.

In addition to analysing each task separately, we also built four different corpora (one for each type of task) for each group. The purpose for doing so was that LFPs are claimed not to be stable with very short texts. Moreover, P_Lex, which only needs an input of 20 words to compute a lambda value, is more reliable when the tasks have more than 80 words. The total amount of tokens for each of the corpora and the average length of each task are presented in Table 2 below. As can be seen in the average length, some tasks produced more output than others, which meant that the profiles or lambdas for these short ones could be biased. Therefore, we also computed the profiles and the lambdas for each of the corpora so as to make sure that the average results that the groups obtained, i.e., the mean coming from the analysis of each task, would not be distorted by the results obtained for the short texts.

	Interview		Storytelling		Roleplay		Composition	
	Tokens Total	Average Length	Tokens Total	Average Length	Tokens Total	Average Length	Tokens Total	Average Length
A1	4848	134.67	3167	87.97	2259	68.45	3175	90.71
B1	8522	207.85	4542	110.78	2905	78.51	3379	96.54
A2	2775	173.43	1465	91.56	818	68.17	1896	118.50

Table 2. Total amount of tokens in the corpora and average length for the tasks in each of the groups.

The analysis with P_Lex offers the possibility of manually classifying all the words the program does not find in its own lists giving six options: mistake, name, number, level 0 word, easy word or hard word. The criteria adopted was to classify words following as closely as possible the suggestions in Nation's lists (Nation, 1996). That is, loan words (*jogging, pub*) and derived forms of words from the first 1,000 words were classified as "easy" (*play-player*), as were also, for instance, family names (*mother, brothers*), while words that do not appear in this list and their derivatives were taken as "hard" (*astonished, invade*). Coordinators such as *and, but* or the word *yes* were classified as "level 0".

Following Horst and Collins (2006), the Anglo-Saxon and Greco-Latin Cognate Indices⁴ of the tasks were also computed using Cobb's version 2.6 of the VP available at <http://www.lexutor.ca>. (Cobb, 2000a). These two indices, which show the percentage of

words from an Anglo-Saxon or Greco-Latin origin that the students employed, were thought to be interesting measures to explore in our context, as our learners also have (like in the Canadian study) mother tongues that are Romance languages (Spanish or Catalan). The presence of cognates in the data was evident, but it was not known up to what point learners use them as a resource in their oral and written productions, nor if it could be a good indicator of lexical growth.

The percentages given by the profiles and the lambda values obtained, as well as the indices explained above, were used to analyse the data statistically with *Statistical Package for Social Sciences v. 11.5*. Two one-way ANOVAS were conducted to ascertain whether there was a difference between the lambda values of the three groups in the four tasks and whether the cognate index for each task differed significantly among the three groups. The alpha level was set at .01 and preliminary assumption testing was also conducted with no serious violation noted.

A one-way between-groups multivariate analysis of variance⁵ was performed to investigate the roles of AO and AE in the LFPs of the learners. Preliminary assumption testing was conducted to check for normality, linearity, univariate and multivariate outliers, homogeneity of variance-covariance matrices, and multicollinearity. The roleplay was excluded from this analysis because first of all, the LFP variables for the roleplay did not follow a normal distribution in this case and secondly, a MANOVA requires having more cases in each cell than dependent variables: as there were fewer participants in A2, including the roleplay would have meant having too similar numbers of dependent variables and cases for the MANOVA to be performed.

III. RESULTS

III.1. General descriptive results

This section presents an account of the results obtained from the analysis of the corpora for each task and group. Figures 1 to 3 present the percentage of tokens and types in each of the tasks in the first 1,000 words (Figure 1), the second 1,000 (Figure 2) and the University Word List (Figure 3).

As can be observed if we compare the three figures, almost 90% of the words that learners produce corresponded to words from the first 1k band, that is, the first 1,000 words in

English. Fewer words belong to the second thousand (between 10 and 12%) or the University word list (about 5%)⁶.

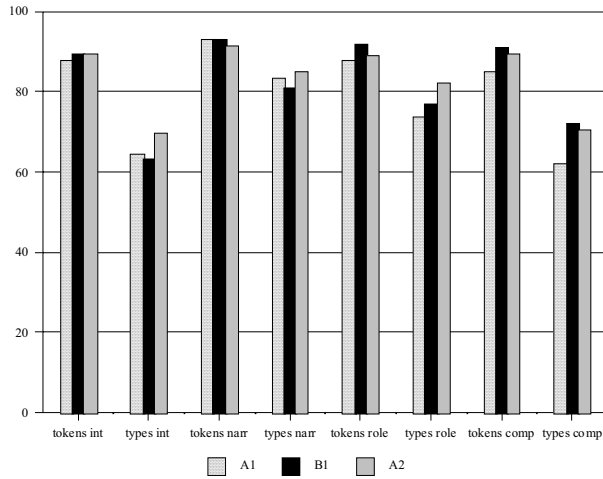


Figure 1. Percentage of tokens and types in 1k band for each group.

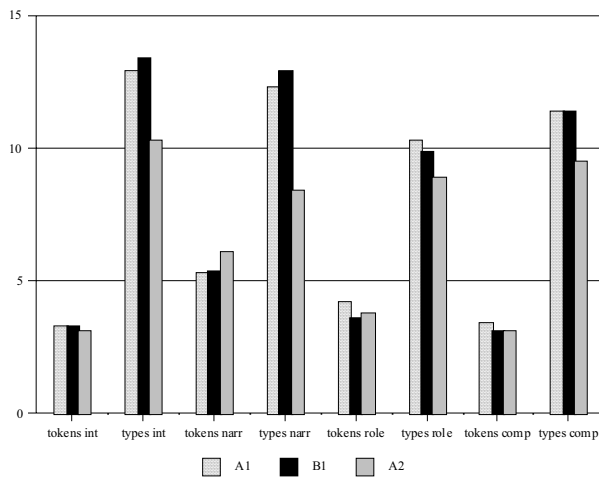


Figure 2. Percentage of tokens and types in 2k band for each group.

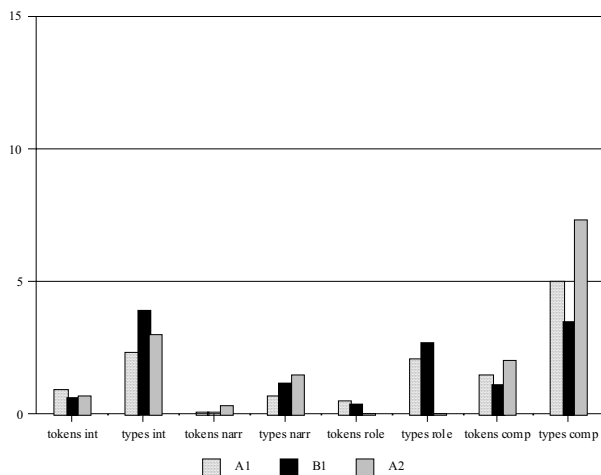


Figure 3. Percentage of tokens and types in 3k band for each group.

The vocabulary profiles offered for each corpus of tasks is shown in Table 3. As can be seen in this table, the three groups performed in a very similar way as the profiles are remarkably alike (Figure 4 offers an example). The fact that the profiles present higher percentages in the “not in the lists” column (4k) is a consequence of including proper nouns in the analysis: the program does not know how to categorise them (we find some in the interviews, roleplays and compositions but not in the storytelling, as shown in Figure 4); therefore, this rise cannot be interpreted as a high proportion of infrequent words occurring in these tasks.

Task	Group	1k	2k	3k	4k
Interview	A1	88	3.3	0.9	7.8
	B1	89.5	3.3	0.6	6.7
	A2	89.5	3.1	0.7	6.7
Storytelling	A1	93.1	5.3	0.1	1.5
	B1	92.9	5.4	0.1	1.6
	A2	91.6	6.1	0.3	2
	A1	87.7	4.2	0.5	7.7

Roleplay	B1	91.8	3.6	0.4	4.1
	A2	89.2	3.8	0	7
Composition	A1	85.2	3.4	1.5	9.9
	B1	91.2	3.1	1.1	4.5
	A2	89.3	3.1	2	5.7

Table 3. Group profiles for each task.

B1 has a few more 1k words (tokens but also types) than the other groups in the composition. The amount of vocabulary in this task from the 2k band is not very low for B1 in comparison with the other two groups, but it has less tokens and types than the others as regards the 3k band. A2 has a slightly higher number of tokens and types from the 3k list in the storytelling and the composition than the other groups.

In addition, different tasks elicited similar proportion of 1k, 2k and 3k words, although the composition seems to elicit some more types from 3k than the other tasks.

III.2. Results from statistical analyses

This section offers a summary of the results from the analyses of variance conducted. The first one-way between-groups ANOVA was conducted to explore the impact of the group the students belonged to on the lexical richness, as measured by lambda values in the four tasks (group means are shown in Table 4). No statistically significant differences were found for any of the variables. The lambda results also corroborate the fact mentioned in the previous section: the composition elicits lexically richer productions than the oral tasks, as the lambda values obtained are the highest for this task. It can also be observed in Table 4 that, although differences do not reach significance, A1 does not have the highest lambdas for any of the tasks.

	INTERVIEW		STORYTELLING		ROLEPLAY		COMPOSITION	
	AS	Cog.	AS	Cog.	AS	Cog.	AS	Cog.
A1	80.62	19.38	85.00	15.00	80.75	19.25	83.81	16.19
B1	84.36	15.64	87.98	12.02	86.80	13.20	86.84	13.16
A2	82.44	17.56	88.84	11.16	81.25	18.75	84.54	15.46
	Lambda		Lambda		Lambda		Lambda	
A1	.16206		.15697		.24467		.44636	
B1	.20171		.15188		.32600		.40229	
A2	.18000		.23813		.25091		.47625	

Table 4. Percentages of Anglo-Saxon and Cognate words as well as mean Lambdas for each group and task.

As far as the use of cognate words is concerned, a statistically significant difference in the Cognate Index was found in the oral tasks: interview [$F(2,92)=4.663, p=.012$], storytelling [$F(2,92)=4.077, p=.020$] and roleplay [$F(2,81)=5.562, p=.005$].

Post-hoc comparisons using the Tuckey HSD test indicated that the mean scores for B1 were significantly different from those of A1, which is something that did not occur in the composition, where no significant differences were found. That is, A1 used a significantly higher number of cognates than B1 in the oral tasks. However, despite reaching statistical significance, the actual difference in the mean between the groups was quite small. The effect size, calculated using eta squared, was not big: .010 for the interview and .008 for the storytelling, in the roleplay the effect size was larger (1.23).

The MANOVA analysis, which looked for any difference in the use of words coming from each of the frequency bands of the LFP (see Table 3), showed a statistically significant difference between A1 and B1 on the combined dependent variables [$F(18,142)=2.4, p=.002$; Pillai's trace=.467]⁷.

When the results for the dependent variables (i.e. amount of words in each band) were considered separately, the only difference that reached statistical significance was the number of tokens in 1k band in the composition [$F(2,381.6)=12.18, p=.000$]. An inspection of the mean scores and Post-Hoc Tukey HSD indicated that B1 produced a higher amount of tokens in band 1k ($M=90.64, SD=4.74$) than A1 ($M=84.04, SD=6.69$) in the written task, which was significant ($p=.000$).

IV. DISCUSSION

The results presented will now be interpreted and discussed taking as a starting point the initial research questions of this study.

Concerning the advantages of age of starting in terms of lexical gains, it can be seen that when AE is kept constant, significant differences are very rarely found after several years of instruction in a formal setting (7 years for B1 and 9 for A1). The lexical richness of groups A1 and B1 was strikingly similar as measured by P_Lex and LFP. The only significant difference was found in the use of cognate words, which were more frequently used by the ES group.

Horst and Collins (2006) found that more proficient learners used fewer cognates and exhibited a wider variety of frequent words. Cobb (2000b) also found out that French learners of English in Quebec relied very heavily on cognate words in everyday life and that was the reason why some vocabulary tests overestimated these learners' actual vocabularies. In the present study, A1 is the group that shows more reliance of Romance-based lexis. However, it should also be taken into account that a greater use of these words does not necessarily mean that their proficiency is much lower. As Lightbown and Libben (1984) acknowledge, the existence of cognates between languages does not imply that learners in instructional contexts will recognise or even use them, especially if there is no particular instruction on this point or if they have never encountered the word before in the target language. What is interesting to note, though, is that cognates appeared more often in the interviews and roleplays, while the composition shows a greater use of 3k words. In the oral data, time to plan the interventions is much shorter than in writing and the need to get the message across and obtain feedback is immediate. It was observed that learners made use of what Granger (1993) calls non-core cognate words in the oral task instead of using core Anglo-Saxon terms (*academy* for *language school*, *eccentric* for *odd*, *liberty* for *freedom*).

Regarding the use of more varied vocabularies in the 1k band found by Horst and Collins (2006), results from a previous study with intrinsic measures, proved that ES did not overtake LS either, nor was their score higher in an English cloze (Miralpeix, 2006). Similarly, other studies from the BAF project have consistently come up with results that indicate that LS tend to be superior in different linguistic abilities. For instance, in a study carried out by Navés, Torras and Celaya (2003), thirteen intrinsic measures of lexical complexity were used to analyze writing performance and it was found that after 726 hours LS also outperformed ES.

These results seem to be coherent with DeKeyser's view (2000) that age effects would depend on the availability of implicit learning procedures: children are better than adults at acquiring the language implicitly, while adolescents tend to benefit more from explicit instruction, which is the one provided at school. This would explain the lack of advantage for the early instructed beginners. However, it should also be noticed that vocabulary is not just an explicitly learned component; Ellis (1994), for instance, described vocabulary acquisition as an implicitly acquired skill as regards learning of forms and as an explicit learning process as regards learning of meaning.

The second research question referred to learners who shared the same AT, but one group (A2) had started earlier and had more hours of exposure than the other (B1). Results showed that there were no significant differences between these groups of learners, the very few found, as we have seen, were between A1 and B1. Therefore, after 800 hours of exposure, learners in A2 perform similarly to those of B1 and they do not surpass them. It should also be remembered that the number of participants in the A2 group is very limited (N=16) and therefore these results should be treated with due caution. However, they are in the same vein as those obtained in other school settings such as Griffin (1993), who saw that at the end of high school, American students that started French between Grades 5 and 8 outperformed those started in Grade 4 despite having received less exposure.

In spite of the fact that none of the differences are significant between A1 and A2, it can be pointed out that the length of A2 productions resembles that of B1 (they even write, for example, longer compositions) and that the amount of cognates decreases in A2. They also have the highest lambda values of the three groups in a couple of tasks. These findings might be taken as indications that A2 was "catching up with" B1, because A2 behaviour resembled more that of B1 than A1 resembled B1. However, an early AO together with some more hours of exposure are not sufficient for ES to overtake LS, who started learning English when they were cognitively more mature.

In formal settings, then, all the formal curricular exposure offered in our context does not appear to be enough to show any possible advantage as far as lexical richness in oral and written production is concerned. Despite starting later, LS probably have a faster rate of acquisition, which 74 hours of extra exposure and starting earlier do not compensate for, at least in terms of productive vocabulary knowledge. The belief that starting at a young age will give an advantage as regards vocabulary knowledge does not find support in this study: in the long term, the fact of starting at an early age does not seem to provide a benefit in a school context in productive vocabulary.

Most probably, findings from SL learning in naturalistic contexts have been generalized to FL learning without taking exposure into account and in vocabulary learning, as well as in other areas of language learning, exposure has a fundamental role, as the next two examples show. First, the chance of learning a word from a single exposure is minimal, there is a strong need for multiple contacts and consolidation. Second, even if students start to learn vocabulary at an early age, lexical knowledge seems to be more prone to attrition than other linguistic aspects, such as phonology or grammar, and this attrition is thought to occur more often at the first stages of learning a language (Schmitt, 2000). LS go through the first stages of learning the language when they are cognitively more mature and, in comparison to ES, they achieve some degree of proficiency faster –and more proficient learners tend to lose less knowledge of the new language than beginning learners-. This might also be a reason why a possible initial advantage for ES does not show in A1 and A2. Even if teaching methodologies are excellent, massive exposure to L2 input will be necessary. It could be concluded that these aspects should be taken into account in curricula planning: language production should be emphasised since the early stages of learning a language and a careful planning of what vocabulary to teach should also be encouraged.

Our results also indicate that no significant differences between A1 and A2 were found, thus suggesting no significant change in terms of production of less frequent words between the ages of 16.9 and 17.7 in a curricular framework. Nevertheless, just because the differences are not significant does not mean that there is no change. We actually found some improvement as can be seen in the descriptive figures. There could be two explanations for such a modest improvement. First of all, the emphasis on grammatical issues in a school setting at this point and a neglect of vocabulary, especially towards the end of Secondary Education when the University Entrance Examination is near. Secondly, it could also be the case that there were other factors in favour of ES that were not the focus of this study, ones that the extrinsic measures used might not have been able to identify. For instance, it is usually assumed that reception precedes production and that they probably develop in different ways (Laufer, 1998). Therefore, ES could have had greater gains in reception or word comprehension abilities or in depth of knowledge of the words (Liu & Shaw, 2001).

However, extrinsic measures can be a good way of assessing learners' development, as this sort of measures include information not available in purely quantitative measures (Daller, Van Hout & Treffers-Daller, 2003), and of knowing which vocabulary our students know and which they need. From the results obtained here, for instance, it can be deduced that if cognates are the words most readily available to learners when speaking, teachers could

introduce the equivalent Anglo-Saxon terms at different points of the syllabus so that learners' speech becomes more native-like in terms of vocabulary. It is also necessary that these extrinsic measures, such as profiles, work on reliable lists. LFP and P_Lex, have as their basis Nation's Vocabulary Lists (Nation, 1996), which were compiled following not only principles of frequency but also other criteria such as coverage or regularity, bearing in mind the learner who acquires a new language. Therefore, the purpose with which the lists are compiled are of vital importance for the reliability of the measures. Recently, different authors have proposed solutions in order to fine-tune intrinsic measures with other information not present in the text itself. For instance, in the measures Daller, Van Hout and Treffers-Daller (2003) propose (Advanced TTR and Guiraud Advanced), types are weighted according to a distinction of basic and advanced vocabulary, as non-basic vocabulary is more difficult because it is only acquired in later stages of the language acquisition process, especially in a classroom setting. Also Vermeer (2004) proposes using a corpus of age-appropriate classroom input from which to draw frequency lists of different frequency levels. As Horst and Collins (2006) have also pointed out, lists based on corpora representing child native-speaker language would probably allow for more meaningful and detailed comparisons. Extensive research in devising corpus for particular tasks (both written and oral, as the storytelling task used in this study), can be a good way of making these measures even more informative.

V. CONCLUSION

Although vocabulary has usually been neglected in age studies, it is a domain that cannot be left aside. We are nowadays far away from the notion that learning a language means learning a collection of words; however, "whichever way you look at it, lexical competence is at the heart of communicative competence" (Meara, 1996:35). That is why studies on vocabulary acquisition should also investigate the role of AO, AT or AE, with the objective of finding those measures that can better gauge vocabulary development and inform both the researcher and the teacher of learners' strengths and needs.

Results of this study indicate that an early AO in formal contexts does not systematically entail having a richer productive vocabulary in the long-run, if we understand 'long-run' as the end of Secondary Education, after at least 7 years of formal instruction. The vocabularies of the three groups analysed in our research do not present crucial differences concerning productive vocabulary in any of the tasks performed, neither in the oral nor in the written language. Therefore, in the light of these results, an early start cannot be considered an

advantage or a handicap in itself. What is worth noticing, though, is that ES, who had 800 hours of exposure to the L2, performed similarly to LS, who had less exposure (726 hours). This fact raises questions about the most appropriate AO and amount of exposure to be successful in language learning in formal settings, which is an issue that will surely stimulate further research.

Acknowledgment

This study was carried out with the assistance of the research grants BFF2001-3384 and HUM2004-05167 from the Spanish Ministry of Education. I would also like to thank Dr Carme Muñoz and the two anonymous reviewers for offering very valuable comments on the paper. I am also indebted to the editors of the volume, their efficiency and careful revision made the task much easier. Needless to say, any shortcomings are entirely my own responsibility.

NOTES

1. Nattinger (1988:62) defines production as the “retrieval of words from memory by using them in appropriate situations”.
2. The Poisson distribution is used in Statistics when the number of trials n is big but, at the same time, the probability of success p is very low, so np has a moderate size. The reason why the Poisson distribution resembles the distribution of data produced by SL learners is that both are usually skewed to the left (the probability of having segments with 0,1,2,3 or even 4 infrequent words is higher than the probability of having 10-word segments with 8,9 or 10 infrequent words). The function that describes this distribution is: $P_N = (\lambda^N \cdot e^{-\lambda}) / N!$, where λ is the average of occurrences and $e=2.71828...$ the basis of natural logarithms.
3. There is a slight variation in the N of roleplays and compositions: some of the participants did not perform the four tasks due to class constraints or they performed very poorly in the roleplay (less than 20 tokens) and the task was discarded for the analysis. Therefore, in A1 there are 33 roleplays and 35 compositions, in B1 37 roleplays and 35 compositions and 12 roleplays in A2
4. Actually, only one of these indices is necessary to know the origin of the words in a text, as the two indices always add up to 100 (eg. $AS= 80\%$, $Cog.=20\%$). This is why the statistical analyses are just carried out with *Cog. Index*.
5. There were two reasons for conducting two ANOVAS and one MANOVA instead of just one MANOVA. First, it is generally recommended not to lump all the dependent variables together unless there is a good theoretical basis for doing so, and in this case, in spite of being extrinsic vocabulary measures, their nature is quite different. Second, it is also recommended to use fairly small numbers of dependent variables (less than 10) in MANOVAS unless sample sizes are large, and in this case one of the samples has less than 30 subjects.
6. There were just minor variations between means computed from the individual profiles for each band and the mean values obtained from the corpus of each task. The same happened with lambda values.
7. Pillai's trace is reported instead of Wilks' Lambda as homogeneity of covariances matrices is assumed and there are unequal values of N in the samples. However, Pillai's Trace and Wilks' Lambda coincided in this analysis.

REFERENCES

- Abutalebi, J., Cappa, S. F., & Perani, D. (2001). The bilingual brain as revealed by functional neuroimaging. *Bilingualism: Language and Cognition*, 4:2, 179-190.
- Burstall, C., Jamieson, M., Cohen, S., & Hargreaves, M. (1974). *Primary French in the balance*. Windsor, Berks: NFER Publishing Company Ltd.
- Cenoz, J. (2002). Age differences in foreign language learning. *ITL Review of Applied Linguistics*, 135-136, 125-142.
- Cobb, T. (2000a). *The complete lexical tutor*. Available at <<http://www.lextutor.ca>>
- Cobb, T. (2000b). One size fits all? Francophone learners and English vocabulary. *The Canadian Modern Language Review*, 57:2, 295-324.
- Coniam, D. (1999). An investigation into the use of word frequency lists in computing vocabulary profiles. *Hong Kong Journal of Applied Linguistics*, 4:1, 103-123.
- Cummins, J. (1979). Cognitive/Academic language proficiency, linguistic interdependence, the optimum age question and some other matters. *Working Papers on Bilingualism*, 19, 198-203.
- Cummins, J., & Swain, M. (1986). Linguistic Interdependence: A Central Principle of Bilingual Education. In J. Cummins & M. Swain (Eds.), *Bilingualism in education: Aspects of theory, research and practice*. London: Longman, pp.80-95.
- Daller, H., Van Hout, R., & Treffers-Daller, J. (2003). Lexical Richness in the spontaneous speech of bilinguals. *Applied Linguistics*, 24:2, 197-222.
- DeKeyser, R. M. (2000). The robustness of Critical Period effects in second language acquisition. *Studies in Second Language Acquisition*, 22, 499-533.
- Ekstrand, L. (1976). Age and length of residence as variables related to the adjustment of migrant children, with special reference to second language learning. In G. Nickel (Ed.), *Proceedings of the fourth international congress of Applied Linguistics*, Vol. 3, Stuttgart: Hochschulverlag, pp. 179-197.
- Ellis, N. (1994). Vocabulary Acquisition: The Implicit Ins and Outs of Explicit Cognitive Mediation. In N. Ellis (Ed.), *Implicit and Explicit Learning of Languages*. San Diego, CA: Academic Press, pp. 211-282.
- Ervin-Tripp, S. M. (1974). Is second language learning like the first? *TESOL Quarterly*, 8:2, 111-127.
- García Mayo, M. P., & García Lecumberri, M. L. (Eds.). (2003). *Age and the acquisition of English as a foreign language*. Clevedon: Multilingual Matters.
- Granger, S. (1993). Cognates: an aid or barrier to vocabulary development? *ITL Review of Applied Linguistics*, 99/100, 43-56.

- Griffin, G. G. (1993). *The relationship between starting age and second language learning*. Unpublished Master Thesis, Faculty of the Dominican College, California.
- Harley, B. (1986). *Age in second language acquisition*. Clevedon: Multilingual Matters.
- Harley, B., & Hart, D. (1997). Language aptitude and second language proficiency in classroom learners of different starting ages. *Studies in Second Language Acquisition*, 19, 379-400.
- Hatch, E. (1983). *Psycholinguistics. A second language perspective*. Rowley: Mass.: Newbury House.
- Horst, M., & Collins, L. (2006). From *faible* to strong: How does their vocabulary grow? *The Canadian Modern Language Review*, 63:1, 83-106.
- Jarvis, S. (2003). Measuring lexical diversity through "exhaustive sampling". Paper given at Second Language Research Forum (SLRF). Tucson, US, October.
- Krashen, S. D., Scarcella, R. C., & Long, M. H. (Eds.). (1982). *Child-adult differences in second language acquisition* (2nd ed.). Rowley, Massachusetts: Newbury House.
- Laufer, B. (1998). The development of passive and active vocabulary in a second language: same or different? *Applied Linguistics*, 19:2, 255-271.
- Laufer, B., & Nation, I. S. P. (1995). Vocabulary size and use: lexical richness in L2 written production. *Applied Linguistics*, 16:3, 307-323.
- Lenko-Szymanska, A. (2002). How to Trace the Growth in Learners' Active Vocabulary? A Corpus-based Study. In B. Kettermann & M. Georg (Eds.), *Teaching and learning by doing corpus analysis*. Amsterdam: Rodopi, pp.217-230.
- Lightbown, P. M., & Libben, G. (1984). The Recognition and Use of Cognates by L2 Learners. In R. W. Anderson (Ed.), *A crosslinguistic perspective for second language research*. Rowley, Mass.: Newbury House, pp.393-417.
- Liu, E. T. K., & Shaw, P. M. (2001). Investigating learner vocabulary: A possible approach to looking at EFL/ESL learners' qualitative knowledge of the word. *International Review of Applied Linguistics*, 39:3, 171-194.
- Long, M. H. (1990). Maturational constraints on language development. *Studies in Second Language Acquisition*, 12:3, 251-285.
- MacWhinney, B. (1995). *The CHILDES Project: Tools for analysing talk*. Hillsdale, N.J: Lawrence Erlbaum.
- Mayberry, R. I., & Eichen, E. B. (1991). The long-lasting advantage of learning sign-language in childhood: Another look at the Critical Period for language acquisition. *Journal of Memory and Language*, 30, 486-512.

- Mayberry, R. I., Lock, E., & Kazmi, H. (2002). Linguistic ability and early language exposure. *Nature Neuroscience*, 417, 38.
- McLaughlin, J., Osterhout, L., & Kim, A. (2004). Neural correlates of second-language word learning: minimal instruction produces rapid change. *Nature Neuroscience*, 7:7, 702-704.
- Meara, P. M. (1996). The Dimensions of Lexical Competence. In G. Brown, K. Malmkjagr & J. Williams (Eds.), *Performance and competence in second language acquisition*. Cambridge: CUP., pp. 33-51.
- Meara, P. M. (2001). *P_Lex* (Version 1.1). Swansea.
- Meara, P. M. (2005). Lexical frequency profiles: A Monte Carlo analysis. *Applied Linguistics*, 26:1, 32-47.
- Meara, P. M., & Bell, H. (2001). P_Lex: A simple and effective way of describing the lexical characteristics of short L2 texts. *Prospect. A Journal of Australian TESOL*, 16:3, pp. 5-19.
- Meara, P. M., Lightbown, P. M., & Halter, R. H. (1997). Classrooms as lexical environments. *Language Teaching Research*, 1:1, 28-47.
- Miralpeix, I. (2006). Age and Vocabulary Acquisition in English as a Foreign Language (EFL). In C. Muñoz (Ed.), *Age and the rate of foreign language learning*. Clevedon: Multilingual Matters, pp. 89-106.
- Milton, J. & Hales, T. (1997). Applying a Lexical Profiling System to Technical English. In *Evolving models of language*. Clevedon: Multilingual Matters, pp.72-82.
- Morris, L., & Cobb, T. (2004). Vocabulary profiles as predictors of the academic performance of TESL trainees. *System*, 32:1, 75-87.
- Muncie, J. (2002). Process writing and vocabulary development: comparing Lexical Frequency Profiles across drafts. *System*, 30:2, 225-235.
- Muñoz, C. (1997). Age, exposure and foreign language learning. *APAC Monographs*, 2, pp. 16-24.
- Muñoz, C. (Ed.). (2006). *Age and the rate of foreign language learning*. Clevedon: Multilingual Matters.
- Nation, I. S. P. (1995). *VocabProfile*. Wellington.
- Nation, I. S. P. (1996). *Vocabulary lists*. Wellington: Victoria University of Wellington. English Language Institute. Occasional Publication n.17.
- Navés, T., Torras, M. R., & Celaya, M. L. (2003). Long-term effects of an earlier start. An analysis of EFL written production. In S. H. Foster-Cohen & S. Pekarek Doehler (Eds.), *Eurosla Yearbook* (Vol. 3). Amsterdam: John Benjamins, pp.103-129.

- Nikolov, M., & Djigunovic, M. (2006). Recent research on age, second language acquisition, and early foreign language learning. *Annual Review of Applied Linguistics*, 26, 234-260.
- Ovtcharov, V., Cobb, T., & Halter, R. H. (2006). La richesse lexicale des productions orales: mesure fiable du niveau de compétence langagière. *The Canadian Modern Language Review*, 63:1, 107-125.
- Oller, J., & Nagato, N. (1974). The long term effects of FLES: an experiment. *The Modern Language Journal*, 58, 15-19.
- Schmitt, N. (2000). *Vocabulary in language teaching*. Cambridge: CUP.
- Scovel, T. (2000). A critical review of the critical period research. *Annual Review of Applied Linguistics*, 20, 213-223.
- Singleton, D. (1995). Introduction: A critical look at the Critical Period Hypothesis in Second Language Acquisition. In D. Singleton & Z. Lengyel (Eds.), *The age factor in second language acquisition*. Clevedon: Multilingual Matters, pp.1-29
- Singleton, D. (1999). *Exploring the second language mental lexicon*. Cambridge: CUP.
- Singleton, D., & Ryan, L. (1999). *Language acquisition: The age factor* (2nd ed. 2004). Clevedon: Multilingual Matters.
- Snow, C. E. (1983). Age Differences in Second Language Acquisition: Research Findings and Folk Psychology. In K. M. Bailey, M. H. Long & S. Peck (Eds.), *Second language acquisition*. London: Newbury House, pp.141-150.
- Snow, C. E., & Hoefnagel-Höhle, M. (1978). Age Differences in Second Language Acquisition. In E. M. Hatch (Ed.), *Second language acquisition. A book of readings*. Rowley: Newbury House, pp. 333-344.
- Spadaro, K. (1998). *Maturational constraints on lexical acquisition in a second language*. Unpublished Doctoral Dissertation. University of Western Australia, Australia.
- Swain, M. (1981). Time and timing in bilingual education. *Language Learning*, 31, 1-15.
- Torras, M. R., & Celaya, M. L. (2001). Age-related differences in the development of written production. An empirical study of EFL school learners. *International Journal of English Studies*, 1:2, 103-126.
- Torras, M. R., Tragant, E., & García, M. L. (1997). Croyances populaires sur l'apprentissage précoce d'une langue étrangère. *AILE*, 10, 127-158.
- Vermeer, A. (2004). The Relation between Lexical Richness and Vocabulary Size in Dutch L1 and L2 Children. In P. Bogaards & B. Laufer (Eds.), *Vocabulary in a second language*. Amsterdam: John Benjamins, pp.173-189.