THE INFLUENCE OF AGE ON VOCABULARY ACQUISITION IN ENGLISH AS A FOREIGN LANGUAGE

Tesi doctoral presentada per

Immaculada Miralpeix Pujol

com a requeriment per a l'obtenció del títol de

Doctora en Filologia Anglesa

Programa de Doctorat: *Lingüística Aplicada*
(Bienni 2000-2002)
Departament de Filologia Anglesa i Alemanya

Directors: **Dra. Carme Muñoz Lahoz i Dr. Paul M. Meara**

Universitat de Barcelona

2008

CHAPTER 6

LEXICAL FREQUENCY PROFILES AND P_LEX

6.1. Introduction

The tasks performed by the learners in English are analysed in this chapter with measures extrinsic to the learners' production. Firstly, their LFPs are computed with *VocabProfile* (Nation, 1995a). Secondly, *P_Lex* (Meara, 2001) is used as an exploratory measure to describe the lexical richness of the texts. Furthermore, the proportion of cognates the learners use is assessed by means of the Anglo-Saxon and Greco-Latin Cognate indices, which are obtained for each of the tasks.

As we have seen in chapter 5, there are quite clear indications that the LS group is not outperformed at the end of secondary education by the ES group. Even though the latter has received more exposure after ten years of instruction in the FL, the differences are not significant in the majority of cases. However, vocabulary knowledge has been tested by means of a cloze and the measures that have been applied to the oral and written language produced by the learners are all intrinsic: the evaluation has been carried out in terms of the words in the learners' production and not according to criteria extrinsic to the texts. If there is a clear moral to be drawn from studies in assessing vocabulary (such as Daller, Milton & Treffers-Daller, 2007; Read, 2000) is that multiple

assessment is necessary, as long as the multiple measures adopted are derived from a coherent theoretical model.

Long-term achievement, in our case, could also be described using extrinsic measures, as learners' produce more after receiving a large amount of exposure to the language. This fact allows for the analysis of the data with two tools that evaluate the texts following criteria external to the texts, which is not possible in earlier stages due to the very few words the learners are able to produce.

6.2. Participants

Participants in the study are students belonging to groups A3, B3 and A4, these groups differed with respect to AO, AT and AE. Their characteristics have been fully described in chapter 4 and are also displayed in Table 6.18. In short, they can be summmarised as follows: Group A3 (N=36) started learning English when they were 8 (Grade 3 at primary school), their average age at testing was 16.3 (when they were in Grade 11) and they had received 726 hours of formal exposure to the language, the same as group B3 (N=41). The latter 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 A3 had received 726 hours of exposure throughout 9 years and group B3 throughout 7. Group A4 (N=16) had the AO (8 years old) in common with A3, and very similar AT to that of B3 (17.7); 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 (AO)	Age at Testing (AT)	Amount of Exposure (AE)
A3	36	11	8	16.3	726 h
B3	41	12	11	17.9	726 h
A4	16	12	8	17.7	800 h

Table 6.18. Participants in the study.

6.3. Data analysis

All the tasks were analysed using two programs[40]: *VocabProfile* (Nation, 1995a) 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, although Laufer (1995) has also asserted that there are no definitive conclusions about LPF sensitivity to text length. 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 (Meara & Bell, 2001). The total amount of tokens for each of the corpora and the average length of each task are presented in Table 6.19 below. As can be seen in the average length, some tasks elicited more output than others, which meant that the profiles or lambdas for the short ones

[40] There is a slight variation in the N of roleplays and compositions: some of the participants performed very poorly (less than 20 tokens) and the task was discarded for the analysis. Therefore, in A3 there are 33 roleplays and 35 compositions, in B3 37 roleplays and 35 compositions and 12 roleplays in A4.

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 attributed to each group, 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
A3	4,848	134.67	3,167	87.97	2,259	68.45	3,175	90.71
B3	8,522	207.85	4,542	110.78	2,905	78.51	3,379	96.54
A4	2,775	173.43	1,465	91.56	818	68.17	1,896	118.50

Table 6.19. 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 that 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', as well as numbers and proper nouns.

Following Horst and Collins (2006), the Anglo-Saxon and Greco-Latin Cognate Indices[41] of the tasks were also computed using Cobb's version 2.6 of the vocabulary

[41] 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.

profile available at http://www.lextutor.ca. (Cobb, 2000a). These two indices, which show the percentage of words from an Anglo-Saxon or Greco-Latin origin that the students employ, 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. 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 variance[42] 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

[42] 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 (Field, 2000) and in this case, in spite of being extrinsic vocabulary measures, their nature is quite different. Second, it is usually 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.

participants in A4, including the roleplay would have meant having too similar numbers of dependent variables and cases for the MANOVA to be performed.

6.4. Results

6.4.1. General descriptive results

This section presents an account of the results obtained from the analysis of the corpora for each task and group. The vocabulary profiles offered for each corpus of tasks are shown in Table 6.20. As can be seen in this table, the three groups performed in a very similar way as the profiles are remarkably alike (Figure 6.14 illustrates this similarity in the storytelling task and the shapes were very similar in the other tasks as well).

As can be appreciated in Table 6.20, there were just minor variations between means computed from individual profiles for each band and the mean values obtained from the corpus of each task[43]. While the first and second rows for each group in the table display respectively the means and the standard deviations when the profiles were computed individually, the third row shows the means for the corpora of the whole group. The difference between the first and the third row for each group and task are very subtle.

[43] The same happened with lambda values.

Figures 6.15 and 6.16 show the percentage of tokens and types in each of the tasks in the first 1,000 words (Figure 6.15) and the second 1,000 (Figure 6.16). As can be observed if we compare these 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%). Table 6.20 and Figure 6.14 show that very few words belong to the University Word List (about 5%) or to band 4 (not present in the lists).

In addition, it is worth mentioning that 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.

		1st 1,000 words (1k)				2nd 2,000 words (2k)				University Word List UWL (3k)				Not in the lists (4k)			
		I	**S**	**R**	**C**	**I**	**S**	**R**	**C**	**I**	**S**	**R**	**C**	**I**	**S**	**R**	**C**
A3	**tokens (individual)**	95.46	93.7	95.11	93.42	3.56	5.68	4.38	3.54	0.77	0.52	0.3	1.54	0.21	0.1	0.21	1.5
	sd	2.17	4.17	3.3	3.7	2.05	4.36	3.32	2.47	0.75	0.23	0.68	1.68	0.43	0.58	0.69	2.68
	tokens (corpora)	95.5	94.5	95.1	93.6	3.3	5.3	4.2	3.4	0.9	0.1	0.5	1.5	0.3	0.1	0.2	1.5
B3	**tokens (individual)**	95.51	94.24	95.4	94.96	3.5	5.12	3.78	3.18	0.54	0.4	0.41	1.12	0.45	0.24	0.41	0.74
	sd	1.86	2.49	2.84	3.25	1.81	2.52	2.72	2.36	0.51	0.2	0.76	1.39	0.7	0.63	1.05	1.20
	tokens (corpora)	95.7	94.3	95.7	95.1	3.3	5.4	3.6	3.1	0.6	0.1	0.4	1.1	0.4	0.2	0.3	0.7
A4	**tokens (individual)**	95.93	93.31	95.83	93.61	3.22	6.06	3.37	3.39	0.61	0.29	0	1.91	0.24	0.34	0.8	1.09
	sd	0.63	3.36	3.30	3.05	1.16	3.34	2.38	1.65	0.47	0.68	0	1.86	0.41	0.58	1.22	0.86
	tokens (corpora)	95.9	93.3	95.5	94.3	3.1	6.1	3.8	3.1	0.7	0.3	0	2	0.3	0.3	0.7	0.6

Table 6.20. Mean percentages and standard deviations of words at different levels for each task and group.
I= Interview, S= Storytelling, R= Roleplay, C= Composition.

B3 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 B3 in comparison with the other two groups, and it has also fewer tokens and types than the others as regards the 3k band. A4 has a slightly higher number of tokens and types from the 3k list in the storytelling and the composition than the other groups.

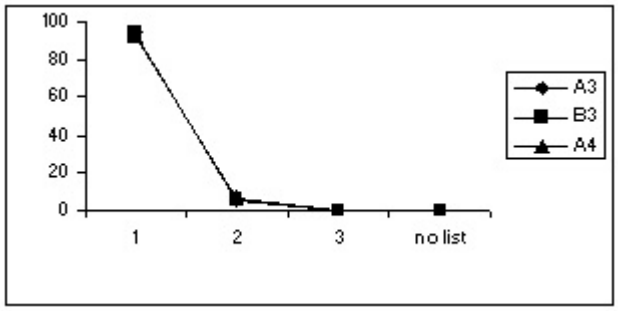

Figure 6.14. Group profiles (LFP) for the storytelling.

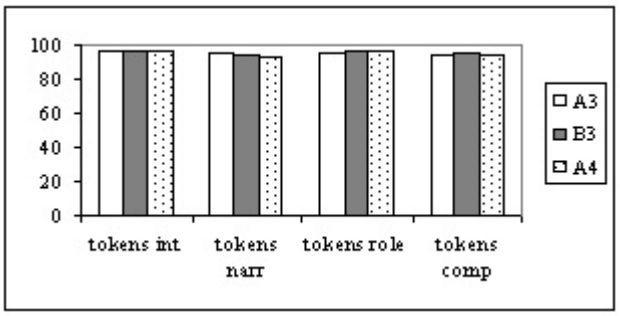

Figure 6.15. Percentage of tokens and types in 1k band for each group and task.

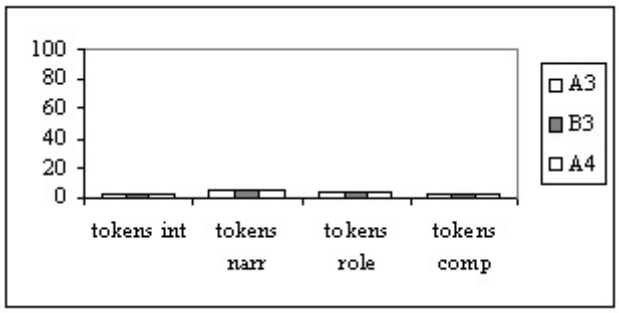

Figure 6.16. Percentage of tokens and types in 2k band for each group and task.

6.4.2. Results from statistical analyses

This section offers a summary of the results from the analyses of variance performed. 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 6.21). No statistically significant differences were found for any of the variables. Although the lambda values obtained should be taken with caution as *P_Lex* is still an exploratory tool, the results seem to 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 particular task, but differences are not significant. It can also be observed in Table 6.21 that A3 does not have the highest lambdas for any of the tasks

	INTERVIEW		**STORYTELLING**		**ROLEPLAY**		**COMPOSITION**	
	AS	**Cog.**	**AS**	**Cog.**	**AS**	**Cog.**	**AS**	**Cog.**
A3	80.62 (7.18)	19.38 (7.19)	85.00 (6.18)	15.00 (6.18)	80.75 (10.71)	19.25 (10.78)	83.81 (7.51)	16.19 (5.90)
B3	84.36 (2.92)	15.64 (2.92)	87.98 (4.53)	12.02 (4.53)	86.80 (4.79)	13.20 (4.80)	86.84 (4.51)	13.16 (4.42)
A4	82.44 (5.42)	17.56 (5.42)	88.84 (5.38)	11.16 (5.54)	81.25 (6.89)	18.75 (6.89)	84.54 (4.78)	15.46 (4.68)
	Lambda		**Lambda**		**Lambda**		**Lambda**	
A3	.16206 (.12)		.15697 (.17)		.24467 (.23)		.44636 (.32)	
B3	.20171 (.12)		.15188 (.12)		.32600 (.24)		.40229 (.27)	
A4	.18000 (.13)		.23813 (.22)		.25091 (.29)		.47625 (.38)	

Table 6.21. Percentages of Anglo-Saxon and Cognate words as well as mean lambdas for each group and task. Standard deviations are found in brackets.

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 [$F(2,92)=4.663$, $p=.012$] and roleplay [$F(2,92)=4.077$, $p=.020$].

Post-hoc comparisons using the Tuckey HSD test indicated that the mean scores for B3 were significantly different from those of A3, which is something that did not occur in the composition, where no significant differences were found. That is, A3 used a significantly higher number of cognates than B3 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).

In the MANOVA analysis, which looked for any difference in the use of words coming from each of the frequency bands of the LFP, no statistically significant difference was found: [F(18,142)=1,077, p=.381; Pillai's Trace[44]=.240; partial eta squared=.120].

Results show that an early AO does not necessarily suppose an advantage for the ES group when AE is kept constant, as LS' productive vocabularies are very similar to those of their younger peers. Given the same AT, ES with some extra exposure to the language do not significantly outperform the LS group either.

[44] Pillai's Trace is reported here instead of Wilks' lambda as it is thought to be more robust when the sample size is small or when there are unequal N values (Pallant, 2001), as it is the case in this analysis.

6.5. Discussion

Concerning the role of AO 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 B3 and 8-9 for A3). The lexical richness of groups A3 and B3 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, A3 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 oral tasks, while the composition shows a greater (though not significant) 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 our previous study with intrinsic measures (chapter 5), proved that ES did not overtake LS either, nor was their score higher in an English cloze. 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.

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 (1994b), 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.

Regarding B3 and B4, the two groups that shared the same AT but differed in AO and AE (A4 had started earlier and had more hours of exposure than B3), results showed no significant differences between them, the very few found, as we have seen, were between A3 and B3. Therefore, after 800 hours of exposure, learners in A4 perform similarly to those of B3 and they do not surpass them. It should also be remembered that the number of participants in the A4 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 A3 and A4, it can be pointed out that the length of A4 productions resembles that of B3 (they even write, for example, longer compositions) and that the amount of cognates decreases in A4. They also have the highest lambda values of the three groups in a couple of tasks (storytelling and composition). These findings might be taken as indications that A4 was 'catching up with' B3, because A4 behaviour resembled more that of B3 than A3 resembled B3 (we will come back to these results in chapter 8). 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 for any possible advantage to be shown in favour of ES, 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 (when comparing A4 and B3) 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 sub-study either: in the long term, 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 generalised to FL learning without taking exposure into account (Muñoz, 2006b) 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, without considerable exposure attrition is likely to occur 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 at Times 3 and 4. Even if teaching methodologies were excellent, massive exposure to L2 input would be necessary (see Jiménez Catalán, 1997-98 for a review of the roles of input and exposure on vocabulary learning). It could then be concluded that these aspects should be taken into account in curricula planning. Although all learners seem to go through a 'silent period', 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 there are no significant differences between A3 and A4, thus suggesting no significant change in terms of production of less frequent words between the ages of 16.3 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, as already introduced in the previous chapter, 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), but this would be pure speculation.

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 NS 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.

6.6. Conclusion

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 results, thus, are in the same line as those obtained in chapter 5 using intrinsic measures: the vocabularies of the three groups analysed in this 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). Could it be then that A4 had actually bigger vocabulary sizes than the other groups? Which group has larger productive vocabularies? How could we give an estimate of how many words these learners know productively? This is what we try to solve in the next chapter.