# 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 4

## RESEARCH QUESTIONS AND METHOD

### 4.1. Introduction

This chapter presents the methodology and procedures used in the present dissertation to answer the research questions proposed. Apart from the participants and the instruments for the data collection, it also accounts for a series of methodological decisions that were taken in the light of a pilot study with the data (section 4.3.4.1). The chapter closes with some preliminary considerations on the analyses that will be performed.

### 4.2. Research questions

It has been shown in chapter 2 that there is a lack of studies relating starting age and vocabulary, especially productive vocabulary in a FL, these studies are even more scarce if we talk about long-term effects. Besides, very few long-term studies succeed in controlling the amount of exposure the learners receive, which is a key factor to reach any conclusion as regards the role of age and the role of exposure in the learning of a FL. Accordingly, then, one of the aims of this project is to find out if an earlier introduction to English in a Catalan/Spanish bilingual context has any positive effects as regards

productive vocabulary (both oral and written) in the long run.

The present study analyses the productive vocabulary of two groups of students who started English at different ages in primary education (age 8 vs. age 11). It aims to see whether an early start confers any advantage on the productive vocabulary knowledge of the students towards the end of secondary education. That is, both groups will be tested at different time periods with the objective of analysing their productive vocabulary development and especially assessing it after 726 hours of exposure to the language (ES and LS) and after 800 hours (ES). This means that learners will have received English lessons for 7 school years (LS) and for 9-10 (ES).

Productive vocabulary, though, is not easy to quantify, especially if it is oral. As it has been seen in chapter 3, different measures have been used along the years to evaluate it and a distinction has been drawn between intrinsic and extrinsic measures. Both will be applied and discussed in this study in chapters 5 and 6. However, the issue of deriving accurate vocabulary estimates from text samples that the learners produce is still at its infancy and far from being solved. There are some theories on which to base these vocabulary estimations that seem to be more valid than others, although how to operationalise the whole process is nowadays a debatable topic. Therefore the last part of this study presents an attempt to estimate the size of the productive vocabulary knowledge of these learners from the vocabulary they use in oral and written tasks.

In summary, the present study wants to answer these research questions:

1) After having received the same amount of exposure, who will have richer productive vocabularies: a group of Early Starters (ES) who started learning English at 8 (Grade 3) or a group of Late Starters (LS) who started at 11 (Grade 6)?

2) Will ES with more exposure than LS have better productive vocabularies at the end of secondary education (Grade 12)?

3) How do different intrinsic and extrinsic measures describe the productive vocabulary of these learners?

4) How can their oral and written productive vocabulary size for some tasks be appropriately estimated?

## 4.3. Method

In an attempt to provide answers to the research questions presented, the methodology implemented in the current dissertation is the following:

4.3.1. Participants

The sample of participants for this study belongs to the BAF (*Barcelona Age Factor*) Project, which studies the effects of age on the learning of English in an instructional context. The subjects were Catalan/Spanish bilinguals who were learning

English as an L3 in state-funded schools in Barcelona. The schools were located in three central districts of the city and the participants belonged to different social classes, from low-middle class to professionals.

Students were divided into two groups as shown in Table 4.2. The first group started learning English when they were 8, hence they will be called Early Starters (ES). These participants followed the new LOGSE curriculum: they studied *Primaria* (primary school) until Grade 6 and then continued into *Secundaria* (secondary school) until Grade 10, when they were 15. After that, *Bachillerato* (also in secondary school) comprises grades 11 and 12 and it is not compulsory for all Spanish students, but they are a requirement for those who want to attend university.

Students in the other group started learning English when they were 11, hence they will be called Late Starters (LS). They were following the former curriculum in the Spanish Educational System, which was applicable from 1970 to 2000. Students started learning English at Grade 6 in the EGB (Educación General Básica), which comprised 8 courses (primary school) and continued all along the BUP (*Bachillerato Unificado Polivalente*), from 14 to 16 years old, and COU (*Curso de Orientación Universitaria*) when they were 17-18 years old and which was a pre-university course.

As the new curriculum took effect from 1992, when the former one was still being implemented, their coexistence made it possible to collect data from students with different Ages of Onset (AO) and curricular exposure, the Ages at Testing (AT) were also different and the design allowed for a variety of comparisons involving these three

variables. Because a high number of schools took part in the study (21 in total)[24],

variables related to teachers and books (like teachers' methodologies or textbooks used)

were considered to be balanced out.

Data was collected from ES and LS on different occasions (see data collection

years in Table 4.2). At Time 1, both groups had received 200 hours of formal exposure:

ES were in Grade 5 (Group A1) and LS in Grade 7 (Group B1). At Time 2 they had

received 416 hours: ES were in Grade 7 (Group A2) and LS in Grade 9 (Group B2). At

Time 3, after 726 hours of instruction, ES were in Grade 11 (Group A3) and LS in Grade

12 (Group B3). Finally, data was also collected at Time 4 from a group of ES in Grade

12 (Group A4), after 800 hours of exposure. Data at Time 4 could not be collected for

the LS group because at Grade 12 students finished their secondary education: some of

them dropped their studies and some continued with university education, therefore

neither their school itineraries nor a possible ongoing formal exposure to English could

be controlled any more.

The present dissertation focusses mainly on Times 3 and 4: A3, B3 and A4. The

first group -A3- (ES, N=57) started learning English at the age of 8 and was tested when

their average age was 16.3. In this study, the subjects that belong to A3 come from two

different grades: Grade 10 (4Eso) and Grade 11 (1Batxillerat). They all started English

at school when they were 8 (Grade 3) and received the same amount of exposure (726h),

but the first group received it along 8 school courses and the second along 9. The

---

[24] The total number of schools in the BAF project for these groups is 29, but there were 21 schools involved in the sample for the present dissertation. For the present study data was collected in a total of 6 primary schools and 15 secondary schools. In some cases data from different grades was collected in the same school and data from one specific grade was usually collected in more than one school. Each of the following grades involved data collections in the number of schools indicated within brackets: Grade 5 (4), Grade 7 (11), Grade 9 (8), Grade 10 (1), Grade 11 (3) and Grade 12 (6).

average age of the first group at testing is 15.6 and for the second group 16.6. In spite

of the intensity difference, no significant proficiency differences were found between the

groups, which assured that they could be merged.

Students in A4 (ES, N=16) also started English at 8 and were all at Grade 12 at

the time of the data collection. Their mean age was of 17.7 years and therefore they had

been learning English at school for about ten years. Group B3 (LS, N=41) began

learning English at the age of 11 and was tested at the age of 17.9. The amount of

exposure was the same as A3: 726 hours (see Table 4.2), although distributed

differently: 7 school years for LS and 8-9 years for ES. For the purpose of this study, we

chose the students with curricular exposure only; those with stays abroad, attending

language schools and repeaters were not included in the sample.

|  | GROUP A ES (AO=8) | GROUP B LS (AO=11) |
|---|---|---|
| Time 1 200h | **A1** AT=10.9 N=31 D.C: 1995-96 | **B1** AT=13 N=29 D.C: 1995-96 |
| Time 2 416h | **A2** AT=12.9 N=47 D.C: 1997-98 | **B2** AT=15 N=22 D.C: 1997-98 |
| Time 3 726h | **A3** AT=16.3 N=57 D.C: 2001-02 | **B3** AT=17.9 N=41 D.C: 1996-97 |
| Time 4 800h | **A4** AT=17.7 N=16 D.C: 2003-04 |  |

Table 4.2. General outline of the subjects in the BAF project.
AO=Age of Onset, AT=Age at Testing, DC=Data Collection.

Although those three groups in bold (Times 3 and 4) are the basis of this dissertation, the seven groups are introduced here because their lexical performance has also been analysed in earlier grades. The reasons for focussing on Times 3 and 4 were the following: first of all, this study deals with long-term effects and therefore, the end of secondary education was thought to be a good indication of 'long term' in the present educational context. Secondly, FL learners do not tend to produce much at low levels, especially at Times 1 and 2. That is why we thought that the analysis at low levels should be accompanied by robust statistical analyses with longer productions at later stages (Times 3 and 4), in this way a more complete picture of the development, in addition to more reliable results, would be obtained.

In spite of focussing on Times 3 and 4, there are two clear incentives to have a descriptive analysis of what is taking place at lower levels. The first is that very few studies attempt at characterising the lexical competence at these levels, especially as regards vocabulary. Therefore, it is important to trace the development from the very first stages of learning the language in a curricular framework, where not much data is available. The second is that a short and mid-term description of these stages may probably help to elucidate what happens in the long run. This becomes especially relevant if we also take into account the fact that a few subjects are longitudinal, that is, some ES and some LS were tested at all data collections (see Table 4.3)[25].

Despite the interesting data that longitudinal subjects may offer, this type of participants were particularly difficult to find. For a number of reasons, some of them

---

[25] In the case of group B3, all the data obtained was cross-sectional because when B1 or B2 had reached the 726 hours of exposure, the former curriculum was extinguishing.

were lost when they changed from primary to secondary school or they changed school during their secondary education and their itineraries could not be followed -some participants, for instance, could be tested at Times 1 and 3 or at Times 3 and 4, but not in all four data collections-. What happened with other participants was that, as they grew older, they started to have some extracurricular exposure, either because they went to language schools, spent some time abroad or repeated one or more courses.

Therefore, out of the 243 participants that appear in Table 4.2 (151 ES and 92 LS), a sample of them (34 ES and 10 LS) are longitudinal subjects, as specified in Table 4.3.

| Group | Longitudinal | N |
|---|---|---|
| ES | T1+T2+T3+T4 | 4 |
| | T1+T2+T3 | 1 |
| | T2+T3+T4 | 9 |
| | T1+T2 | 4 |
| | T2+T3 | 1 |
| | T3+T4 | 15 |
| LS | T1+T2 | 10 |

Table 4.3. Types and number of longitudinal subjects in each group.

Not all the analyses in the chapters that follow deal simultaneously with all the groups presented due to several reasons. First of all, not all the data analysed is of the same nature, mainly because of the different proficiency levels of the students. Therefore, there are some measures that can -and must- be applied to all the data because, as Witalisz points out, "quantitative measures can be very useful for

comparison within the same study as long as they are consistently applied to all subjects" (2007:115), but there are some others that can only be applied to some groups. Secondly, some of the analyses will focus on the long-term effects while some others will analyse short and mid-term effects of starting earlier or later. Thirdly, because of the low number of longitudinal subjects, some of the data cannot be analysed statistically, but we consider the information that these subjects could provide extremely valuable and useful for the reasons just mentioned above. Therefore, as the number of participants and the use of particular tools can vary in the chapters where data analyses are reported, other relevant information and specific details (cross-sectional or longitudinal subjects, number of groups etc.) will be given for each particular study.

It should finally be mentioned that oral data from six NSs was collected for very specific purposes -to compare, using a computer tool, their productive vocabulary estimates with those of SL learners in the storytelling task-. The detailed account will be given in chapter 7.

4.3.2. Instruments

Students performed three oral and one written task and filled in a background questionnaire and an English cloze. The aim was to elicit both free and controlled productive vocabulary. Table 4.4 below presents the tasks performed, which can also be found in Appendix A.

### 4.3.2.1. Background questionnaire

A questionnaire written in Catalan (the school language) was used to elicit information about when students started receiving English classes and for how long instruction had taken place, to make sure that the hours of exposure were kept equal for both groups. It also elicited information about extracurricular exposure, if any, and extensive biographical and linguistic information about the learners (see Appendix A.1).

### 4.3.2.2. Oral tests

*Free productive vocabulary*

As regards free productive vocabulary, learners were asked to perform three tasks: first of all they were asked about their daily lives, hobbies, family etc. by a researcher in a semi-guided interview. The aim of this task was to elicit the maximum number of answers and to create a situation as natural and interactive as possible. The actual questions appear in Appendix A.3.1.

Secondly, in a picture-elicited narrative (storytelling) the learners were asked to look at six pictures with no text showing a boy and a girl preparing food and going to a picnic with their little dog, who hides in the basket they carry. They discover at the end that the dog has eaten their sandwiches. Afterwards, the learners were asked to tell the story to the researcher, who was instructed to intervene as little as possible (the task is shown in Appendix A.3.2). The story, from Heaton (1966), has been found to

discriminate adequately between different proficiency levels in a variety of skills. For instance, Álvarez (2006) identified a series of stages in the telling of the story that learners went through as their proficiency developed, each set of stages was characteristic of a particular level. Also Muñoz (2000, 2006b) found differences on textual cohesion depending on the level of the student that told the story.

Finally, a roleplay task was performed in pairs with students of the same class. In this task, 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 to 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 s/he had to remind the learners of the topics to discuss or bring the task to an end (see Appendix A.3.3). Samples of the oral data are presented in Appendix B (B.1 for the interview, B.2 for the storytelling and B.3 for the roleplay).

### 4.3.2.3. Written tests

The subjects were asked to perform two tasks: to write a composition and to fill in an English cloze.

*Free productive vocabulary*

Students wrote a composition in class about themselves, the maximum time allowed for them to write was 15 minutes. The title they were given was: 'Me, my past

and my future' and they were not allowed to use any dictionary or textbook or ask for any kind of help. A sample is shown in Appendix B.4.

*Controlled productive vocabulary*

A cloze test about the Little Red Riding Hood (that appears in Appendix A.2) was used to assess the learners' written controlled productive vocabulary. This cloze contained 30 gaps the students had to fill in (25 for the students in Grade 10[26]). Reliability of the cloze test was assessed by computing Cronbach's Alpha for a group of ES (N=51) and a group of LS (N=51). The results, .8860 and .8674 respectively show that the test has internal consistency. It was also shown to have a good discrimination and difficulty rate in subjects with different proficiency levels in the BAF Project.

| Background questionnaire | | |
|---|---|---|
| Oral tests | Free | Semi-guided interview<br>Storytelling<br>Roleplay |
| Written tests | Free | Composition |
| | Controlled | Cloze |

Table 4.4. Tasks performed by the students.

---

[26] Having a version with 25 items instead of 30 was an operational decision (see Appendix A.2 for the items that were discarded). A reliability analysis performed showed that there were 5 items whose difficulty index for advanced groups was close to 1 and their discrimination index was extremely low (<.19). Taken together, these indications confirmed that the items were redundant in the assessment of the students' performance (del Rincón et. al, 1995).

There were some of the subjects who could not perform all the tasks due to several reasons[27], out of a total of 243 cases, the following tasks were analysed: 243 interviews, 243 storytelling tasks, 236 roleplays, 234 compositions (therefore 956 tasks with free productive vocabulary) and 232 clozes (controlled productive).

Several general comments should be made about the tests used. First of all, they are called 'tasks' in the general sense defined by Crookes (1986:1, cited in Norris et al. 1998): 'A piece of work or an activity, usually with a specified objective, undertaken as part of an educational course, at work, or used to elicit data for research'. As McNamara (2000) points out, this definition does not only include the 'real-language tasks' but also those in which the communication of meaning is submitted to pedagogic or, as in this case, research concerns. For example, the type of tasks used here are a means for eliciting learners' use of vocabulary.

In addition, the oral interview and the roleplay can be considered communicative tasks (Muñoz, 2006b), as they take into account the social roles that testees might assume in real-world settings and fulfill the following requirements: meaning is primary, task completion has some priority, learners are not given other people's meanings to regurgitate, there is a communicative problem to solve and there is some sort of relationship to comparable real-world activities (Skehan, 1998a). However, in this case, the assessment of the task is not done in terms of the outcome (i.e. if the aim of the communicative situation is achieved or not), but in the very specific terms of the vocabulary employed. It should also be noted that, although the three oral tasks were

---

[27] Mainly because they were not present in class the day of the data collection and could not be tested later or because there were problems with the oral recordings.

designed to elicit learners' production, they also involve comprehension skills to understand what is being asked (in the interview and the roleplay) and when help from the interviewer is needed to continue with the task (as it may be the case with the storytelling).

All the tests can also be considered integrative, as they integrate knowledge of systematic features of the language (pronunciation, grammar, vocabulary...) with the understanding of the context to achieve communication, i.e. it is assumed that 'knowing' a language should be determined by the students' ability to operate in a specific sociolinguistic situation. Discrete-point tests, on the contrary, focus on knowledge of the linguistic system for its own sake. The cloze would deserve a special comment, though. Most researchers agree that the cloze requires the student to integrate different types of knowledge and that it could replace a test of productive skills, being apt for assessing language skills at different proficiency levels (Oller, 1972; Fotos, 1991). Some claim, however, that it can only be considered an integrative measure if the reader has to read beyond the immediate sentence (Alderson, 1983), which is the case of the test used here, otherwise the cloze would  measure the same as a discrete point test of grammar or vocabulary does.

Lastly, in terms of method, the cloze would be considered a paper-and-pencil language test, which are typically used to assess either separate components of language knowledge (grammar or vocabulary for instance) or receptive understanding (for example of a text). However, the composition and the oral tasks would be regarded as performance tests, as "the language skills are assessed in an act of communication, [in which] a more or less extensive sample of speech or writing is elicited from the test

92

taker [and where] the samples are elicited in the context of simulations of real-world tasks in realistic contexts" (McNamara, 2000:5).

### 4.3.3. Procedure

Written tests were administered by an external researcher to intact classes in each of the schools. The tests that have been introduced were part of a battery of tests in the BAF Project, which also included  a listening comprehension, a multiple-choice grammar test and a dictation, as well as four more tests in the L1 of the subjects.

Only some of the subjects who performed the written tasks were chosen to do the oral ones. The selection criteria gave priority to those who, according to the questionnaire,  had started English at the specific ages and  that had just received curricular exposure. The battery of oral tasks was performed outside class in a quiet room for recording. Only the researcher and the subject selected were present, except in the roleplay task, for which another student from the same class joined them. The oral tasks lasted for 20 minutes approximately: 10 minutes for the interview, 5 for the storytelling task and 5 for the roleplay.

### 4.3.4. Analyses

This section introduces the kind of the analyses performed with the data, not without reflecting, first of all, on the methodological decisions that were taken before carrying out the analyses.

*4.3.4.1. Methodological decisions*

*4.3.4.1.1. Oral data*

As Broeder et al. (1988) (hereafter Broeder et al.) acknowledge, inventorying the lexicon of language users is not a straightforward issue. Further complication is added when we are dealing with oral data elicited in communication with other speakers. Although this type of tasks tend to resemble more those present in the real-word, its analysis is more complex due to the nature of communication itself:

> "Developments in discourse analysis and pragmatics have revealed the essential interactivity of all communication. This is specially clear in relation to the assessment of speaking. The problem is that of isolating the contribution of a single individual (the candidate) in a joint communicative activity." (McNamara 2000:83)

In addition to analysing oral communicative tasks, it should be also be born in mind that participants performing the tasks in our study were L2 learners, therefore they might sometimes ask for help. According to findings by Fathman and Precup (1983) and Grañena (2006), most of the communicative strategies learners make use of are related to lexical items that are unknown in the target language. For instance, the appeals for assistance tend to be lexical, which frequently means that some of the words produced by learners were actually provided by the interviewer so that the task could be performed, especially at low levels of proficiency.

Therefore, as regards oral data, it became necessary to make a selection of the items that were going to be included in the analyses. The tasks were transcribed using CHILDES (MacWhinney, 1995) and revised at least two times by a different researcher

and all the transcripts were checked and adapted in order to ensure that the learners'
amount of types and tokens produced was neither overestimated nor underestimated.

The basic unit for the present analysis was the single word because, as Folse
(2004) points out, the group of single words includes the bulk of the vocabulary of any
language, and it includes not only more items but also more frequently used items
(which are also the ones used for learners that do not have a high level of proficiency).
Very few phrasal verbs were found in the data (e.g. *come out, go away or take out*) and
it was also decided to separate compound verb tenses of the type *has eaten* or *will go.*
The main reason to do so was the high amount of mistakes in compound and
prepositional verbs, like *go to inside the basket, the dog is eat, a place to eating*.

As it was of crucial importance to decide on which words would be taken into
account for the analysis, a pilot study was conducted with one oral task: the storytelling.
The narrations of students belonging to six groups (A1, A2, A3, B1, B2 and B3)[28] of
different ages and proficiency were analysed (N=30 per group). The purpose of this
analysis was to identify those phenomena in the oral language that may be problematic
to assign words to one speaker or the other (e.g. self and other repetitions) or represent
an incomplete knowledge of the word (e.g. mistakes, inventions) and that would make
frequency counts difficult. The study was conducted with different groups so that the
inclusion or exclusion of different aspects identified would not benefit or go against any
group, that is, some phenomena could be more prototypical at a given level and its
inclusion or exclusion might have overestimated or underestimated the productive
vocabularies at a particular stage.

---

[28] This study was carried out before data from A4 was collected during spring 2004.

The storytelling was chosen for several reasons: although it might be classified as a monologic task (the subject is supposed to tell the story), the researcher had to intervene at some points as the best learners were considered to have merely an intermediate level. Furthermore, this task did not follow the question-answer pattern of the other two, which makes decisions easier in the interview and roleplay and in this sense some issues in the storytelling were considered more controversial and questionable than in the other oral tasks. Finally, it was also chosen because the phenomena observed in this task were also present in the interview and roleplay and the purpose was to have a band, as representative as possible, to be applied in the analysis of the three tasks (some examples in the other tasks are also given below together with some from the storytelling).

Consequently, this analysis gave rise to a data-driven classification with different categories whose inclusion or not in the analysis that followed should be decided on. Percentages of occurrence were calculated for each category and they appear in Table 4.5. The first row shows the groups and the second row the total amount of spontaneous English tokens produced by each group. For each category and group the following figures are presented: the raw figure (or the occurrences of that particular category), the percentage of occurrence of the category in relation to the total amount of spontaneous English words in each group and finally the percentage of each category in relation to all the occurrences of that category in all groups.

| | A1 | B1 | A2 | B2 | A3 | B3 | TOTAL |
|---|---|---|---|---|---|---|---|
| Spontaneous English words | 973 | 1,662 | 1,998 | 2,829 | 2,978 | 4,208 | 14,648 |
| **INCLUDED** | | | | | | | |
| Non-immediate repetitions | 3<br>0.30%<br>20% | 2<br>0,12%<br>13.33% | 5<br>0.25%<br>33.33% | 3<br>0.11%<br>20% | 2<br>0.07%<br>13.33% | 0 | 15<br>0.10%<br>100% |
| Wrong pronunciation | 34<br>3.49%<br>38.20% | 10<br>0.60%<br>11.24% | 7<br>0.35%<br>7.86% | 13<br>0.46%<br>14.61% | 9<br>0.30%<br>10.11% | 16<br>0.38%<br>17.98% | 89<br>0.61%<br>100% |
| Words the interviewer has used | 9<br>0.92%<br>13.23% | 15<br>0.90%<br>22.06% | 14<br>0.70%<br>20.59% | 6<br>0.21%<br>8.82% | 13<br>0.44%<br>19.12% | 11<br>0.26%<br>16.18% | 68<br>0.46%<br>100% |
| Words with clue | 0 | 0 | 2<br>0.10%<br>50% | 0 | 1<br>0.03%<br>25% | 1<br>0.02%<br>25% | 4<br>0.03%<br>100% |
| Mistakes | 4<br>0.41%<br>3.70% | 24<br>1.44%<br>22.22% | 17<br>0.85%<br>15.74% | 15<br>0.53%<br>13.89% | 27<br>0.91%<br>25% | 21<br>0.50%<br>19.44% | 108<br>0.74%<br>100% |
| **NOT INCLUDED** | | | | | | | |
| L1 words | 398<br>40.90%<br>23.76% | 468<br>28.16%<br>27.94% | 341<br>17.07%<br>20.36% | 174<br>6.15%<br>10.39% | 229<br>7.69%<br>13.67% | 65<br>1.55%<br>3.88% | 1,675<br>11.43%<br>100% |
| Lexical inventions | 6<br>0.62%<br>20.69% | 9<br>0.54%<br>31.03% | 8<br>0.40%<br>27.59% | 0 | 1<br>0.03%<br>3.45% | 5<br>0.12%<br>17.24% | 29<br>0,20%<br>100% |
| Immediate repetition | 55<br>5.65%<br>18.21% | 83<br>4.99%<br>27.48% | 65<br>3.25%<br>21.52% | 41<br>1.45%<br>13.58% | 31<br>1.04%<br>10.26% | 27<br>0.64%<br>8.94% | 302<br>2.06%<br>100% |

Table 4.5. Traits found in the storytelling task at different proficiency levels.

This quantification was used to back up the decisions that had to be taken, as it shows how representative of the data each category is; it also helped to see if its inclusion or exclusion might influence the results. Besides, as a way of having a theoretical background against which the inclusion or exclusion of each category could be assessed, the aforementioned article by Broeder et al. and Richards and Malvern (2000) (hereafter Richards and Malvern) were chosen. Most of the studies analysing

oral data do not make explicit which words they take into account or, when they do it, they give very general indications, which undermines the results and makes replication impossible. Richards and Malvern give some indications of the way their interviews were analysed as regards vocabulary. Broeder et al. also provide information about the organization of their data file, the records obtained for each word and the process to build the file, as they claim that "only researchers very familiar with the data to be analysed can recognise and handle the problematic items in the utterances of the language learners studied" (1988:30), only in this way can problems of faulty segmentation be solved and conclusions on size, variety and development of the learners' lexical lexicon be drawn more consistently[29]. This was the objective of the following analysis, which tries to identify the phenomena that should or not be contemplated in the main analyses of the present dissertation. Therefore, in addition to all the spontaneous words in English, the analysis in the present dissertation included also the following five categories:

1) Non-immediate repetitions

In non-immediate repetitions the subject does not know the word in English and the interviewer gives him/her the word, which the subject repeats but not in the next turn (there is at least one turn in between). Example 1 shows an instance of an immediate

---

[29] Lately, researchers as Tidball and Treffers-Daller (2007:140-142) also give clear indications of the standardisation of the data conducted in their study. They analyse learners' production in French, which is different from English as regards inflections and morphology. These differences between languages sometimes affect the counting of types/lemmas and segmentation.

repetition (the first 'basket'), and an example of a non-immediate repetition (the last

'basket'). The first is not included in the analysis, as explained below, while the second

is.

> Example 1 (0925 A3)
> (At the beginning of the storytelling)
> *CAR:  look the # no sé the box no sé cómo llamarlo esto . [+ s]
> *INT:   the **basket** .
> *CAR:    the **basket** .
>       [...]
> (At the end of the storytelling)
> *INT:   Why are they surprised?
> *CAR:  Doncs because the dog is inside the **basket** . [+ c]

The repetition can consist of exactly the same word, as in Example 1 or it may

vary, as it is the case with 'takes out' and 'take' in Example 2. This would be an instance

of what Heath (1980), quoted in Peters (1985), calls 'repetition with variation'. When

referring to language segmentation, Peters points out that as a child's linguistic skills

develop s/he is heard not just imitating a phrase but also manipulating it in different

ways and this phenomena also takes place in child SL acquisition. In the data examined,

non-immediate repetitions represented just a 0.10% of all the tokens produced in

English.

> Example 2 (0091 B2)
> *HUG:  frightened er a watch the dog .
> *INT:   aha .
> *HUG:  er sacan . [+ s]
> *INT:   takes out .
> *HUG:  **takes out** er the coffee I watch no the sandwich .
> *INT:   aha .
> *HUG:  er the boy er [/] er the [/] the <to watch> [/] to watch of xxx **take** [/]
>       take that take xxx the coffee and the dog I go to home [//] house .
> *INT:   very good .

2) Wrong pronunciation

Sometimes the student pronounces the word in an incorrect way that hinders communication, as it is the case with 'map' and 'picture' in Examples 3 and 4. Those instances that could be understood by the context in which they appeared were included in the analysis.

> Example 3 (0316 A2)
> *CLA:          er mother and girl and boy looking mm a
>         **map** [= pronounced as **meip**], no ?
>
> Example 4 (0326 A2)
> *RAM: er this **picture** [= pronounced **/piktur/**] have got a: family +...

However, words that could not be understood even with the help of the context (or just understood due to the subject's translation or the visual prompt) were not taken into account. They do not just reflect a pronunciation problem but also a very biased and fragmentary knowledge of the word.  This is the case in Example 5 ('hangred' for 'hungry') or 'esmolting' for 'smelling'. Broeder et al. indicated that the degree to which word-forms deviated from the pronunciation norm should be normalised and their indication was that phonemes that deviate from the norm while having no morpheme value[30] should be counted. The same convention is followed in the present study as it is considered that if phonemes have no morpheme value the meaning will normally be properly conveyed. Examples 5 and 6 would not follow this principle:

---

[30] By 'having no morpheme value' we understand the units that do not have meaning. E.g. -ed attached to a bare infinitive would have morpheme value (past tense), while the 'ei' in /meip/ instead of /map/ would have no morpheme value, /ei/ does not have meaning.

Example 5 (0492 B1)
*X:     er the gran boy mm bueno a ver [**hangred**], están **hambrientos**, ¿no?
*INT:  aha

Example 6 (0745 B3)
*S:     er the dog arrive[i] <at the at the top of> [/] at the top of the table and [/] and
         is **esmolting** ?
*INT:  **smelling** .

Although this category represents a 3.49% of all the tokens included in the

lowest grade (A3), it just constitutes a percentage around 0.5% in more proficient

groups.

3) Words the interviewer has used

Sometimes the student does not ask for the word but we observe that the

interviewer has used it before, for instance in an attempt to encourage the subject to

continue or in a general question aimed at providing the subject with things to say

whenever s/he is stuck at a particular point. It did not happen often, as the category

represents a 0.46% in the pool of English data.

For instance, 'mother' (in Example 7) was included in the analysis even if the

interviewer used the word before, as well as 'feel' in Example 8, as it is quite possible

that the subject was acquainted with these words before.

Example 7 (0349 A2)
*INT:        er what is **mother** doing here ?
              [...]
*VAN:       mm and **mother** +...
*INT:        yes .
*VAN:       +, xxx goodbye +...

Example 8 (0840 B3)
*E:      ah!, the the dog had had eaten all the food
*INT:   and what do the children say or **feel**?
*E:      the children **feel** very surprised

In Example 7 the participant may be acquainted with the word because it is a very common word and in Example 8 because s/he answers the question correctly and knowing the meaning of the verb here is the key to answer properly. Therefore, if the subject does not show that s/he does not know the word, cases where the interviewer had used the word before were not excluded from the analysis.

4) Words the interviewer gives a clue about

In very few cases, where the learner found difficulties to continue with the story, the interviewer gave a clue to the learner about the word, for instance giving a usual collocation as it can be seen in Example 9 ('hello and goodbye') or giving him/her the first letters of a word (Example 10).

Example 9 (0151 A3)
*RAQ:        the [/] the mother # mm cómo se dice <se despiden> ["] . [+ s]
*RAQ:        [=! low tone] or +...
*INT:        ok how you say **hello and +…**
*RAQ:        **++ goodbye .**

Example 10 (0712 B3)
*J:      in the basket and <they are> [//] they surprise and then she realise that the
         basket is [/] is vacía no em surt . [+ c] [+ s]
*INT:   **emp**(ty) .
*J:      **empty** and because the dog er had eated [/]eated the food .

5) Mistakes

Mistakes were corrected and included in the final word counts. They were normally the result of overegularisation as shown by 'childrens' in Example 11.

Example 11 (0116 B1)
*DIA: the **childrens** going to mountain .

The same happened in the other oral tasks: 'homeworks' or 'borned' (instead of 'homework' and 'born') in the interview, 'furnitures' and not 'furniture' in the roleplay.

As evidenced in Table 4.5, all the categories that were included in the analysis did not represent a high proportion if all the production is taken into account, nor did they severely affect any particular group as regards the amount of tokens computed for the analyses. Some groups tend to have higher percentages than others (A1 has a 3.49% of words pronounced incorrectly while B2 has 0.46%), but even in these cases the figures do not diverge considerably, with total amounts that do not surpass a 0.74% for all the groups together.

There were also several words that were not included in the final word-counts, they belong to categories 6-8. Categories 6 and 7 were excluded because they were not formed by English words. If category 8 had been included, it would have probably overestimated learners' production, especially at low levels. These categories are defined and exemplified below:

6) L1 words

Those words that were non-L2 words, like Catalan/Spanish words within or without an English utterance, were not taken as part of the L2 productive knowledge.

Example 12 (1461 A3)
*MAR: **els nens** prepare it for a excursion . [+ c]
      [...]
*MAR: watch **un plano** . [+ s]

Neither 'els nens' (the children) nor 'un plano' (a map) in Example 12 were included in the analysis. L1 words represented a 11.43% of the total number of tokens produced by all groups (a 40.90 % in A1 and a 1.55% in B3 as shown in Table 4.5), but the study of code-switching was beyond the scope of the present dissertation (see instead Navés, Miralpeix & Celaya, 2005) . This decision was different from the one taken by Broeder et al., who counted non-target words if they belonged to target language utterances, but they do not include non-target language utterances. However, for the present study, it was considered that the use of L1 was a clear indication that the subject could not produce the words in the target language, even if it was in an English utterance as in Example 12.

7) Lexical inventions

There was a low number of non-L2 words that were invented, for instance 'cistel' and 'cest' for 'basket' (Example 13), 'veik' for 'cow' (Example 14) and 'picolos' for 'children' (Example 15).

Example 13 (0405 A2)
*DAN:          brother and sister <are sun day> [?]  xxx the dog a **cistel** [?] .
(0158 A2)
*HEL:          bueno on the jam in the: mm the: xxx un moment . [+ s] [+ c]
*HEL:          in the **cest** .

Example 14 (0566 B1)
*F:            ah sí!, the girl .. diu good-bye a the mother . [+ c]
*INT:          OK, and now ?
*F:            the **veik** .
*INT:          cows .
*F:            the cows eating mm ...

Example 15 (0950 A2)
*AGU:          the mm **picolos** no ? [+ i]
*INT:          yes the children these are boy and a girl good .


Other instances can be 'sorpresive' for 'surprising' in the storytellings or forms

like 'impersonality' in the interviews.


8) Immediate repetition


Immediate repetition or imitation took place when the interviewer gave the

subject a particular word that the subject did not know and the student repeated it

immediately in the next turn. It is evident when it is clear that the subject cannot produce

the word alone (i.e. s/he asks for it, says it in L1...). Example 16 provides an example

of immediate imitation of the word 'map'.


Example 16 (0208 A3)
*EST:          second mm # **what is this** [=! signalling the picture] ?
*INT:          what is what ?
*INT:          a map .

        *EST:         ah yes mm ah the: mother is: showing a **map** # mm # for: # saying <to the childs .

Sometimes, the immediate repetition varied slightly (as it also happened with non-immediate repetitions), but none of the immediate imitations were computed for the analysis, even if they included some sort of variation. This is what can be seen in Examples 17 ('to put' and 'put'), 18 ('countryside' and 'country') and 19 ('cow' and 'cows').

Example 17 (1080 A3)
*INT:        what do you want to say ?
*ION:       estan posant el menjar dintre la bossa . [+ c]
*INT:        **to put** .
*ION:       **put** the [/] put the bag mm after the mother # eh talked them the way +/

Example 18 (0529 A2)
*INT:        in the **countryside** .
*LOL:       +, in the **country** mm breakfast .

Example 19 (0146 B1)
*MAR:      it's mountain of look vacas . [ + s]
*INT:        **cow**
*MAR:      **cows** it's boy of the girl mm <berenar> com es diu ["] ? [ + c]

Incorrect immediate repetitions (as in Example 20) were not taken into account for the final counts either:

Example 20 (0196 A2)
*SIL:  com es diu que parlaven ["]? [+ c]
*INT:   **talk** ["] .
*SIL:   **dolk** [= talk is pronounced as /dolk/] and mother .

This classification is thus different from that of Broeder et al. in the sense that they counted word repeats (both self-repeats and other repeats) in their study as part of the productive word supply of the L2. The reason to do so is that, according to them, the distinction between imitated-non imitated words is arbitrary and unworkable. The evidence they put forward, though, can also suggest other alternative explanations for not including this immediately repeated word as part of the productive supply.

Firstly, they assert that imitation can take place from early utterances, not just the previous one. Certainly, as explained above, one can never be sure if something produced spontaneously by the interviewer was being imitated or was already known by the learner, in any case, it represented a 0.46% of the data. Secondly, they point to the fact that imitation may not take place successfully. However, when that happened, incorrect imitation was not taken into account for the present study.

Thirdly, they consider that it is possible that at the moment of successful imitation the word form in question becomes part of the learner's lexicon and that therefore it should be counted as such: as a word known by the learner. They consider that imitation is the confirmation that a form has been acquired.[31]

Nevertheless, in the type of data being analysed in the present study, it is usually one-word forms that students ask for and therefore immediate imitation is correct nearly all the time without necessarily meaning that the form has been successfully acquired, as evidenced by the fact that they may ask for the same form later on. Hence, it is not

---

[31] This idea of imitation being a sign that a form has been acquired can also be found in L1 grammar studies. For instance, in the research by Kuczaj and Maratsos (1975), imitation was essential to know if an aspect was acquired or not: a child had to repeat long sequences that could not be memorized as a chunk; thus correct imitations meant that the subject had acquired the target form, not that s/he had just memorised it, even though they would not produce these forms spontaneously. However the present study does not deal with long sequences or chunks but with single words.

always the case that the word form in question is incorporated into the learner's lexicon at the moment of successful imitation. It is more probable that the word is considered as having been integrated into the lexicon if it is just not repeated immediately but used later on in the conversation (with or without variation).

The fourth reason that Broeder et al. give is that in computerised concordance and frequency lists of large stores of data imitation or repetition cannot be controlled. Actually, it is true that in order to identify these phenomena  manual work or a specific type of tagging is required[32]. They suggest that the number of hapaxes[33] are a negative indication for productivity, as most of the time they are direct imitations. While this might hold in some cases, it does not always happen with young learners' corpora, where words produced just once are not usually direct imitations but the only words the subject knows in a particular task. All in all, as these authors themselves acknowledge, the inclusion of word repeats of any kind (immediate or not, of words known or not) will probably lead to an overestimation of the total productive vocabulary of the learner.

Therefore, self-repeats like the one in Example 21 were not taken into account in the word count, as in Malvern and Richards' study.

Example 21 (0454 B2)
*FER:  preparing a basket **with** [/] **with** some bread **and** [/] **and** sandwiches .

---

[32] This is probably why they also count L1 words if they are inside target language utterances, a thorough non-computerised revision is needed to delete them from large data files.

[33] Tokens that just appear once in the whole text/corpus.

Similarly to Broeder et. al and Malvern and Richards, false word starts (when the learner interrupts himself or is interrupted by somebody else) were not included. According to Broeder et al., false word starts reflect incapacity as they may be imitations of word segments of the NSs, but this is not the case in the data being analysed here, where false word starts can more often be interpreted as a way to hold the ground in the conversation or a device to have time to plan the next part of the utterance.

Following also these two studies, laughter and fillers ( hesitations, pause markers like *mm, eh, ah*) and onomatopoeias were filtered out and were not included in the final analyses. Non-completed words, which normally occur in the shortened form (e.g. *disco-discotheque, vet-veterinary, cos-because*) were turned into their completed forms, and words with more than one spelling (e.g. *programme-program*) were consistently changed into one. Also the inconsistencies with spelling and phonetic variants of the same word were standardised (e.g. *yeah-yes, gonna-going to*) and contractions were turned into their complete forms (e.g. *isn't-is not, she's-she is*). Following Broeder et al., compound nouns were counted as one lexical entry (e.g. living-room, high-school). Lastly, contrary to Richards and Malvern, homographs were not kept, i.e. they were not taken into account for the final count. Thus 'may', for instance, might be taken as a month or as a verb. Anyhow, similarly to what happened in Laufer and Nation (1995),  there were very few cases of homography in the data (basically the pronouns or determiners *her* and *his*). Another reason for not computing homographs was the presence of mistakes like the one in Example 22, where 'eat' cannot be categorised as a verb but it is not the noun either, as the correct one would be 'food'.

Example 22 (0402 A2)
*JUA:  mm prepare the **eat** .

*4.3.4.1.2. Written data*

English cloze

The contextually appropriate scoring method was followed to score the English cloze (Oller, 1972). 1 point was awarded for each correct answer and no points were given when the answer was incorrect. Therefore, the highest score for this test was 30 (25 for those students from Grade 10, see note 26). The final scores were turned into percentages for the statistical analyses.

There were several errors that were not taken into account. First of all, errors of grammar: present tense instead of past tense or plural form of present tense instead of the singular form (e.g. *meet-meets*) were not computed as mistakes, as we saw them as more closely related to grammatical accuracy than to vocabulary knowledge. However, negative forms instead of affirmative forms were not accepted, as it shows that students have not understood the content. Spelling mistakes (*'wat'-'what'*, *'bery'-'very'*) were not discounted either and adjectives ended in 's' (e.g. *'bigs', 'blacks'*) were also given 1 point. Also words like '*a hundred*' (referring to age) were taken as one unit as it shows understanding of the context but other instances like '*the red*' were not accepted: when two words appeared in one blank, it was taken as void, even if one of them was right.

Composition

For the analysis of the written production, the same conventions adopted for the oral data were followed as regards contractions, non-words in the L2, inconsistencies with spelling and compound nouns. Spelling mistakes that allowed us to understand the word were corrected (e.g. *'writter'*-*'writer'*, *'nouse'*-*'nose'*, *'serie'*-*'series'*). In order to anonymise the compositions and to avoid overestimating the total amount of tokens (for instance by having a very large amount of different proper nouns), the conventions devised by Navés (2006) for proper nouns were adapted and followed in the data analysis (see Appendix C).

*4.3.4.2. Description of the studies and statistical analyses performed: Some preliminary considerations*

Three main studies were conducted with the data, which are reported in chapters 5, 6 and 7 respectively. Chapters 5 and 6 are of a descriptive type while in chapter 7 an inferential study is carried out.

The SPSS statistical program (*Statistical Package for Social Sciences* v.11.5) was used to analyse the results obtained. If not indicated otherwise, an alpha level of .05 was chosen as the significant level. Also the non-significant and significant results have the exact probability levels reported in each case. Normality tests were carried out for each study, although in a general data distribution study an outlier was identified in group A3[34] and as

---

[34] It was seen in the 5% trimmed mean that it was especially this subject who had an obvious influence on the group mean.

a result the N dropped from 58 to the final 57. Assumptions other than normality for the specific statistical tests (Hatch & Lazaraton, 1991) were checked previously to the performance of each test. The particular descriptions of the statistical analyses (parametric or non-parametric) together with their respective results will be found in the corresponding chapters. The data which could not be statistically analysed (either because the low *N* or because the nature of the study was descriptive) will be presented in turn in the corresponding chapters as well.

A final note should be devoted in this section to mention that other computer programs were employed in the data analyses. Chandler's *Longman Mini Concordancer* (LMC) version 1.01 (1989) was used to obtain different types of word lists. It works on MS-DOS Operative System as well as the *Lexical Frequency Profile* (LFP) computed with *VocabProfile* (Nation, 1995a)*,* which was used to obtain the vocabulary profiles for the different learners and tasks according to Laufer and Nation (1995) in chapter 6. Also in chapter 6, *P_Lex* v. 1.1 (Meara, 2001) was applied in order to obtain the lambda values of the texts and *D_Tools* (v. 1.0) was specially created to analyse data in .txt format (Meara & Miralpeix, 2004). Both *P_Lex* and *D_Tools* are written in Delphi 6, as are also *V_Size* (Meara & Miralpeix, 2004)*,* the exploratory tool used in chapter 7, and *V_Tools* (v. 0.2), which helped to create the profiles for the estimations study. A detailed account of the reasons to use these programs, their creation (if it is the case) and development will be given in the respective chapters, as well as the way they were applied and the results obtained.