UNIVERSITAT DE BARCELONA

Master Thesis

VOCABULARY ASSESSMENT IN STANDARD TASKS

Student: Federico Zuddas

Master: Applied Linguistics and Language Acquisition in Multilingual Contexts University of Barcelona, English Department Supervisor: Dr Imma Miralpeix February 2010

Table of Contents

Abstract

Research on tasks carried out so far has usually neglected lexical aspects of learners' production and it has normally dealt with English as a target language. Besides, only few studies present a native speakers' baseline to compare the results obtained by the learners. The aim of this study is to analyse how two narrative tasks can influence lexical performance and how this performance can be assessed with intrinsic and extrinsic vocabulary measures. A total of 35 Italian native speakers and 2 groups of 35 Catalan/Spanish learners of Italian (intermediate vs. advanced levels) took part in the study by writing two different stories. Results show that the tasks with more elements elicit more vocabulary and more lexically diverse output than the task with less elements. Results also indicate that the two tasks used can discriminate across proficiency levels and shed light on research related to measurements issues.

KEYWORDS: *Lexical Frequency Profile*, lexical richness, lexical sophistication, narrative tasks, vocabulary assessment.

1. Introduction

Since the introduction of task-based learning in the 80s, the role of tasks in second language learning, teaching and testing has been the focus of interest in many studies. One of the areas of investigation has extensively dealt with the effects that the manipulation of certain task features has on task performance. Other studies have concentrated on the role of tasks in second or foreign language (FL) assessment. Research has also dealt with vocabulary acquisition and production, as tasks are valid and effective tools to promote the acquisition of words and to assess vocabulary gains when learning a language. However, little attention has been given to vocabulary assessment and measurement in the literature on tasks published so far.

Therefore, the main interest of the present study is the relationship between standard tasks and vocabulary assessment, with attention to the different measures and tools used in research to assess vocabulary in task performance.

1.1. Task effects on language performance and acquisition

Over the last ten years, part of the research on tasks has taken a cognitive approach to task performance. The cognitive complexity of a task and the ways attention is deployed during task completion have been topics of interest in SLA studies (Gilabert, 2005, 2007; Laufer & Hulstijn, 2001; Robinson, 2001; Robinson & Gilabert, 2007; Moonen, 2008) and quite recently there has been a debate in the literature about the effects of task types on the linguistic performance. Most of the research carried out on task-complexity effects on learners' performance has focused on different dimensions of learners' production, such as fluency, accuracy and structural and lexical complexity (Skehan, 2009b). Different theories have been taken as the ground upon which tasks are classified and learners' performance analysed. Of these theories, we would like to highlight two.

Skehan & Foster (2001) proposed the *Limited Attentional Capacity Model* and considered cognitive difficulty as a particularly significant characteristic of task design. According to this model, humans have limited information processing capacity and must therefore prioritise where they allocate their attention. If a task demands lot of attention to its content, less attention will be available to be devoted to the language required to accomplish the task. Therefore, more cognitively demanding tasks would result in poorer performance because less attention would be allocated on linguistic form.

Robinson's *Cognition Hypothesis* (2001) analysed the factors that can interact and influence task performance: task complexity, task condition and task difficulty. The factors concerning task complexity are represented as dimensions (plus or minus of a feature), but also as a continuum in which the endpoints stand for the presence or the absence of a certain feature. Robinson states that the dimensions of task complexity are design features of tasks, and can be manipulated in order to increase or lessen the cognitive load of a learner during task performance. He distinguishes between two groups of dimensions that can be manipulated to increase task complexity: resource-depleting and resource-directing dimensions. The *Cognition Hypothesis* claims that increasing task complexity can have different results depending on what dimensions of the task are manipulated. Increasing task complexity along the resource-depleting dimension will lead to greater fluency, while increasing task complexity along resource-directing dimension will lead to gains in accuracy and lexical complexity, but at the expense of fluency.

There are many studies, grounded on the cognitive paradigm of SLA and on the theories of cognitive processes in an L2, that have investigated the effects of pre-task planning (Ellis, 2005); online planning (Yuan & Ellis, 2003) and several other aspects of task design, task condition and task performance (Foster, 2001; Foster & Skehan 1996; Foster & Tavakoli, 2009; Gilabert, 2005, 2007; Robinson, 2001; Skehan & Foster, 1999; Tavakoli, 2009b; Tavakoli & Foster, 2008). In most of these studies, learners performed oral tasks and only in a few experimental studies task effects were analysed in written production (Kuiken et al., 2005; Kuiken & Vedder, 2007, 2008). In the analyses of performance, vocabulary is usually taken into account as a factor that could possibly vary.

It is also worth mentioning that if the hypotheses and studies above dealt with the effects of task designs on linguistic aspects of performance in general, there are two other hypotheses that deal specifically with tasks and vocabulary acquisition. The first has to do with task effects on the retention of vocabulary and was proposed by Laufer & Hulstijn (2001). The *Involvement Load Hypothesis* claims that three main components are crucial for vocabulary retention when a learner is performing a task: need, search and evaluation. The three elements induce the necessary involvement that helps learners better retain vocabulary while performing a task. The involvement load is defined as the combination of the presence or absence of the three involvement elements. According to the *Involvement Load Hypothesis*, words that are processed with higher level of involvement are retained better than words that are processed with less involvement load.

The second has also to do with task effects on lexical acquisition (retention and recall in this case) and was proposed by Westhoff (2004). The *Multi-feature Hypothesis* claims that differences in learning impact between tasks are to a large extend due to differences in the way that target language is manipulated in working memory during task performance. Westhoff distinguishes three characteristics of mental actions that are expected to ease the activation and foster the retention of vocabulary in the target language. He states that tasks that elicit mental action involving more features of the target language (in more different feature categories, in common combinations, simultaneously and frequently) will enhance the retention and ease of activation of the target language. In other words, it predicts that the acquisition of vocabulary is more effective when learners are engaged in tasks that present a combination of different features. The *Multi-feature Hypothesis* was tested by Moonen (2008) by comparing the effects of two different tasks, a rich one (in this case a writing task) and a poor one (a listening). The results of the study provided evidence that rich tasks led to better word retention and retrieval than the poor ones.

In sum, in task-based studies, vocabulary has been seen as one aspect in performance that could be influenced by the manipulation of the task design or as an element that could be more or less successfully retained and activated due to the type of task performed. In the next section, we will focus more specifically on how vocabulary has been measured and assessed in the task-based studies that have taken lexical aspects into account.

1.2. Tasks and vocabulary assessment

We mentioned in the previous section that in some studies certain task features are manipulated in order to check the effects in the learners' output, and one of the aspects measured is vocabulary. However, this measurement is not important just for research purposes. The role of tasks in language assessment is becoming more prominent, as the importance of the tasks as valid tools to assess second language (SL) performance has started to be widely recognised:

> "In line with the demands for language instruction to be based more on productive skills, second language assessments based more on receptive skills are likely to be replaced by performance assessments, in particular task-based ones." (Celik 2004, pp. 418-419)

Also Webb (2002) stresses the importance of different tasks to assess both productive and receptive vocabulary knowledge. Therefore, it is crucial to fall back on reliable measures that allow researchers not only understand the effects of tasks more precisely, but also to compare their results and establish valid forms of assessment that could also be useful for practitioners. In a vast majority of task-based studies, the amount of tokens is used as a measure of fluency and Guiraud's Index (and lately D by Malvern & Richards, 1997) have been the only ways to gauge lexical diversity.

Moreover, as pointed out by Skehan (2009a), in the last 20 years most of the research into SL learning tasks has focused only on SL learners. Apart from Foster (2001), which is a replication of Foster & Skehan (1996), it is not usual to find studies on task design or task assessment that include a native speaker (NS) dimension. According to Tavakoli & Foster (2008:463):

"[...] a proper perspective on task performance by learners of a second language needs a baseline native speaker's performance (Foster, 2001). Knowing how task design can influence a native speaker's fluency, complexity and lexical choices is an important triangulation for understanding the measures we take for non-native performance."

This, of course, has many practical implications; that is, for instance: how can we interpret a D of 45.5 if we do not know which D a NS would obtain when performing the

same task? Skehan (2009a: 107) adds that: "it is difficult to disentangle whether performances which are reported are the result of different variables which are being manipulated (e.g. tasks characteristics, tasks conditions) or simply SL speakerness of the participants". He also notices how lexical aspects of task performance are often omitted or dealt with using a restricted set of measures. In his meta-analytic research, Skehan shows that there is a need for experimental studies with a focus on lexical aspects of learners' and NS performance. He notices how rarely different lexical measures are used in a single experimental study and stresses the fact that there is little published on the relationship between measures of lexical diversity and lexical sophistication. He also claims that "this is a serious omission. The lexis-syntax connection is vital in performance models such as Levelt's, and lexis represents a form of complexity that has to be assessed in SL speech performance if any sort of complete picture is to be achieved." (2009b: 514).

We also feel there is a great need in the literature for studies focusing on the two issues outlined above: 1) standard tasks and how they are performed by both learners and NS of a given language and 2) lexical measures: research on vocabulary assessment has been extensive in the past decade but none of its outcomes has been adopted by studies on tasks. The only three studies that have very recently dealt with these points are Tavakoli & Foster (2008), Foster & Tavakoli (2009), and Skehan (2009a).

In Tavakoli & Foster (2008) 40 ESL learners in London and 60 EFL learners in Teheran were asked to tell different stories with the prompt of six drawings each. The study is set out to explore how differences in narrative structure (loose or tight) and storyline complexity (with or without background events) affect learners' output. Different measures were used to analyse the effects of task features on different dimensions of learners' speech production. The authors predicted that narratives with foreground and background events would be associated with greater lexical diversity. However, their prediction was only partially supported. By measuring lexical diversity using D (Malvern & Richards, 1997), they did not obtain clear evidence for their prediction. According to the authors, it is possible that the independent variable (+/-background events) might not be reliably connected to lexical diversity and that it could actually be the number of events what influences learners' production from a lexical point of view.

Partially responding to the lack of literature mentioned above, Foster & Tavakoli, (2009) repeated the experiment of Tavakoli & Foster (2008) with a group of 45 NS, who performed the same tasks as in the first study. As far as lexical diversity is concerned, the results followed the same patterns of the former study: there was no evidence that

background/foreground events significantly affected the performance in terms of lexical diversity. Results also showed that the learners in London, who were more proficient, were far closer to the NS than learners of English in Teheran.

In the meta-analysis that Skehan (2009a) conducts, enlightening conclusions are reached, such as that narrative tasks "provoke the most consistent difference in lexical performance between NS and NNS" (2009a: 119), which is not that evident with other types of tasks. He also analyses which lexical measures are more often used in task-based vocabulary assessment and suggests to explore the interrelationship between lexical richness and lexical sophistication, a central theme in his meta-analysis.

1.3. Vocabulary measures in task-based performance

In the SLA literature, lexical performance is generally assessed with *text internal* and *text external* measures (Daller et al., 2003). Other researchers have used the same distinction but with different terminology and divided measures into *intrinsic* (text-internal) and *extrinsic* (text external), depending on the source upon which the text is assessed (Meara & Bell, 2001).

Intrinsic measures are used when the assessment is carried out only in terms of the words that appear in a text. The most commonly used intrinsic index of lexical richness is the Type/Token ratio, which, as pointed out in Vermeer (2000), is sensitive to differences in text length. With measures such as Guiraud's Index the differences in text length are compensated as the total number of types is divided by the square root of the total number of tokens. Nowadays Malvern & Richard's D (1997), seems to be the best solution to problems encountered in quantifying vocabulary diversity. According to Mckee et al. (2000), the parameter D is shown to be a valid and reliable measure of vocabulary diversity, as it avoids sample size problems found with previous methods.

Extrinsic measures, on the other hand, assess the vocabulary used in a text in relation to language external to the learners' or speakers' production. In this case the measures are computed to assess to what extent the speakers draw upon a more varied lexicon by comparing output from the speaker with an external corpus of words. Some researchers refer to these indexes as "lexical sophistication" measures (Read, 2000). Two extrinsic measures of this kind are *Lexical Frequency Profiles* (*LFP*) and *Lambdas* (λ).

The *LFP*, probably the most well known extrinsic measure in vocabulary assessment, was developed by Nation (1995). Using four different word frequency lists with the

VocabProfile program, it is possible to calculate the percentage of words that belong to each one of the lists. The first list contains the most frequent 1,000 words in a language; list two contains the next most common 1,000 words; word list 3 is made of the next 1,000, and finally word list 4 contains all the words not belonging to any of the previous word lists. Therefore, the *LFP* gives information on how much a learner draws upon frequent and infrequent words while performing a task. The measure is presented as a valid tool to assess vocabulary growth over time and also as a tool that distinguishes between proficiency levels in a target language (Laufer & Nation, 1995).

A similar tool recently designed to assess vocabulary sophistication is Meara's *P_Lex* (2001). The program resembles *LFP* in the sense that it uses word lists external to the text to carry out the assessment. However, results are given in just one parameter instead of different percentages. The profile it computes first shows the proportion of 10-word segments containing 0 difficult words, the proportion containing 1 difficult word, the proportion containing 2 difficult words, and so on up to 10. The programme, which is based on the assumption that difficult words are infrequent occurrences in a text, calculates the theoretical Poisson curve which most closely matches the actual data produced from the text. The value obtained is called *Lambda* (λ), and indicates the degree of lexical sophistication the text presents.

Although there are studies in which intrinsic and extrinsic measures are used to assess non-native speaker (NNS) performance (Miralpeix, 2007), there are no studies on narrative tasks effects on learners' performance that make use of any of these lexical measures except from Skehan (2009b). The establishment of a set of narrative tasks for which lexical values or ranges could be obtained for different proficiency levels (and NS performance) can be a step forwards in the research on task complexity and task effects on learners' performance, as well as in task assessment.

We should notice, though, that most of the tools devised to measure vocabulary were originally conceived for English. Furthermore, most of the research conducted on task-based assessment has English as a target language as well. Kuiken et al. (2005) and Kuiken & Vedder (2007, 2008) are one of the few exceptions, as the languages analysed in their studies were Italian and French. However, apart from using a questionable variant of the *LFP*, they assess vocabulary by means of Type/Token measures and the peculiar tasks used in their studies make it almost impossible to compare results with other studies in the field. In the present piece of research, we wonder to what extent the tools designed for English can be adapted and used to measure vocabulary in a Romance language such as Italian.

Therefore, the present study is set out with the aim to fill a gap in the literature on task effects on learners' lexical performance in narrative tasks. For this purpose, it uses the same tasks as in other studies (Tavakoli & Foster 2008; Foster & Tavakoli 2009; Tavakoli, 2009b). Nevertheless, the main focus will be on vocabulary measures in written output that can be used in task assessment. It presents a NS baseline which may help to put different measures into perspective and may help to better understand the results obtained in relation to the learners' proficiency levels. Finally, it also wants to see if tools designed for English can be adapted to other languages, Italian in this case.

More specifically, the research questions this study aims at answering are presented in the section below.

2. Research Questions

RQ1 Does a narrative task with more events elicit more vocabulary and lexically richer language than a task with less events? RQ2 Do these tasks discriminate effectively between proficiency levels? RQ3 How do the measures proposed and the adapted software tools behave with Italian?

3. Method

3.1. Participants

Participants in the study come from two different proficiency levels: Intermediate (G1) and Advanced (G2). There is also a group of NS (G3). Groups G1 (N=35) and G2 (N=35) are 70 Catalan/Spanish bilingual students of Italian at the *University of Barcelona* (UB) and at the *Escuela Oficial de Idiomas de Barcelona* (EOI). The groups differ in their level of proficiency in the target language: in G1 there are beginner students belonging to the courses *Lengua II* at the UB and *Lengua Italiana II* at the EOI. These students have received six months of formal instruction in the target language and very few of them have been to Italy for a short holiday. In G2 there are proficient students belonging to the last course of Italian at the UB (*Lengua Italiana IV*), and the last two courses of the EOI (*Lengua Italiana IV* and *V*). Most of them have spent time in Italy and have had extra exposure to the target language in a naturalistic environment. Group G3 is formed by 35 Italian NS who have

finished at least Secondary studies. Participants were assigned to G1 or G2 according to their level of proficiency, which was assessed by the institution where they attended Italian classes. The fact that students have to pass a language level exam to be admitted/registered at each institution/subject was a reliable indicator of their proficiency at the moment of data collection.

Group	N	Proficiency Level
G1	35	Intermediate
G2	35	Advanced
G3	35	Native Speakers

Table 1. Participants in the study

3.2. Instruments

Two narrative tasks were chosen for the experiment: *Walkman* and *Picnic* (see Appendix 1). Each of the tasks consists of six prompts describing a story. Both tasks are defined as complex by Tavakoli (2009a; 2009b) as they present both foreground and background events. *Walkman* and *Picnic* were chosen for the present study because they differ in the amount of elements they present: the former has more elements and the latter has less.

Although students were already placed in different proficiency levels by their institutions, it was decided to devise a questionnaire (see also Appendix 1) in order to obtain information both on students' linguistic background and the type and amount of exposure to the target language they had received. The purpose of its administration was to have groups as homogeneous as possible in terms of proficiency and exposure to the target language.

3.3. Procedure

Learners and NS were presented with the two narrative tasks introduced above. They were given the tasks at two different times within a time period of three weeks, in a random order so as to avoid any possible sequencing effect in the sample. Students were asked to tell the story in the comic strips as if they were telling it to somebody who could not see the

drawings. They had 20 minutes to write their description and they could have a copy of the comic strip while writing. They were not allowed to use dictionaries or to be helped by the teacher or other students and they were also asked to write as much as they could. In G1 and G2, the tasks were performed in class as they were taken by teachers as part of the curriculum. The questionnaire was administered after the students had finished with the first task. Altogether, 210 compositions and 105 questionnaires were collected for the present study.

3.4. Analysis

3.4.1. Intrinsic measures

All the composition were typed and saved in txt files. The transcription was made following some conventions in order to process the texts with the software tools we adapted. Then they were analysed with *D_Tools* version 2.0 (Meara & Miralpeix, 2007) in order to obtain the D values, which provide an index of how much a learner varies his vocabulary in a text. Using the *VocabProfile* software, the total amount of tokens and types for each composition was computed in order to assess which task elicited more vocabulary and to calculate Guiraud's Index.

3.4.2. Extrinsic measures

3.4.2.1. Measures for "target-like use"

Analyses were also conducted to compute the percentages of target-like vocabulary use. The compositions from the NS were processed to obtain two corpora of words required to perform the two tasks. Using these corpora the compositions of the learners in groups G1 and G2 were processed with the adaptation of the *VocabProfile* to calculate, for each composition, the percentage of tokens and types belonging to the NS corpus and thus to assess how much learners' use of the language resembled that of NSs. This percentage was called "target-like vocabulary use", and it was operationalised in these measures:

Target-like vocabulary use: tokens belonging to the NSs Corpus x 100 tokens
total nº of tokens

Target-like vocabulary use: types belonging to the NSs Corpus x 100 Types
total nº of types

A target-like version of Guiraud's Index was obtained by dividing the number of target-like types (the types belonging to the NS corpus for the same task) divided by the square root of the number of target-like tokens (the tokens belonging to the NS corpus for the same task).

Target-like Guiraud's Index: Number of target-like types √
number of target-like tokens

3.4.2.2. Lexical Frequency Profile

Two *LFPs* were obtained for each composition using the software V*ocabProfile* (Laufer & Nation, 1995). One profile was based on the entire Italian Corpus and one was based on the NS corpus obtained for the specific task. This decision was taken on the assumption that, in order to assess the *LFP* of a learner that performs a particular task, it would be especially informative to rely on the set of words that a native would use when performing the same task. It was decided to create word lists using types instead of word families grounding on the fact that knowing a word does not necessarily mean knowing all the words of a certain word family (Bauer & Nation, 1993). This is necessarily true for Italian, which compared to English presents far more inflections in its words due to word gender, number, or to the conjugations of the verbs.

The word frequency lists for Italian were based on the *Colfis corpus* (Laudanna et al., 1995). This corpus was chosen because it is based on written Italian and also because it is quite recent if compared with other existing corpora, which present a vocabulary different from the one used nowadays (Basti, 2007). The *Colfis* list was adapted to be used with the software, and all the words were listed following the principles used to transcribe the compositions as mentioned above (see also Appendix 2).
The *LFP* for Italian also uses three lists. The first list comprises the first 1,000 more frequent words in Italian (1K). The second list contains the words listed between 1,001 and 2,000 (2K) and the third list contains the words from 2,001 and 3,000 (3K). The fourth list

(4K) is formed by all the words not present in the first three. Using this set of lists, *LFPs* were computed for all the compositions.

In order to have the *LFP* based on the NS corpora, the corpus of 35 compositions for *Walkman* and the one for *Picnic* were used. As the two corpora are far less big than the *Colfis*, the division into frequency lists was changed as follows: List 1 contains all the words that occurred more than 50 times in the NS corpus for the specific task; List 2 contains all the words that occurred more than 10 times and less than 50 times in the corpus and List 3 contains the rest of the words used by the NSs. Finally, List 4 contains the words not present in the NS corpus but that are used by the learners. This division was not arbitrary but determined by the distribution of frequencies in the corpora.

3.4.2.3 P_Lex: Lambdas

Following the same procedure, 90 compositions (the 2 compositions of 15 representative participants in each group) were analysed with *P_Lex* (Meara & Bell, 2001), which provides an index of lexical sophistication of a text. This measure is claimed to be more effective with texts with more than 100 words (and the majority of the compositions in this study were actually longer). The first analysis was conducted using the adapted *Colfis* list of word frequency. Lambdas were also computed for the same compositions using the word frequency lists obtained from the NS corpora.

3.4.3. Statistical analysis

Using the SPSS statistical program for data analysis, normality was assessed and parametrical tests carried out as no serious violations of normality were found. To answer RQ1, a paired sample t-test was performed to compare the values obtained from the two tasks. To answer RQ2, a one-way Anova was conducted to compare scores of the three groups on the different measures. No statistical test was conducted for RQ3, as it was conceived to be answered by an exploratory study, however *LFPs* and descriptive statistics from *Lambdas* are examined.

4. Results

This section presents the results obtained from the analysis of the two tasks for the three groups presented. RQ1 was posed to analyse if a task with more events (*Walkman*) would elicit more vocabulary and a richer language than a task with less events (*Picnic*). As shown in Table 2, *Walkman* elicited more words and more lexically rich language. The number of tokens is always higher for this task. Moreover, G3 obtained higher values than G2 and both groups scored higher than G1 in the two tasks.

Group	N	Proficiency Level
G1	35	Intermediate
G2	35	Advanced
G3	35	Native Speakers

	G1		G2		G3	
Measures	P	W	P	W	P	W
Tokens	110.54 (44.92)	126.37 (37.09)	173.71 (44.83)	185.83 (44.95)	228.17 (91.16)	242.11 (88.25)
Types	64.71 (20.62)	74.06 (17.86)	103.23 (19.30)	113.83 (19.84)	129.63 (43.60)	141.09 (44.98)
D						

Table 2. Mean values of tokens, types, D and Guiraud's Index for the two tasks. Standard deviations are presented within parenthesis.
(P=*Picnic*; W=*Walkman*)

A paired sample t-test indicates that there are significant differences in the results obtained for the two tasks in all the measures: Tokens: t(104) =2.90, p= .004; types: t(104) = 4.20, p= .000; D: t(104)= 4.87, p= .000 and Guiraud's Index: t(104) =5.76, p=.000.

Table 3 presents the results from the analysis on the target-like vocabulary use with the measures proposed in 3.4.2.1. The mean percentage of target-like tokens in G2 is higher for *Walkman*, while the opposite happens in G1, where higher percentages are obtained for *Picnic*. However, the differences in the percentages between the two groups are not considerable. Both groups seem to rely to a great extent on the same set of words that the NSs used when performing the same tasks. There is a clear difference in the scores obtained for

the Target-like Guiraud's Index. Results show that G2 students (advanced) present higher scores than G1 students (intermediate) in both tasks. Moreover, in both groups the target-like Guiraud's Index is higher for *Walkman*, which is the task with more events. If we compare the results of the Guiraud's Index and those of the target-like Guiraud's Index, they follow the same tendency.

Group	N	Proficiency Level
G1	35	Intermediate
G2	35	Advanced
G3	35	Native Speakers

	G1	G2	G3

RQ2 was posed to discover if the two tasks discriminated effectively between proficiency levels. A one-way Anova (Table 4) was conducted to compare scores of the three groups on the different measures and it was seen that there were significant differences between groups in all of the measures for both tasks. A post-hoc Sheffé test indicated that the significant differences were found between the three groups, thus the measures discriminate between the three proficiency levels.

Group	N	Proficiency Level
G1	35	Intermediate
G2	35	Advanced
G3	35	Native Speakers

	G1		G2		G3	
Measures	P	W	P	W	P	W
Tokens	110.54 (44.92)	126.37 (37.09)	173.71 (44.83)	185.83 (44.95)	228.17 (91.16)	242.11 (88.25)
Types	64.71 (20.62)	74.06 (17.86)	103.23 (19.30)	113.83 (19.84)	129.63 (43.60)	141.09 (44.98)

Table 4. Results from the ANOVA for the measures used in the tasks for all groups. (P=*Pichie*, W=*Walkman*).

RQ3 wants to examine how software tools created to analyse English behave with Italian. The measures used in this section are all extrinsic measures obtained from comparison to parameters external to the learners' output and the two main software programs which were designed to process English were used to process Italian: *VocabProfile* and *P_Lex*. Figures 1, 2 and 3 show the percentages of tokens used by the three groups in the two tasks when compared to the *Colfis* corpus. The first bar (in blue) indicates the percentages for *Picnic* and the second (in red) the percentages for *Walkman*. Results provide evidence that all the groups produced an output in both tasks with very high percentages of words belonging to the first 1,000 words. The percentage of 1,000 words is higher in *Walkman* than in *Picnic* and for both tasks learners from G1 use more words from 1K than advanced learners (G2) and NS (G3). On the contrary, the percentage of words belonging to the fourth list (infrequent words) is higher when the level of proficiency rises, that is, NS show higher percentages than advanced learners, and both groups have higher percentages than the intermediate one. However, in this case only in G1 the percentage of infrequent words is higher for *Walkman* than for *Picnic*.

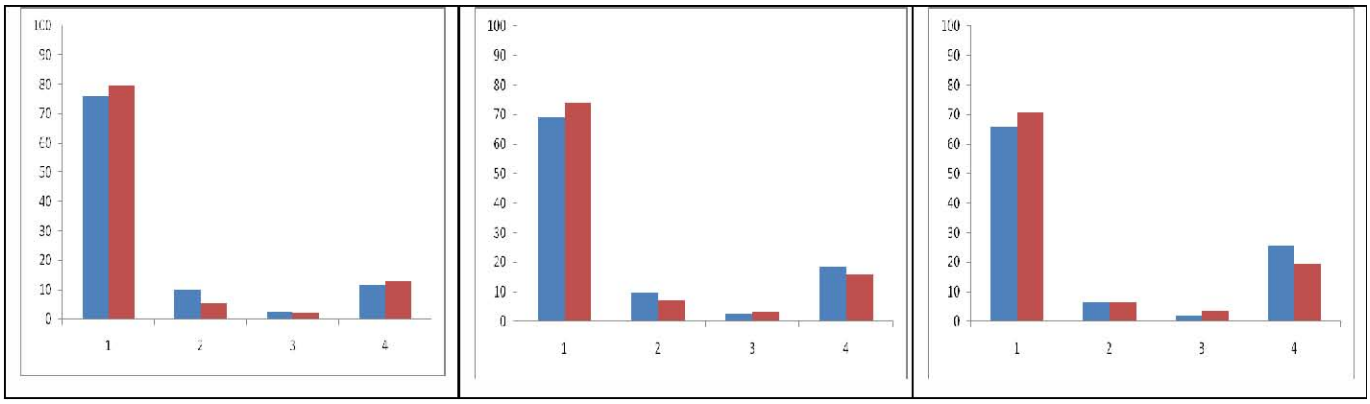


Figure 1. *LFPs* for G1 Figure 2. *LFPs* for G2 Figure 3. *LFPs* for G3

Figures 4 to 6 show the profiles processed with the lists obtained from the NS corpora specifically for these two tasks. Again, the percentage of words belonging to the first list is higher when the level of proficiency is lower, but the difference between the three groups are less marked than in the previous analysis. This time the words belonging to list two and three present higher percentages for all groups. As far the use of infrequent words is concerned, results show similar patterns with the previous analysis. In other words, advanced learners (G2) used more infrequent words in both tasks than intermediate learners (G1) and the

percentage is higher in *Picnic*. Intermediate learners (G1) use less infrequent words than learners from G2, but the percentage of infrequent words is higher in *Walkman*. Note that in this case the percentage of infrequent words is not presented for the NS group, as the analysis was conducted with reference to the corpus created from their compositions.

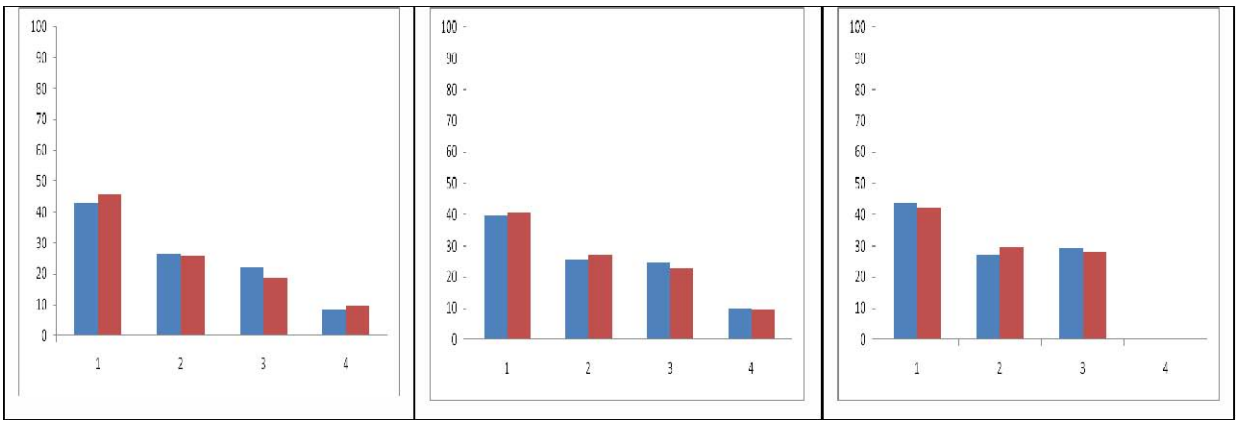


Figure 4. *LFPs* for G1 Figure 5. *LFPs* for G2 Figure 6. *LFPs* for G3

The results in Table 5 show the scores for *Lambdas* obtained when using the *Colfis* lists. It can be seen that the higher the mastery of the language, the higher the *Lambda* score. In other words, more proficient learners use more infrequent words in their compositions. It is interesting to observe that this time there are no regular patterns regarding the comparison of the two tasks. *Lambda* scores are higher in *Walkman* for G1, almost equal scores for the two tasks are obtained for G2 and the NS show the highest *Lambdas* in *Picnic*.

Group	N	Proficiency Level
G1	35	Intermediate
G2	35	Advanced
G3	35	Native Speakers

Table 5. Lambda scores using *Colfis* lists. (P=*Picnic*, W=*Walkman*)

The results in Table 6 present the scores for *Lambda* using the word frequency lists obtained from the NS corpus. In *Walkman* the values are higher for G3 (3.72) than for G2 (3.41) and G1 (3.1). However, for *Picnic* both G2 and G1 scored higher than G3.

Group	N	Proficiency Level
G1	35	Intermediate
G2	35	Advanced
G3	35	Native Speakers

Table 6. Lambda scores using the NS corpus. (P=*Picnic*; W=*Walkman*)

5. Discussion

This study was set to analyse the elicited production in two different tasks with the focus on vocabulary and lexical measures. A special attention was put on the tools used to assess vocabulary and their behaviour with Italian, which is not the language they were designed for. In this section we will draw some interpretations of the main findings related to the research questions of this study.

With regard to the relationship between task type and performance, there is clear evidence that *Walkman*, which presents more elements, elicits more words than *Picnic*. The number of tokens and types is higher for *Walkman* in all groups. All the standard intrinsic measures used to assess lexical richness showed that the task with more elements also elicits a richer vocabulary than the task with less elements, and in all cases the differences are significant. It is noteworthy that the mean scores for D of G2 (advanced learners) are closer to those of G3 (NS) than to those of G1 (intermediate). The same patterns are present for *Picnic*, where G2 and G3 have means of 84.55 and 89.31 respectively; which are higher than G1 that has a mean for D of 61.59. Results of the Guiraud's Index are in line with D results. Again, Guiraud's Index is always higher for *Walkman*, and G2 and G3 present higher means than G1. These results indicate that our group of advanced learners is not actually far away from performing the task like NSs regarding vocabulary, and this is important because NS may show higher/lower Ds in other tasks, but they would not be good points of reference in this case (and probably in task-based assessment in general).

Foster & Tavakoli (2009) assessed the oral performance of learners and NS using the same tasks in the present study. In line with the results shown above, their learners in a naturalistic environment (which also had a higher proficiency than at-home learners) presented means for D closer to the NSs and the means were higher in *Walkman* in all the groups. These results were not predicted in their hypothesis, which was based on the assumption that task performance was affected by storyline complexity and narrative structure. Their results do not show any clear relation in these two tasks dimensions. Nevertheless, they suggest in the conclusion that something other than background events and loose narrative structure (this specific dimension present in *Walkman* and not in *Picnic*), was causing differences in vocabulary production. They ascribe the grater lexical variety to the fact that *Walkman* has more events. This suspicion is confirmed in our data.

The results of the present study and those of Foster & Tavakoli (2009) can be interpreted in line with the Robinson's *Cognition Hypothesis,* which states that increasing task complexity along resource directing dimensions will lead to greater lexical complexity. In fact, *Walkman* presents more elements than *Picnic*, and the reasoning demands are higher in the former one.

RQ2 aimed at investigating whether the two tasks used could discriminate between proficiency levels. Results show that there was a clear distinction between the scores obtained by NSs, advanced learners and intermediate learners in all the measures. We think these results may be useful in promoting the creation of a set of standard tasks which would be a good tool both for language testing and assessment and for applied linguistics researchers, especially if we take into account Yu's (2009) results, which show that writing prompts may affect the lexical diversity of the written discourses. This idea could be linked also with the results obtained for native-like extrinsic measures that we were discussing above, as there is a point that seems to be particularly interesting: although the target-like Guiraud's Index (see also Table 3) is higher for G2 than for G1, the percentage of target-like types and tokens used by both groups of learners is close and quite high. We can assume that both groups rely on a set of words that is very close to those used by NSs as they present percentages of tokens around 90% and percentages of types around 85%. What is revealing about this data in comparison with the one obtained from NS is that it shows how lexical performance is taskconstrained. This may actually be common knowledge, as Gardner (2007:253) has pointed out:

> "In a collection of texts that share the same topic, the lexis employed will be used with the same meaning and the target vocabulary will be more fixed or more predictable."

However, this data shows that lexical performance is actually task-constrained to a very large extent. Actually, variability in amount of types and tokens that are native-like would just be around a 10% here. Therefore, the idea that the more proficient a learner is, the more target-like vocabulary he will use will actually not be of much use when assessing task performance. G1 and G2 present more or less the same percentages of target-like vocabulary use, although they belong to two different proficiency levels that other measures are able to discriminate between. From the data in this study, we can say that the NS baseline can be useful for measures such as profiles, for instance. Results also point out at the fact that maybe looking at the 10% of variation would be more useful to assess the vocabulary in a task than to look at the 90% of items which are normally present in all levels. For example, there are very specific places where more advanced learners systematically make use of particular words while less proficient learners do not. We exemplify this statement with a fragment of a composition of a learner from G1:

> [...]*una bestia sta guardando lui ma lui è molto tranquillo leggendo il diario* [...] *(a beast is looking at him, but he is very calm while reading the diary)*

The learner uses the word *bestia* (beast) while NS and more proficient learners used *tiger*. Then he uses the word *diario* (diary) instead of *giornale* (newspaper), probably due to the influence of the L1 in this case.

In RQ3, we wanted to explore if tools created to analyse English could be helpful to evaluate productions in Italian. Most of the studies on task performance and vocabulary acquisition dealt with English as a target language. The present study was carried out using tools that had to be set to process Italian as a target language and therefore some reflections are made about its functioning and appropriateness. The first tool used was the program *D_Tools* in order to obtain D values of lexical richness. We have already seen that after applying some changes to the txt documents in Italian in order to process them, D values were obtained and the results showed that the measure was able to discriminate between tasks and across proficiency levels with significant differences. However, if we compare this measure in two different languages (English and Italian), we realise that the highest D value

reported in Foster and Tavakoli (2009) for *Walkman* is 45.67 for the group of NS, while for Italian the highest D value in the same task performed by the NS is 100.66. What is more, even the lowest score reported for group G1 (intermediate) in *Picnic* is higher (55.48). These figures suggest that this measure is language dependent. This is a major point to take into account in the assessment of tasks: measures in different languages, even those that assess lexical richness, may vary. Task-based assessment, which will probably be widely used in a variety of languages in the near future should not forget this issue .

In order to obtain extrinsic measures of lexical sophistication, we made use of two lists. One was the *Colfis* list of word frequency in Italian and the other was obtained from the composition by the NS group. The lists compiled were used to draw *LFPs* and to compute *Lambda* values for all the groups.

Very few studies have used different extrinsic measures to assess vocabulary. Miralpeix (2007) used both *LFPs* and *P_Lex Lambdas* in her study in which English was the target language. The *LFP*s she obtained, as in other studies conducted with English as the target language, showed that in different tasks learners used almost 90% of the words belonging to the first 1,000 words in English. These results are different from those in the present study, in which the higher percentage of 1K words is 79.54 and the lowest is almost

70. This can be explained considering the differences between English and Italian and the word lists used in this study. Italian is a Romance language and has more inflections, for this reason the distribution of words is different in the 4 frequency lists. For example, as a single verb in Italian has inflections according to the tense and the pronoun he refers to, the composition that presents more verbs in the simple present tense (more frequent words in the list) will present a 1K bar higher than the compositions written in the past perfect tense (*passato remoto)*, as this tense is less used in Italian and the verbs inflected in this particular tense are far less frequent than those in the simple present. However, they will add more words to bands different from the first one.

The *LFPs* obtained using the *Colfis* lists show how the three groups rely to a great extent on a set of words belonging to the first 1K words in Italian. Comparing the three groups, the higher the level of proficiency, the lower the percentage of 1K words. Moreover, all groups used a higher percentage of 1K words in the more complex task. The *Colfis* list seems then to be a good list to obtain the profiles as it presents a steadily decreasing percentages in the results from 1K words to 4K.

However, the profiles with NS data are more informative, as words do not just cluster around 1K band. The distribution of words in this case has shown to be around 40-45% in 1K and 20%, 20%, 10% in the others, respectively. This means that the shape of the profile will vary depending on the lists we are comparing the text against. This is important to take into account, as it should also be noted that different curves and slopes of the profiles can influence the estimation of vocabulary sizes that take vocabulary profiles as their starting point (Miralpeix, 2008).

As far as *Lambda* scores are concerned, those obtained using the list of the *Colfis* corpus are higher for NS than for advanced learners, and both groups scored higher than intermediate learners for both tasks (see also Table 5). Thus, with higher levels of proficiency the level of lexical sophistication is higher in line with the results obtained for measures of lexical richness. Even though no statistical analysis was conducted for this measure, the fact that the results obtained using the *Colfis* are in line with those of lexical richness is a good indication that this list could be useful for vocabulary assessment.

The same measure obtained using the NS corpora for the two tasks does not show a clear distinction. One possible explanation can be put forward considering how the measure is computed. *P_Lex* divides the text into segments of 10 words and in each segment it looks for infrequent words. In this study we used lists adapted for two particular tasks, thus, the number of words that are infrequent is very low because most of them appear in the corpus of NS. It would be recommendable then to use a standard lists to compute *Lambdas*, as a general list reflects what happens in the language as a whole. NS lists are very task constrained and may not be adequate to compute *Lambdas*. Moreover, the NS lists were obtained from a corpus of only 35 compositions of NS for each task, which is a limitation of this study. As a consequence, the set of words that were used to create the lists was limited. It would be interesting to verify if with bigger NS corpora the results could differ.

6. Conclusions

Vocabulary in task-based assessment has quite often been neglected in SL studies. The research on tasks and vocabulary usually limits the analysis of learner's production to few measures of lexical richness and very few studies deal with extrinsic measures of lexical performance. Moreover, researchers have investigated mainly the acquisition of English as a FL and the tools used have been designed to deal with English.

This study intends to fill a gap in task-based vocabulary assessment providing a NS baseline to compare the results. It investigates learners' performance in Italian as a FL measuring several intrinsic and extrinsic measures of lexical performance. In order to do this, tools designed to analyse English were adapted to do the same with Italian.

Results provide evidence that a task with more elements elicits more words and richer vocabulary than a task with less events. They also make evident that, in the tasks used, intrinsic measures can discriminate across proficiency levels for both tasks providing a useful standard for future research in vocabulary assessment in Italian. Results also show how tools designed for English can work with this Romance Language and some reflections have also been provided in relation to lexical indexes, for instance: that some vocabulary richness measures can be language-sensitive, that lambdas may be more reliably computed using standard language lists (like *Colfis* in Italian), that indexes can vary due to the way the profiles are modelled and the word lists selected, that profiles using NS lists can be more informative or, finally, how important it is to interpret figures in relation to NSs' results for the same tasks.

Among the limitations of the study, the following can be mentioned: we are fully aware of the fact that the NS norms cannot be always taken to be representative of the highest levels of language performance and that the use of NS as models may not always be the best option in all language areas. As pointed out in Norris et al. (1998), these are always difficult aspects to decide on when trying to create valid procedures in language testing. Other shortcomings in this study include the number of tasks used, the amount of participants and of NSs from which the corpora was obtained and the lack of software tools already available to conduct analysis with Italian as a target language. Additionally, apart from the NS group, only two proficiency levels were assessed in the present work. Further research could be conducted using different types of tasks and comparing the results obtained in different target languages and in different proficiency levels. It would also be interesting to see if the oral performance for the same tasks would yield different results on lexical richness and sophistication.

In spite of the mentioned weaknesses, the present study can be an initial step of a possible investigation on vocabulary assessment in task performance, especially in languages other than English for which specific software tools are not available. These tools need not to be just adapted, but they should also have a solid theoretical basis that would allow a meaningful computation of results. This work would like as well to raise awareness of issues of validity and reliability of new (and not-that-new) measures that need to be systematically

investigated with larger amounts of data from different tasks and proficiency levels, and therefore this project would also like to promote more research in a field that needs further exploration.

Acknowledgement

I would like to thank Dr Imma Miralpeix for supervising this study and for offering time and valuable suggestions. I would also like to thank Dr Montserrat Casas Nadal, Diana Berruezo Sánchez and Carla Cesararo from the Department of Italian Philology at the University of Barcelona, and Nuria Picola from the Italian Department of the Escuela Oficial the Idiomas de Drassanes (Barcelona) for allowing me to collect the data in their classes and institutions.

References

Basti, I. (2007). Recenti acquisizioni della lessicografia statistico-computazionale italiana e anglosassone. Unpublished Degree Thesis, Università degli Studi Gabriele D'Annunzio di Chieti e Pescara.

Bauer, L. & Nation, P. (1993). Word families. *International Journal of Lexicography,* 6 (4), 253-279.

Celik, M. (2004). An investigation of second language task-based performance assessments. *Applied Linguistics*, 25 (3), 416-419.

Daller, H.; Van Hout, R.; & Treffers-Daller, J. (2003). Lexical richness in the spontaneous speech of bilinguals. *Applied Linguistics*, 24 (2), 197-222.

Ellis, R. (2005). *Planning and task performance in a second language*. Amsterdam: Benjamins.

Foster, P. (2001). Rules and routines: A consideration of their role in task-based language production of native and non-native speakers. In Bygate, M.; Skehan, P. & Swain, M. (Eds.). *Researching pedagogic tasks: Second language learning, teaching and testing*. (pp.75-93). New York: Longman.

Foster, P. & Skehan, P. (1996). The influence of planning and task type on second language performance. *Studies in Second Language Acquisition,* 18, 299-323.

Foster, P. & Tavakoli, P. (2009). Native speakers and task performance: comparing effects on complexity, fluency and lexical diversity. *Language Learning*, 59 (4), 866-896.

Gardner, D. (2007). Validating the construct of 'word' in applied corpus-based vocabulary research: A critical survey. *Applied Linguistics*, 28 (2), 241-265.

Gilabert, R. (2005) Task complexity and L2 narrative oral production. PhD Dissertation: Universitat de Barcelona.

Gilabert, R. (2007). Effects of manipulating task complexity on self-repairs during L2 oral production. *International Review of Applied Linguistics in Language Teaching* (IRAL), 45, 215-240.

Kuiken, F.; Mos, M.; & Vedder, I. (2005). Cognitive task complexity and second language writing performance. *EUROSLA Yearbook,* 5, 195-222.

Kuiken, F. & Vedder, I. (2007). Task complexity and measures of linguistic performance in L2 writing. *International Review of Applied Linguistics in Language Teaching* (IRAL), 45, 261-284.

Kuiken, F. & Vedder, I. (2008). Cognitive task complexity and written output in Italian and French as a foreign language. *Journal of Second Language Writing*, 17, 48-60.

Laudanna, A.; Thornton, A.M.; Brown, G.; Burani, C. & Marconi, L. (1995). Un corpus dell'italiano scritto contemporaneo dalla parte del ricevente. In S. Bolasco, L. Lebart & A. Salem, *III Giornate internazionali di Analisi Statistica dei Dati Testuali* (pp.103-109). Roma: Cisu.

Laufer, B. & Hulstijn, J. (2001). Incidental vocabulary acquisition in a second language: The construct of task-induced involvement. *Applied Linguistics* 22 (1), 1-26.

Laufer, B. & Nation, I. S. P. (1995). Vocabulary size and use: lexical richness in L2 written production. *Applied Linguistics*, 16 (3), 307-323.

Malvern, D. & Richards, B. (1997). A new measure of lexical diversity. In A. Ryan & A. Wray (Eds.). *Evolving Models of Language* (pp. 58-71). Clevedon: Multilingual Matters.

McKee, G.; Malvern, D.; & Richards, B. (2000). Measuring vocabulary diversity using dedicated software. *Literary and Linguistic Computing*, 15 (3), 323-337.

Meara, P.M. & Bell, H. (2001). P_Lex: a simple and effective way of describing the lexical characteristic of short L2 text. *Prospect. A Journal of Australian TESOL,* 16 (3), 5-19.

Meara, P.M. & Miralpeix, I. (2007). *D_Tools*, v. 2.0, Swansea: Lognostics.

Miralpeix, I. (2007). Lexical knowledge in instructed language learning: The effects of age and exposure. *International Journal of English Studies*, 7 (2), 61-83.

Miralpeix, I. (2008). *The influence of age on vocabulary acquisition in English as foreign language*. PhD Dissertation: Universitat de Barcelona.

Moonen, M. L. I. (2008). *Testing the multi-feature hypothesis. Tasks, mental actions and SLA*. PhD Dissertation: University of Utrecht.

Nation, I. S. P. (1995). *Vocab Profile*. Victoria Univeristy of Wellington.

Norris, J.M.; Brown, J. D.; Hudson, T.; & Yoshioka, J. (1998). *Designing second language performance assessments*. Hawaii: University of Hawaii at Manoa.

Read, J. (2000). *Assessing Vocabulary*. Cambridge: Cambridge University Press.

Robinson, P. (2001) *Cognition and second language instruction*. Cambridge: Cambridge University Press.

Robinson, P. & Gilabert, R. (Eds.). (2007) Special issue on task complexity, the cognition hypothesis and second language instruction. *International Review of Applied Linguistics in Language Teaching* (IRAL), 45 (3), 161-284.

Skehan, P. (2009a). Lexical performance by native and non-native speakers on languagelearning tasks. In Richards, B.; Daller, H.; Malvern, D.; Meara, P.; Milton, J. & Treffers-Dallers, J. (Eds.). *Vocabulary studies in first and second language acquisition: The interface between theory and application* (pp.107-124), Palgrave: Macmillan.

Skehan, P. (2009b). Modelling second language performance: Integrating complexity, accuracy, fluency and lexis. *Applied Linguistics* 30 (4), 510-532.

Skehan, P. & Foster P. (1999). The influence of task structure and processing conditions on narrative retellings. *Language Learning*, 49 (1), 93-120.

Skehan, P. & Foster, P. (2001). Cognition and tasks. In P. Robinson (Eds.). *Cognition and second language instruction* (pp. 183-205). Cambridge: Cambridge University Press.

Tavakoli, P. (2009a). Investigating task difficulty: learners' and teachers' perceptions. *International Journal of Applied Linguistics*, 19 (1), 1-25.

Tavakoli, P. (2009b). Assessing L2 task performance: Understanding effects of task design. *System*, 37 (3), 482-495.

Tavakoli, P. & Foster, P. (2008). Task design and second language performance: The effect of narrative type on learner output. *Language Learning*, 58 (2), 439-473.

Vermeer, A. (2000). Coming to grips with lexical richness in spontaneous speech data. *Language Testing*, 17 (1), 65-83.

Webb, S. A. (2002). *Investigating the effects of learning tasks on vocabulary knowledge*. PhD Dissertation. Victoria University of Wellington.

Westhoff, G. (2004). The art of playing a pinball machine. Characteristics of effective SLAtasks. *Babylonia* (3), 58-62. Retrieved from www.babylonia-ti.ch

Yu, G. (2009). Lexical diversity in writing and speaking task performance. *Applied Linguistics*, 24 (1), 1-24.

Yuan, F. & Ellis, R. (2003). The effects of pre-task planning and on-line planning on fluency, complexity and accuracy in L2 monologic oral production. *Applied Linguistics*, 24 (1), 1 27.

APPENDIX 1

1. Tasks
Walkma
n

Picni
c

1

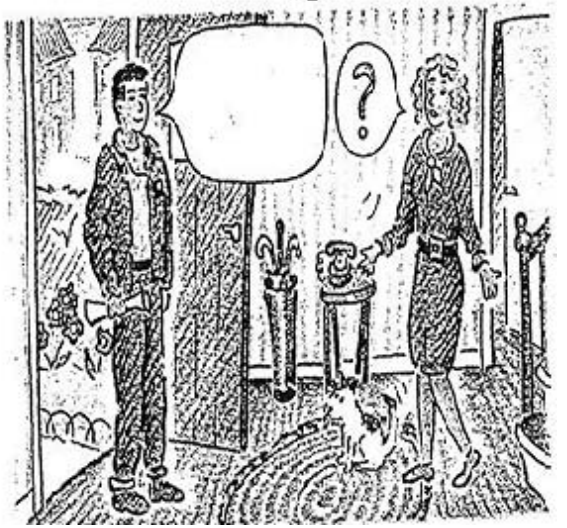
2

3

4

5

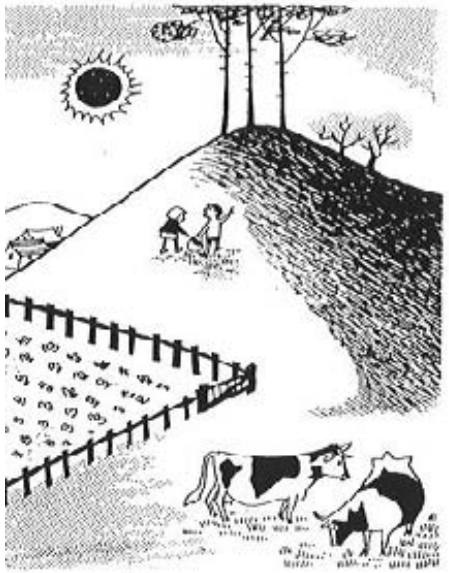
6



1

2

3



4

5

6

2. Questionnaire

Scuola/Istituto_____

Classe/Gruppo_____

Nome:_____

Cognome:_____ Età: _____ anni.

Lingue che parlo: _____, _____, _____,

A che età hai iniziato a studiare la lingua italiana? A _____ anni Da quanto tempo studi la lingua italiana? _____ Sei mai stato in Italia ? ☐ Sì ☐ No Per quanto tempo? _____

Grazie per la tua collaborazione!

APPENDIX 2

1. Transcription Procedure

All the composition were typed and saved in txt files. The criteria followed in order to process the texts with *D_Tools*, *VocabProfile* and *P_Lex* can be summarized as follows:

a) all the accented words were typed without writing the last accented vowel (i.e.: città=citt; puó=pu);

b) mispellings were corrected (i.e.: ragazo=ragazzo; diseño=disegno; vinyetta= vignetta);

c) the homographs were counted as a single word (i.e.: cara= expensive or dear);

d) all the accents and similar signs and symbols were omitted (i.e.: po'=po; dell'altra=dell altra);

e) whenever the learners gave proper names to the characters of the picture prompts, the names were substituted with those previously chosen in order to have a single proper name for all the characters and to reduce the number of proper names in the NS corpora;

f) invented words (words without meaning) were not considered for the transcription.

2. Word lists adaptation

The word frequency list from the *Colfis* corpus was adapted in order to obtain three lists with the first 1,000 most frequent words in the first list (1K), the words from 1,001 to 2,000 in the second list (2K) and the words from 2,001 to 3,000 in the third list (3K). The original corpus contained different entries for homographs and it was decided not to consider them in this analysis, as the programs were not able to recognise which meaning the words had. It was also decided to erase from the *Cofils* corpus the chunks of words, as every word in each chunk was already present in the list.

As far as the NS list is concerned, all the compositions from the NS group were used in order to obtain the corpus for each task. The corpus for *Picnic* contains 9,001 tokens and 1,433 types; the *Walkman* corpus has 10,412 tokens and 1,599 types. The word frequency lists were created assigning the words to three different lists according to the number of occurrences of the words in the corpus. List 1 is made with the words that occurred more than 50 times in the corpus. List 2 includes the words which occur more than 10 times but less than 50 times in the corpus. List 3 is made with the rest of the words.

APPENDIX 3

Examples of compositions.

DESCRIZIONE

UN RAGAZZO che passeggia su un marciapiede e un tratto guarda le abitazioni con dei giardinetti fioriti in come si affaccia da un recinto. Continua la sua strada un attimo di lui accade un incidente fra due automobili: ~~una delle due auto~~ una Lancia ed una mucca la mucca perde una ruota ma ~~fortunatamente~~. Il ragazzo continua la sua strada senza voltarsi. Voltando scorgono una gioielleria dove un ladro ha appena rotto la vetrina. Il ragazzo continua ed ascoltano la musica nelle cuffiette senza accorgersi di niente. Attraverso un incrocio vicino ad un bancomat e un attimo di lui accade una rapina nella banca con polizia e sparatoria ma lui continua senza accorgersi di niente. Arriva ad un parco vede su una panchina e continua a leggere il giornale ascoltando sempre la radiolina con le cuffiette e non si accorge che sta piovendo una tigre. Infine rientra a Casa e la mamma lo guarda sbalordito come se gli avesse fatto una domanda difficile: in casa ed effettivamente c'è un caos con la mamma, come oggetti che ci sono in terra con bicchieri, un portaombrelli in terra e un tavolo una sedia per terra ed il vaso rovesciato. Lui è sulla porta dietro si vede un bel giardinetto fiorito. Ha in mano un giornale e la mamma gli chiede se si è accorto di tutto quello che il telegiornale raccontava: dell'incidente delle due auto, della rapina alla gioielleria della sparatoria in banca, della tigre nel parco, perché tutto ciò accadeva nella strada che percorreva per tornare a casa.

DESCRIZIONE

Nella Cucina la mamma prepara un the e lo mette nel Thermos c'è una finestra chiusa con delle tendine, un Tavolo con delle stoviglie in come che grande i Due Bimbi in maschio ed una femmina che preparano per una colazione sul prato la bambina affetta il pane per poi condirlo con burro e Marmellata, il Bimbo mette la merenda nel Cesto la mamma poi dà una carezza ai bimbi nel frattempo il coniglino è salito sul Tavolo dove c'è la Cesta con la Merenda ed incomincia a mangiare. I Bimbi escono per strada e salutano la mamma. Si arriva in un Campo dove ci sono degli alberi e due Mucche, la giornata è bella c'è il sole ed un grande orto un Cane una bella montagna. I bimbi si vedono felici e la bambina esulta ma quando aprono il Cesto si accorgono che nel cesto c'è il coniglino che nel frattempo Ha Mangiato la loro Merenda ed infine il bimbo Tiene il Thermos in Mano e TENTA e due guardano Stupiti dentro il cesto Vuoto. In Lontananza il coniglino come felice nel Prato.