

Investigation of Vocabulary in a Corpus of Written Production of Greek Learners of English

Chrysanthi Tiliakou*, Katerina T. Frantzi^

Laboratory of Informatics, Department of Mediterranean Studies

School, of Humanities, University of the Aegean

Rhodes, Greece

Abstract

This work focuses on the exploration of a learner corpus for the identification of the amount and level of vocabulary used in the writing production of learners and the exploration of the extent the vocabulary assigns these samples to the expected level of proficiency. The objective is to record the vocabulary appearing and to categorize it according to the Common European Framework levels. The corpus consists of 69 pieces of writing by 12 and 13-year-old learners of English in Greece, produced in the writing module of the KPG Exam, an exam battery for learners of Foreign Languages in Greece, for levels A1-A2. As a corpus-assisted study, key-word analysis is applied, identifying lexical items that do not present the same relevant frequency compared to the reference corpus. Since lexical criteria may account for L2 development, the analysis of the texts focuses on the lexical aspects of English and is performed using the English Vocabulary Profile Online tool (Cambridge University Press, 2015). The results showed that the larger part of content words and especially nouns, verbs, and noun-verb combinations used belonged to A1 level while a smaller amount was of A2 and a few words of B1 and B2 levels. The occurrence of the higher-level words correlated with the higher graded writings and the respective learners' higher proficiency.

Keywords: learner corpus, writing module, vocabulary, language levels

1. Introduction

Learner corpora have opened new pathways to SLA, since systematically collected, machine-readable, L2 material, produced orally or in written form by language learners, can now be processed via different software tools. Oral or written scripts, of different L1 learners and of different language proficiency levels constitute a valuable reservoir of interlanguage. Hence, interlanguage can be researched intricately and in a wider range. Only then can common features and processes be explored when “numerous learners’ linguistic data can be studied and compared” (Leńko-Szymańska, 2015, p.116). Several projects like English Profile have undertaken this endeavor and by detecting language use frequencies at different stages of language development they derived “criterial features”, features that signal different language levels (Tono, 2013). Profiling research offers an overview of the modifications of the learner’s performance from one level to the next. Studies have spotted that lexical/semantic, morpho-syntactic/syntactic, functional, notional, discourse, and pragmatic features are criterial

depending on their use or non-use, accuracy of use or frequency of use (Salamoura & Saville, 2009, p. 34, as cited in Tono, 2013). Profiling findings therefore can guide teachers and lesson designers in effectively and purposefully tackling learners' needs to help them advance from one language level to the next.

The delineation of levels in the Common European Framework for Languages rendered the concept of language proficiency measurable, which in concert with the instances of language use and language competences as they appear in the descriptors offer an invaluable tool to “practitioners to specify objectives and describe achievements” conditioned to different learner needs and characteristics (C o E, 2001, p. 5). Accordingly in language programs learners deal with the linguistic component of the Communicative Language Competence primarily. Within the linguistic component one of the competences to be achieved is the lexical competence, to know and be able to use “the vocabulary of a language” (C o E, 2001, p. 110).

Vocabulary learning and teaching is at the center of Second Language Acquisition. It is the foundation of learning a second or foreign language, as people empirically gather from their first attempts to communicate with speakers of other languages. Besides, a learner's progress becomes evident by the width of his/her vocabulary (Lessard-Clouston, 2013, Chapter 1). According to Nation (2008) successful acquisition of vocabulary involves both receptive and productive skills and requires the knowledge of the form, the meaning and the use of a word. Language proficiency on the other hand describes students' “communicative ability in real life” in reading, listening, writing and speaking on a continuum of levels from A1 to C2 (CoE, 2020, p. 29). Hence investigating this ability is a tool to understand and address the needs of students. Moreover, the relation of vocabulary and language proficiency is another functional piece of information in the development of student's interlanguage. Particularly the skill of writing is strongly associated with the size of the vocabulary a student has mastered (Staer, 2008)

In the current study ESL students' written productions within a KPG A level test exam battery are explored to examine the following:

1. Are the students in the expected band of proficiency level according to the syllabus for the 6th grade of primary and the 1st of secondary school in the Greek public school's system?
2. Is the students' use of vocabulary and particularly their use of verbs and nouns in their written productions in line with the achieved level of proficiency of their written production?
3. To what extent the use of higher-level vocabulary (specifically verbs and nouns) by a specific student agrees with the proficiency level of the student (as it is revealed in the rating of the whole test battery) or is it a transitional threshold towards higher- level proficiency bands?
4. Did students demonstrate control of the higher-level vocabulary?

This work is supported by APOLLONIS,¹ “the Greek Infrastructure for Digital Arts, Humanities and Language Research and Innovation”, recently formed by the union of two existing ESFRI-related national research infrastructures: CLARIN:EL,² the CLARIN-related

¹ <http://apollonis-infrastructure.gr>

² <https://www.clarin.gr/en>

Greek network for language resources, technologies and services and DARIAH-GR/DYAS,³ the DARIAH-related Greek network for digital research in the Humanities”.

1.1 Vocabulary knowledge

Linguistic competence according to CEFR comprises several aspects: General linguistic range, Vocabulary range, Grammatical accuracy, Vocabulary control, Phonological control, Orthographic control. It is interesting to note that for levels such as pre- A1 there are no descriptors for Vocabulary range, and for both pre- A1 and A1 there is no descriptor for Vocabulary control. In other words, no Vocabulary range is expected to exist in pre-A1 level and Vocabulary control not even in A1. The learners control their vocabulary when they can select the most suitable utterance in a given communication setting (CoE, 2020, p.129-132). Range can be related to vocabulary breadth, meaning the number of words a learner knows without taking into consideration how well s/he knows them. Vocabulary depth is “the knowledge a learner may have about how these words may work, their nuances of meaning and subtleties of combination” (Milton, 2010, p.219). This dimension then denotes the competence of the language users to use collocations, and idioms successfully (Gyllstad, 2007, as cited in Milton, 2010). One last dimension was added by Meara (1996, as cited in Milton, 2013) and developed by Daller et al. (2007, as cited in Milton, 2013), which focused on the ease with which a person can recognize and/ or use a word appropriately. It is the dimension of fluency.

The dimensions of vocabulary knowledge can explain performance variance among learners. If a learner has high vocabulary breadth but low depth and low fluency may be distinguished by a learner of the same vocabulary breadth with high depth and high fluency who will be more likely to form less frequent collocations for instance and faster. These dimensions may explain divergence in learners’ performance even of those with the same vocabulary range.

There has been extensive research to explore the contribution of lexical knowledge to learner’s performance. As several studies suggest (Staehr, 2008; Milton et al., 2010; Schoonen, 2010, as cited in Milton, 2013) “a single measure of vocabulary knowledge” can be “a good indicator of skill and level in a foreign language” (Milton, 2013, p.72).

2. Design and Methodology of the study

2.1 Rationale

Taking into consideration that children in the last grade of primary school are expected to have reached level A2 (Odigies Dimotiko, 2020), and at the end of junior high school level B1 (Odigies Gymnasio, 2020), it is fair to assume that level A2 can be mainly encountered in most of written productions of children of the last grade of primary and the first of secondary. Since all four skills need to be developed to attain a specified level, the writing skill, which is considered the most difficult should have been developed to the point at least of level A1. Furthermore, since vocabulary is in the core of language learning, it is interesting to examine to what extent the knowledge and use of nouns and verbs, the basic forms taught in levels A1

³ <http://www.dyas-net.gr/?lang=en>

A2 is observed in the students' texts. Given the fact that learners of different levels are included in our cohort, it is quite important to explore the existence of patterns of use of certain content words corresponding to the different levels.

2.2 Data

Accordingly, 69 students' writing pieces were studied. The pieces were written by students of the last grade of primary and the first of secondary from areas in Thrace and in the South Aegean. The writing tasks were part of the KPG exam battery in English of May 2012. The filled-in tests were used as research material for my doctoral dissertation "Classical Item Analysis for language tests in the context of multilingualism in education and towards the development of language proficiency of Greek speaking and Turkish speaking bilingual students of Public and Minority Schools in Greece" (Tiliakou, 2016).

All the collected compositions in this study were responses to the same topic. The task was part of the 2nd of 4 modules in total. It was the longest of the 4 activities composing this module. Two of them are considered A1 level and the other 2 A2 level (Dendrinou & Karavas, 2013). The learner corpus under discussion came from the A2 level part and it was an email to a friend where students were asked to tell them that they had moved to another city, why they moved and how they spent their free time. It was a guided task of 70 -80 words. KPG test was distributed to students during school hours and was supervised by the class teachers. After being collected the handwritten texts were typed into a word document which was then turned to a text document in order for it to be analyzed with a tool. The pupils came from the same national educational system, and their presumed language level was A1- B1 according to the estimations of the ministry which were mentioned in the previous section. So, even the authority designing the syllabus accepts variability among students' proficiency as a given. Consequently, the need to objectively gauge learners' levels becomes more obvious. The corpus was conditioned by another criterion, the number of words. Therefore, very short productions were excluded from the research, as texts shorter than 30 words cannot be reliably analyzed, as they are hardly considered to form a text. The small sample was also more convenient as the study involved a lot of manual tagging.

2.3 Level Tagging of Nouns and Verbs and Type- token ratio

My approach was to use the Antconc tool, where using the wordlist of the most frequently used words I picked out the verbs and nouns. Then I noted the level of each word from the English Vocabulary Profile online, depending on the use of the words in their occurrences in the text displayed in the file view of Antconc. 116 A1 verbs and nouns were found, 59 A2, 35 B1 and 7 B2. I used the text inspector tool but I resorted to the English Vocabulary Profile online tool too. The nouns and verbs were categorized to different levels according to the English Vocabulary Profile. The next step was to focus on the 25 texts which contained the B1 and B2 verbs, nouns and verb noun combinations in order to explore whether the existence of these words corresponded to the level of the written production or emerged as irrelevant.

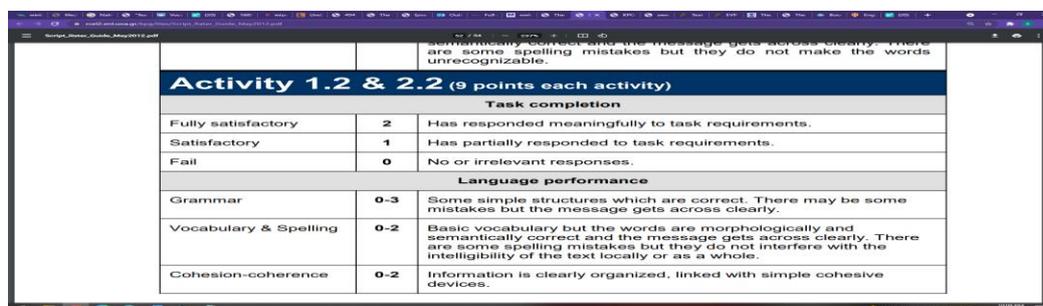
The metric of Type/ Token ratio (TTR) was also accounted for, which is traditionally used to gauge lexical diversity in learner productions. As the study focuses on the examination of vocabulary it was deemed that this criterion of a range of different words used in a text, where the wider variety is an indicator of higher lexical diversity should be reflected (McCarthy & Jarvis, 2010, as cited in Owen, Shrestha, & Bax, 2021). To measure the TTR, the statistics part

of the Text Inspector tool was used. According to Biber (1988, as cited in Granger & Wynne, 1999) a high TTR denotes a more informational style while a low TTR reveals a more involved style.

2.4 Assigning CEFR levels by raters according to KPG criteria

The texts were rated by two experienced teachers of English following the evaluation criteria stated explicitly in the Script Rater Guide. Along with the grid clarifying guidelines were provided in the Guide concerning task performance and language performance. Moreover, samples of fully, moderately satisfactory or unsatisfactory appeared in the Guide proposing solutions in a more targeted way. The raters' assessments consequently were mostly convergent, another reason being that their assessment would assign scripts to a narrow band of the scale namely A1-A2.

Figure 1: Rating Grid of the activity studied



Activity 1.2 & 2.2 (9 points each activity)		
Task completion		
Fully satisfactory	2	Has responded meaningfully to task requirements.
Satisfactory	1	Has partially responded to task requirements.
Fail	0	No or irrelevant responses.
Language performance		
Grammar	0-3	Some simple structures which are correct. There may be some mistakes but the message gets across clearly.
Vocabulary & Spelling	0-2	Basic vocabulary but the words are morphologically and semantically correct and the message gets across clearly. There are some spelling mistakes but they do not interfere with the intelligibility of the text locally or as a whole.
Cohesion-coherence	0-2	Information is clearly organized, linked with simple cohesive devices.

Source: (S. Cert. of Lang. Prof., 2012 p.50)

3. Results

The texts with B1-B2 nouns, verbs and noun-verb combinations were 25. The analysis of these texts with the English Profile tool led to the following:

In the total of 69 written productions there were 4120 tokens and 554 types. I examined the wordlist sorted by frequency and focused on the content (or lexical) words. Content words are the words that carry semantic meaning, the crucial part of information being conveyed in a sentence. Function words on the other hand express grammatical functions, in other words they convey abstract information and, in a sense, regulate the content words of a structure, rendering such dimensions as number, modality, definitiveness and others (Corver & van Riemsdijk, 2001 p.1). 115 A1 verbs, nouns and verb noun combinations were found, 58 A2, 35 B1 and 7 B2 such items (see Appendix 2). The rest were names of countries or cities like Australia, Greece, words related to nationality like English, words like snowman, basketball, sport or proper names like Casuarina. None of these words yielded any results on the English Vocabulary Profile tool. The B1 and B2 words appeared in 25 texts as the file view of Antconc revealed. These texts were checked for their grade, and whether the whole of the test had been assigned A1 or A2 level in the intergraded KPG test. The word moved was not counted as another B1 word, as it was included in the prompt of the task and almost all the students of the corpus used it.

Figure 2: Prompt of the activity under examination

ACTIVITY 2.2

Imagine you have just moved to another city with your parents. Send an email (70-80 words) to your friend Ann to tell her about it. Sign as Alex (not with your real name).

Φαντάσου ότι μόλις μετακόμισες σε μια άλλη πόλη με τους γονείς σου. Στείλε ένα email (70-80 λέξεις) στην φίλη σου την Άνν, και πές της σχετικά με τη μετακόμισή σου. Υπόγραψε ως Alex (όχι με το πραγματικό σου όνομα).

In your e-mail tell her:

- which city you've moved to & where it is
- why your family had to move there
- if you like it & how you spend your free time

Source:(M. of Education, 2012 p.5)

Figure 3: A1 tests with B1-B2 words graded from 5 to 17 out of 18.

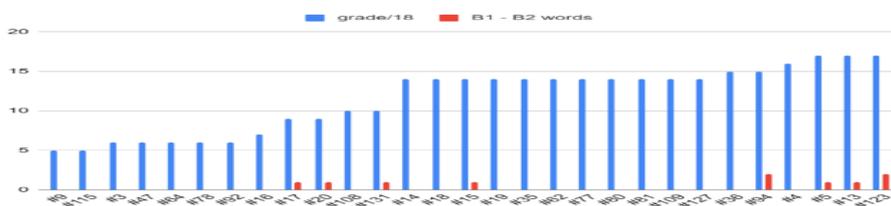
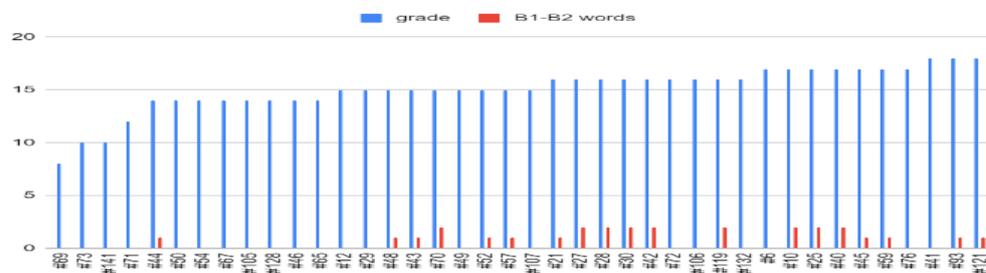


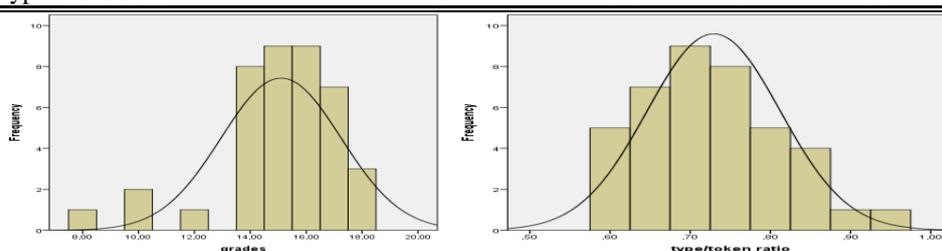
Figure 4: A2 tests with B1-B2 words graded from 8 to 18 out of 18.



(SD=2.146) ranging from 8 to 18. The type/token ratio mean value equaled 0.73 (SD=0.083) ranging from 0.60 to 0.93. In both cases SD was low but none of these measures had symmetrical distribution (graph 1)

Table 1. Measures of central tendency and dispersion of grades and type/token ratio.

	Mean	Median	Std. Deviation	Range	Minimum	Maximum
grades	15,10	15.00	2.146	10.00	8.00	18.00
type/token ratio	0.73	0.72	0.083	0.33	0.60	0.93



Graph 1. Frequency Histogram of grades and type/token ratio

According to table 2, 50% of the B1-B2 variable contained none whereas the percentage of occurrence for 1 and 2 words was similar: 27.5% for 1 word and 22.5% for 2 words.

Table 2. Frequency table of B1-B2 words

	Frequency	Percent	Valid Percent	Cumulative Percent
Valid	,00	20	50,0	50,0
	1,00	11	27,5	77,5
	2,00	9	22,5	100,0
Total	40	100,0	100,0	

According to Pearson's correlation coefficient (table 3) the grades of A2 texts correlate significantly with the type/token ratio ($\rho=-0.324$, $p=0.041$). This negative and weak association shows that text A2 grades are expected to decrease for higher values of type/token ratio. On the contrary, the association between grades and B1-B2 words is positive, medium in strength and statistically significant ($\rho=0.426$, $p=0.006$) and shows that text A2 grades are expected to increase on B1-B2 word count. Finally, the correlation between type/token ratio and B1-B2 words, was positive, weak, and non-statistically significant ($\rho=0.144$, $p=0.375$).

Table 3. Pearson's correlation between grades, types/token ratio and B1/B2 words.

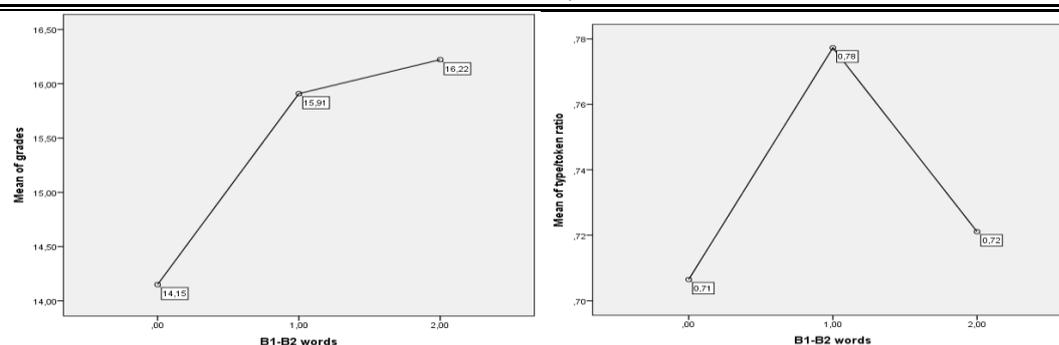
		grades	type/token ratio	B1-B2 words
grades	Pearson Correlation	1		
	Sig. (2-tailed)			
type/token ratio	Pearson Correlation	-,324*	1	
	Sig. (2-tailed)	,041		
B1-B2 words	Pearson Correlation	,426**	,144	1
	Sig. (2-tailed)	,006	,375	

*p-level=0.05 **p-level=0.01

Regarding the 2nd research question, it is obvious that the type/ token ratio does not lead to any conclusive results. It may be due to limitations concerning the word count of the texts that the type/ token ratio, level of scripts and level of vocabulary do not correlate. Mean comparison: According to one way ANOVA with grades and type/token ratio acting as dependent variables and B1-B2 words acting as independent it was found that (table 4) mean grade values is statistically different for B1-B2 words categories, that is 0,1 and 2 ($F=4.733$, $df=2,39$, $p=0.015$). On the contrary, this is not true for type/token ratio where the mean values were not statistically different for B1-B2 words categories ($F=2.879$, $df=2,39$, $p=0.069$). Further investigation with the help of graph 2 showed that moving to larger number of B1-B2 words will result higher grade values but in the case of type/token ratio the one-word category had the highest value (Mean=0.78). Still, in both cases the 0-word category showed the lowest values.

Table 4. One way ANOVA for grades and type/token ratio against B1-B2 words

		Sum of Squares	df	Mean Square	F	Sig.
grades	Between Groups	36,585	2	18,293	4,733	,015
	Within Groups	143,015	37	3,865		
	Total	179,600	39			
type/token ratio	Between Groups	,036	2	,018	2,879	,069
	Within Groups	,233	37	,006		
	Total	,270	39			



Graph 2. Mean values of grades and type/token ratio for B1-B2 categories.

Accordingly, it is concluded that research question 3 is confirmed, namely that the use of higher-level vocabulary (B1- B2 verbs and nouns) by a specific student agrees with the proficiency level of the student and could be a transitional threshold towards higher- level proficiency bands as grades correlate with the existence of this vocabulary.

3.1 Discussion

The first step of the analysis was the selection of the words to be examined from the wordlist of the most frequent words of the corpus. The most frequent word was “moved” with a frequency of 77. In the English Profile search the B1 level of the word was selected (change place: go to a different place to live or work) and not the A2 level (change position or make something change position). However, since it was a word used extensively in the prompt and the guiding notes it was not taken into consideration as B1 vocabulary used by the students. Nevertheless, what is interesting to note is the correct grammatical use of the word. It was used

correctly only in 25 texts. Twenty- three of these were in A2 tests, 2 in A1 borderline A2 tests and only one #20 was in an A1 test, but it included the use of one B1 word. We may assume that in all other cases although the meaning was clear there was not consolidated knowledge of the correct grammatical use of the verb, needing preposition *to* instead of *in*, nor confidence to use the phrases exactly as they appeared in the guiding notes, knowing however that they master the word and would use it anyway. Therefore, although the level A1 meaning was known from every student, and the B1 meaning could be easily assumed in the guiding notes, it was within their breadth of vocabulary knowledge, but it was the A2 students who displayed depth of vocabulary knowledge since they knew its collocational and syntactic features (Shen, 2008). We may also argue that the prompt enhances writing ability, without distorting results, spreading equally the benefits to lower as well as higher level students. According to relevant literature from a prose model prompt, which is a sample of a paragraph of what students are asked to write to guide them (Bahrebar & Mohammadi Darabad, 2013, p.5) “one might assume that these students simply copied the letter or altered it only slightly”. That was not the case as it is not in this study. In both cases, even if the relevant cues of the KPG writing task are shorter than a prose, there were written productions with grammatical or even spelling mistakes, which “could have been avoided by copying the model” (Bahrebar & Mohammadi Darabad, 2013, p.10). The type of guided prompt used in this KPG test activity, deriving probably from the increased needs of young EFL learners falls within the idea that the prose model prompt acting as a stimulus for writing positively affects writing quality (Bahrebar & Mohammadi Darabad, 2013). The number of nouns recorded in this study is roughly three times more than the number of verbs. It has been claimed that the acquisition of nouns precedes the acquisition of verbs in L2 (Dietrich, 1990 as cited in Viberg, 2002). It has also been stated that the lexicalization of nouns is a more straightforward process than that of verbs across languages. Verbal semantic fields in a basic language level consist of one or a few “nuclear” verbs with universal functions to convey the same semantic meanings of motion, perception or possession, whereas nominal semantic fields of the same level consist of a lot more items we perceive them through our senses in our natural world (Viberg, A., 2002 p.41). The number of verb noun combinations is even smaller, confirming the findings in the literature which consider collocation use as a challenge in ESL learners regardless of their L1 backgrounds and their level (Laufer, Waldman, 2011). In the present corpus among the 9 attempts to convey the meaning of “make friends” only twice the student used the right verb “made”. Similarly, learners of English with Swedish L1 and learners with Chinese L1 confused make and do in relevant combinations (Wang and Shaw, 2008 as cited in Laufer, Waldman, 2011). Therefore, misuse of collocations is expected from young students, since it is a problem in advanced students as well. It is an area where learners will become more competent quite later in their journey of acquisition (Laufer, Waldman, 2011).

In order to answer the 4th research question, learners’ control over B1 B2 words was examined and it was concluded that these words were suitably used, with their intended meaning:

- *I want to **go back** to my home.* The phrasal verb is used properly to mean return (B1). It is used once more in the Corpus successfully.
- *Every weekend I **hang out** with my friend and we go to the park.* In two more occurrences the phrasal verb is used identically, accurately and meaning spend a lot of time with someone (EVP online)

- *As you know I moved at Rhodes (Lindos).* This hedge (B1) is appropriately used in the opening of the email.
- *I like *live here but I lost all my friends.*
- *Finally opposite my new house exist a big park and I play with my brother there!*
- *I spend my free time fixing stuff*
- *.....because my father take a good job as reporter.....*
- *We had to move there because my mother got *transferred there.*
- **The summer I welcome why can fantastic game.*
- *I haven't *hear from for donkey's years!*

In these occurrences of B1 verbs and expressions the meaning conveyed is the meaning that these words bear. Except welcome which does not make sense. Exist too is not the optimum choice for what the writer wants to express, and it is a clear L1 influence. The meaning “to be present or real” does not apply to a park as the fact of being real and present cannot be questioned as the notion of an abstract noun would be. Concerning proper grammatical use *exist* does not follow the Subject verb order and the tense is not used accurately in *take* and *haven't hear from*. *Transferred* is misspelled. Concerning the B1- B2 nouns they conveyed the meaning that they were supposed to, based on the English profile findings. The problems lay in grammar and usage.

“It also has many bins in which you can do recycling”. Do is wrongly used here as recycle does not combine with a verb. L1 influence is obvious as the phrase is literally translated from Greek.

*“My family had to move there because of *ikonomic *crises”.* The spelling errors are obvious. However, the words are used correctly. The spelling mistakes are anticipated at this level, particularly in less frequent words, such as these (B2). In the Rater's script guide, it is confirmed that such errors are not penalized as long “as they do not interfere with the intelligibility of the text locally or as a whole” (St. Cert. of Lang. Comp., 2012, p.21). The precondition is that the word should convey the “intended meaning” and be intelligible (St. Cert. of Lang. Comp., 2012, p.21). *“There is a lot of green”.* In this case the meaning according to English Profile which seems to be the intended one is “covered with grass or other plants”. However, it is the adjective form that has this meaning whereas in this text it is used as a noun. The writer probably meant greenery, and as the Greek equivalent of green includes the meaning of “covered with grass” he/ she resorted to green. Hence, the notion of the word is familiar to the student, it is its usage which is not appropriate and would need clarification.

“I go to ballet and jazz lessons”. The meaning is obvious and the word is used properly combined with lessons. The word is B1 in BrE and B2 in AmE (EVP online)

“I go to a field near my house and I'm playing soccer” “....there are a lot of cinemas, football field and swimming pools”. The meaning and usage are correct in these occurrences.

“It is a block of flat near the sea”.* In the three instances of this item (B1) there are only two minor spelling mistakes, the singular in *flat*, which is probably wrongly assumed by the singular of the whole construction and another occurrence misspelled as **blocs*.

*“In Athens it *snows very often and I play snow fights...”.* Fight as a noun is used here to convey people using physical force to hurt others according to the B1 definition In English Profile. Of course, it is snowball and not snow fight, but we focus on fight which is appropriately used.

“....because in the sea have a lot of crocodiles and *shrarks”.* Crocodiles (B2) is accurately spelled here but not sharks (B1). However, their meaning is the

obvious. The “have construction” is not the best choice here but it does not affect the use of the words in question. “*Anywhere in the **future** I find new friend and I play with ***their**”.* Meaning (B1) and usage are accurate here. “*It has a big park where the children can play in the **ground**”.* According to English Profile in BrE there is the notion of “an area of land used for a particular purpose or activity” (B1), with which the word is used here. “*To tell the truth this place is fantastic,...*” The meaning of the phrase (B1) is in agreement with the usage in this phrase as the writer said at the beginning: “*I didn’t want to move there*”. So the “real facts about a situation”(EVP online) are different than the first thoughts. “*My ***famile** have moved here because my father ***take** a good job as **reporter** in famous magazine.* The word reporter is appropriately used in accordance with its meaning (B1) “*We had to move because there was too much **Racism**. I’m hoping things will be different here*”. Although there is not much context, the reader deduces that the word is used with the intended meaning (B2). The sentence that follows and the Greek equivalent of the word which is very similar attest to the same conclusion. “*We go every week in the **mall**”.* Mall (B1) signifies a distinct place which is difficult to be confused. Its usage demands a motion verb as there is. The preposition is not the one accompanying motion verbs. In this case as throughout this corpus students’ overuse the preposition in, often inaccurately. “*My parents and I do ***housesword**”.* The assumption is that the word is housework (B1). It is combined with the right verb. The meaning of the whole phrase could be related to an email about moving. The spelling however, is borderline unacceptable.

- “*I’ll wait forward for you ‘re letter! **Hughs* and kisses**, Alex”.*
- *If you want come to visit me in Xanthi. **Greetings**, Alex*
- **Don’t forget to write to me and tell me your opinion.** The expressions (B1) are correctly used as an email closing to a friend. The word opinion bears its intended meaning (B1) and it is used accurately. The words work and job are used extensively in the corpus and inaccurately most of the times. Students are using them interchangeably. It is a problem however which has been recorded before in learner corpora. In the study of the semantic field of work (Martelli, 2004 as cited in Torsello, Ackerley & Castello, 2008) Italian learners make similar mistakes when using not only work and job but also career, employment, occupation. The collocations produced were often inaccurate, a fact attributed to their belief that they can be used interchangeably (Martelli, 2004 as cited in Torsello, Ackerley & Castello, 2008).

3.3 Limitations

The written scripts were written as part of a whole A level test, of which the modules of Reading and Listening Comprehension had to be examined in part during the same class hour with the Writing module. This must have affected the students’ time management, who may have felt more confident, or more tired over their total answers and did not assign the optimal time to provide a longer text in the examined task, which was one of the four activities of the Writing Module. The fact that the learners lived in small cities, away from large urban centers is another limitation as a more representative sample would include learners from different socioeconomic backgrounds.

4. Conclusion

The text and interlanguage analysis of the sixty-nine writing pieces produced by 12 and 13-year-old Greek ESL learners revealed that the vocabulary used was of A1- A2- and B1 level. Noun and content verbs and their combinations were explored assisted by the English Profile tool. A number of B1 and B2 types were detected occurring predominantly in the A2 level band. Some of the scripts could be considered as emerging B1 level, based on the B1 – B2 types included, and their rating according to specific CEFR criteria. Appropriateness of lexis in the 69 scripts was confirmed as the content words were used according to their intended meaning.

References

- Bahrebar, S., & Mohammadi Darabad, A. (2013). Writing in the foreign language classroom: The effects of prompts on the Iranian learners of English. *International Journal of English Language Education*, 1(3). <https://doi.org/10.5296/ijelev1i3.3492>
- Cambridge University Press. (2015). *English Profile: The CEFR for English*. <https://www.englishprofile.org/wordlists/evp>
- Council of Europe. (2001). *Common European framework of reference for languages: Learning, teaching, assessment*. Cambridge, U.K: Press Syndicate of the University of Cambridge
- Council of Europe. (2020). *Common European Framework of Reference for Languages: Learning, Teaching, Assessment. Companion Volume*.
- Dendrinos, B.& Karavas, K. (Eds). (2013). *The Kpg Handbook: Performance Descriptors and Specifications*. RCeL, National and Kapodistrian University of Athens
- Granger, S. and Wynne, M. (1999). Optimising Measures of Lexical Variation In EFL Learner Corpora In Kirk, J. (ed.) *Corpora Galore*. Amsterdam and Atlanta: Rodopi, 249-257.
- Laufer, B. and Waldman, T. (2011), Verb-Noun Collocations in Second Language Writing: A Corpus Analysis of Learners' English. *Language Learning*, 61: 647-672. <https://doi.org/10.1111/j.1467-9922.2010.00621.x>
- Leńko-Szymańska, Agnieszka. (2015) The English Vocabulary Profile as a benchmark for assigning levels to learner corpus data in Callies, M.& Goetz, S. (Eds), *Learner Corpora in Language Testing and Assessment*, 115-140, John Benjamins DOI:10.1075/scl.70.05len
- Lessard-Clouston, Michael. (2013). *Teaching Vocabulary*. TESOL International Organization.
- Milton, J. (2010). The development of vocabulary breadth across the CEFR levels. In Bartning, I., Martin, M. & Vedder, I. (Eds.), *Communicative Proficiency and Linguistic Development: intersections between SLA and language testing research* (pp.211- 232). European Second Language Association.
- Milton, J. (2013). Measuring the contribution of vocabulary knowledge to proficiency in the four skills. In Bardel, C., Lindquist C., Laufer, B. (Eds.), *L2 Vocabulary Acquisition, Knowledge and Use new perspectives on assessment and corpus analysis*. (pp.57-78). Second Language Association.
- Ministry of Education: English Language Certification. (2012). *LEVEL A1 & A2 on the scale set by the Council of Europe: MODULE 2 WRITING*.
- Owen, N., Shrestha, P., & Bax, S. (2021). Researching Lexical Thresholds and Lexical Profiles Across the Common European Framework of Reference For Languages (CEFR) Levels Assessed in the Aptis Test. ARAGS Research Reports Online, 1.
- Shen, Zhifa. (2008). *The Roles of Depth and Breadth of Vocabulary Knowledge in EFL*

