FORMULAIC SEQUENCES IN THE WRITTEN PRODUCTION OF L2 LEARNERS OF ENGLISH

Mar Gutiérrez-Colón
Universitat Rovira i Virgili

Elisabet Pladevall
Universitat Autònoma de Barcelona

ABSTRACT: This article investigates the use and function of written formulaic sequences and its relation to second language proficiency level (English). This is a cross-sectional study in which data are gathered from 138 students from 5th and 6th grade (from 10 to 12 years old) from 8 different schools in Catalonia. The paper analyzes: 1) the relationship between number of formulaic sequences and proficiency level of the students and 2) distribution of formula types depending on the students' level. For the purpose of this analysis of formulaic sequences, an original classification has been created. The data show that there is a higher use of formulaic sequences in lower levels of proficiency and that each level presents a different distribution of formula types.

KEYWORDS: Second Language Acquisition, English as a Foreign Language, formulaic sequences, classrooms setting.

FÓRMULAS LINGÜÍSTICAS EN LA PRODUCCIÓN ESCRITA DE APRENDICES DE INGLÉS (L2)

RESUMEN: En este artículo se investiga el uso y la función de secuencias de fórmulas escritas y su relación con la competencia lingüística (Inglés como segunda lengua). Es un estudio transversal en el que se recogen los datos de 138 alumnos de 5 º y 6 º cursos de educación primaria (de 10 a 12 años), de 8 escuelas de Cataluña. El documento analiza: 1) la relación entre el número de secuencias de fórmulas y nivel de rendimiento de los estudiantes y 2) la distribución de los tipos de fórmula en función del nivel de los estudiantes. Para poder analizar las fórmulas, se ha creado una clasificación original. Los datos muestran que hay un mayor uso de fórmulas en los niveles inferiores de lengua y que cada nivel presenta una distribución diferente de los tipos de fórmulas.

PALABRAS CLAVE: Adquisición de segundas lenguas, inglés como lengua extranjera, fórmulas lingüísticas, aprendizaje de lenguas en el aula.

1. Introduction

Most studies on formulaic sequences (Krashen & Scarcella, 1978; May, 1982; Wray, 2000; Schmitt & Carter, 2004; Bishop, 2004; Wood, 2006) coincide with the idea that both written and spoken discourse contain a large percentage of formulaic language. However, formulaic sequences are also commonly used by proficient L2 learners as well as native speakers. The role of formulaic sequences in second language learning has been mainly studied with regard to spontaneous spoken language and to our knowledge, studies on the use and form of formulaic sequences in the written production of students are scarce. According to Weinert (1995: 180), the most recurrent themes in the study of formulaic language are: a) the evidence of L2 learners' use of formulaic language and its function, b) to what extent this use is target language (TL) or interlanguage (IL), c) the role of formulaic language in classroom second language development d) how TL idiomaticity develops. These strings of words have been learnt by the speaker, and most of the times, when talking about second language learning, researchers on the field conclude that the level of syntax of these formulaic chunks is higher than the general language proficiency of the learner and hence they are usually associated with L2 learners with low proficiency levels (this phenomenon has been studied in depth by Weinert, 1995). According to Weinert (1995: 186-187) there is ample evidence of both creative construction processes of language in children when acquiring the L2, and children's use of formulaic language. Krashen suggests that the use of formulas in children is not a central process in a naturalistic acquisition (Krashen and Scarcella, 1978), while Wong-Fillmore (cited in Weinert 1995: 188) believes that probably the "linguistic environment of the classroom and the playground may have encouraged the use of formulas by requiring early production". Research on this topic shows that most studies of formulaic language have dealt with children, and that most studies which were conducted on adults involve untutored learners. These studies on adult learners conclude that children appear to use more formulas than adults (Weinert, 1995: 189).

In this article, we will focus on the use and function of written formulaic language and its relation to L2 proficiency level. As Schmitt and Carter (2004) put it, "given the evident importance of formulaic sequences in language use, convincing explanations of the mechanics of their acquisition must become an essential feature of any model of language acquisition" (2004: 12). And for this reason we would like to contribute to this research topic with a new point of view and analysis of formulaic sequences: written formulas of young second language learners of English in a classroom setting.

The article is organised as follows: section 2 presents definitions and uses of formulaic sequences and a brief overview of formulaic language in SLA. Section 3 presents the hypotheses and research questions which motivated the study together with its methodological aspects, namely criteria for identification of formulas, participants in the study and classification of essays and formulas. Section 4 presents the results obtained and section 5 analyses and discusses them. Finally, section 6 concludes the study and points out aspects of future research.

2. Formulaic sequences

Following Weinert criteria for labelling formulaic language (1995: 182), we will use the terms formulas, prefabricated or ready-made language, chunks, etc. interchangeably. There does not seem to be agreement on the definition of what constitutes a formulaic sequence and what specific characteristics these sequences should have which make them distinct. Wood (2002) states that "the consensus seems to be that formulaic language sequences are multi-word units of language which are stored in long-term memory as if they were single lexical units" (p. 31), or as Bishop (2004) puts forward, formulaic sequences "are not generated grammatically, but they are stored and deployed as holistic units" (p. 15). However, that a formulaic sequence is a single lexical unit remains unclear. Hakuta (1976) and Krashen and Scarcella (1978) distinguish between *routines* and *patterns*. They define *routines* as whole utterances learnt as memorized chunks (e.g. *I don't know)*, and *patterns* as utterances which have one or more open slots (eg. *can I have a ...*). Ellis (1983) also suggests that formulaic speech can consist of entire scripts, such as greeting sequences, which the learner can memorize.

In another study, Ellis (1994: 86) also states that what enables researchers to distinguish formulaic speech from creative speech in a learner's language corpus is the well-formedness of formulas as they do not involve creativity and hence the number of mistakes decreases. Ellis (1994) also mentions that each formula is related to the particular language function that the learner wants to communicate: "Learners, like native speakers, learn formulas because it reduces the learning burden while maximizing communicative ability" (1994: 86). Other studies also relate the use of formulas to communicative functions. According to Adolphs and Durow (2004) formulaic sequences in L2 learning are relied on as a quick means to be communicative and this can lead to quicker integration into a peer group.

For Coulmas (1981), formulaic speech is "essential in the handling of day-to-day situations" (p. 4) because it helps covering these parts of speech for which our

proficiency level is still not ready, which suggests that the more proficient a student is the fewer formulas he/she will produce. Myles (2004) states that there is a large body of evidence in child L2 both instructed and naturalistic acquisition that shows that chunks are prominent in the early stages because L2 learners use shortcuts in order to bypass the lengthy processes of the acquisition of morphosyntax and processing skills (2004: 155), and this is precisely what will become evident in the present study. Yet and according to the author, the correct productions of beginners do not necessarily mean that they have acquired the syntactic representations of the target language. The clearest example to her theory is the case of unanalysed chunks, which make students go beyond their grammatical competence (2004: 140).

Evidence in second language learning research has shown that formulaic speech is used as a learning strategy in children and beginner second language learners, that is, the use of unanalyzed chunks is a common phenomenon in early stages of first and second language learning (Ellis 1994; Skehan 1998). Ellis (1994) suggests that the use of formulaic speech is one of the earliest developmental stages of second language acquisition, before what he called the *semantic and structural simplification stage* and after *the silent period stage*. In his case study, Hakuta (1976: 332) concludes that at least in the initial stages of second language learning, prefabricated patterns constituted over a 50% of the subject's utterances. Ellis, (1996) states that second language learners acquire the grammar rules of the language through implicit analysis of the chunks they have memorized. Specifically, students learn the sequential position of words once they have memorized and used these chunks. Vihman (1982) even believes that "the use of formulas by children learning a second language is not so remote from ordinary adult language use" (p. 280). Wood (2002) believes that when acquiring a second language, both children and adults attend to formulaic sequences in language input, adapt them for their own use and later on they segment and analyze them.

According to Ellis (2005) the importance that formulaic expressions have nowadays in the field of second language learning is accepted within the researchers' community, since these expressions seem to be the basis for the future development of a rule-based competence. He believes that learning chunks at the very beginning of the second language acquisition process is more important than learning grammar itself, because in this way students are provided with opportunities to perform pragmatic meaning. For him, the learnt formulas the students use enable them to express functions of language which are beyond their language knowledge and thus maximize their communicative ability. Nevertheless, we should also say that Ellis strongly believes that both the acquisition of formulaic

expressions and a rule-based knowledge is the basis of a complete language curriculum. This study contributes to the definition and identification of formulaic speech within the students' interlanguage.

3. The present study

3.1. *Criteria for identification of formulaic sequences*

Although some recent research has focused on the presence of formulaic sequences in written texts (Bardovi-Harlig, 2002; Tode, 2003), most previous existing literature in the field focuses on formulaic speech in oral texts, which makes the present analysis of written production more relevant. As mentioned above, the identification of formulaic expressions is far from simple given the various types, uses and definitions of formulas.

In her article, Myles (2004) suggests that in the early stages of the learning of a language, there is no semantic structure in the student's productions, but semantic mappings. For her when the students start learning a second language, their first stage of development is to associate semantic content with words or formulaic sequences. She argues that nouns are easier to acquire than verbs and thus there is a hierarchy in the acquisition of primary syntactic categories. Verbs start structuring the sentences only in a second phase of development. Chunks of language are therefore "learnt before the constituents from which they are made" (p. 163).

It seems that the formulaic sequences are at the interface between linguistics and pragmatics (Girard and Sionis, 2003), since their origins seem to be, among others, "the learner's communicative need to perform the language functions encoded in FS [formulaic sequences] with ease and fluency" (2003: 244). In their article, Girard and Sionis divide the structuring components of FS into three groups: syntax, phonology and morphology.

According to Myles et al., identifying chunks is not an easy question. There are many different criteria used both by first and second language researchers. She proposes a criteria for L2 chunk identification:

1. Greater length and complexity of sequence compared with other learner output.
2. Phonological coherence, that is, fluent nonhesitant encoding without a break in the intonation contour.

3. They tend to be used inappropriately (syntactically, semantically, pragmatically) as numerous examples of overextensions clearly show.
4. They are generally used in the same form, with no parts substitutable, that is, learners are not able to change any elements in the sequence.
5. They tend to appear well formed and to be grammatically advanced compared to the rest of the learners' language.
6. They usually occur in situationally specific ways or are predictable in context; the classroom situation is particularly rich in routines that are heavily context dependant. (Myles 1999: 51)

For the purposes of this research, we have used and adapted this criteria. Finally, three general characteristics have been applied in deciding whether a sequence was a formula:

1. Greater length and complexity of sequence compared with other learner output.
2. Many times, the sequences are used inappropriately (syntactically, semantically, pragmatically), although they tend to appear well formed and to be grammatically advanced compared to the rest of the learner's language.
3. They usually occur in situationally specific ways or are predictable in context since the classroom situation is particularly rich in routines that are heavily context dependent.

In the essays to be analysed, formulas also appear to be strings of words used as fillers, that is, to fill in the whole page with a long composition without adding much content or logical coherence to the text. They also tend to be extremely repetitive and they include both *routines*, learnt as memorised chunks, and *patterns* (see section 2), thus allowing for some creativity by inserting different lexical items in the formulaic structure (e.g. my favourite animal is a cat/a dog/an elephant, etc.).

3.2. *Aims and hypotheses*

This study aims to explore the formulaic sequences in the written production of primary school students in Catalonia. The research questions that guide this study are the following:

– Is the presence of formulaic expressions related to the second language proficiency level of the students?

– How can these formulaic expressions be classified in order to further describe them and relate them to the level of the students who produced them?

The hypotheses of research that we propose for this study are:

- Hypothesis 1: We expect to find the highest number of formulaic sequences in the texts of the lowest levels of proficiency of the students.
- Hypothesis 2: Each proficiency level presents a different distribution of formula types. This might give us an indication of the relationship between types of formulas and L2 proficiency level and hence relate formula types to language development.

3.3. *Participants and classification of essays and formulas*

The compositions that have been analysed in this study belong to 138 students from 8 different schools in Catalonia. The schools were chosen at random and students from either 5th or 6th grade in each school were asked to write an essay. Which class participated in the study was decided by the school itself and all students in the selected classes participated in the task. The topic of the compositions was the same for all schools: *Welcome back to school!*

The average of English language hours per week is three, and although the method used in class is clearly communicative, teachers tend to follow a book and do grammar-based exercises in class. The languages used in class are both Catalan and English. The final exams are set up in a written form, and they are generally based on some grammar/vocabulary questions plus a composition. Some teachers also include a listening comprehension exercise. Nevertheless, the final marks are based on a continuous assessment method, for which each school has its own percentages and rules.

The essays were written by 138 students who were between 10 and 12 years old. Although these students belong to two different academic courses (5th and 6th year of Primary School) their foreign language proficiency level is in general quite similar, and for this reason we decided to treat them as a single group. A significant heterogeneity was expected and found, since the students belonged to eight different schools with different kinds of input. The classification of essays in different proficiency levels was done according to Wolfe-Quintero, Inagaki and Kim's taxonomy (1998). They considered that the syntactic complexity, the fluency

and the accuracy of the student's writings could group and define the measures of analysis used in the literature to describe L2 writing development. More specifically, we adopted some of the measures of L2 writing analysis and modified some others from Celaya Villanueva, Pérez-Vidal and Torras Cherta (2000-2001). As a result of this classification, we found three different groups according to their level (L1, L2, L3).

3.4. *Specific measures of analysis*

The basic measures of analysis were words and units. The term *unit* included simple sentences and coordinated and subordinate clauses. The term *unit* used here is different from the term *T-unit*, defined as "an independent clause and all its dependent clauses" (Hunt, 1965 cited in Polio, 1997) and extensively used in the L2 writing literature. Two coordinated sentences are counted as two units in the present study and a main clause with a subordinate clause are also counted as two units here, whereas they would be counted as one *T-unit*. The term T-unit was not chosen for the present analysis due to the low lexical and syntactic complexity of the essays analysed.

As for syntactic complexity, coordinated and subordinated units were counted and regarding fluency the total number of units/essay, words/essay and words/unit were calculated for each essay. L1 words in Spanish or Catalan or proper nouns were not considered. With respect to accuracy, *disregarded units* and error-free units were analysed. The term *disregarded unit* was used to refer to those units that were unintelligible for the reader (e.g. sentences entirely written in the learners' L1, totally impossible word order, invented words or expressions, etc) . These units were not considered and hence not counted in the measures of complexity, fluency or correct units but provided significant information about the level of the essay. For instance:

(1) My doing is *carta* (= *letter*) (Essay 177 – Level 1)

(2) In the profetion (= *the teachers*) is Mar, Antonia, Elsa the profetion is *nostre* (= *our*) class *angles (= English*) (Essay 148 – Level 1)

However, those units that included errors but were intelligible for the reader were indeed counted and analysed. For example:

(3) School is colour orange (Essay 166 – Level 2)

(4) I no sister (Essay 182 – Level 2)

The number and percentage of error-free units provided complementary but not determinant information in the classification of the essays into the three levels. Since some of the essays contained a very low number of units, which were not linguistically complex and hence correct, a high percentage of error-free units cannot be taken into account on its own, as it would be misleading and must be taken into consideration together with other measures of analysis.

In order to classify the essays the most important factor was on the one hand the presence/absence of disregarded units and of coordinated and subordinated units. Average results on fluency will show an increasing tendency in number of units/essay, number of words/essay and number of words/unit across the three level groups, although the increase in the number of words/unit is far less remarkable.

Level 1 (L1) contains 60 essays and is mainly characterised by the presence of disregarded units and the generalised absence of coordinated and subordinate units. Those essays which contained a very low number of coordinated or subordinated units together with disregarded units are also classified under this group. These essays contained a low number of units and a low percentage of accurate units. As for average results on fluency, L1 shows an average number of units/essay of 6.83, an average number of words/essay of 33 and an average number of words/unit of 4.81.

Level 2 (L2) contains 69 essays and is mainly characterised by the absence of disregarded units and the generalised absence of coordinated and subordinate units. Those essays with either presence of coordinated units and absence of subordinate units or absence of coordinated units and low presence of subordinate units are also classified under this group. As for average results on fluency, L2 shows a considerable increase with an average number of units/essay of 10.13, an average number of words/essay of 55.75 and an average number of words/unit of 5.42.

Level 3 (L3) contains 9 essays and is mainly characterised by the absence of disregarded units and the full presence of both coordinated and subordinate units in every essay. As for average results on fluency, L3 shows an average number of units/essay of 14.11, an average number of words/essay of 81.66 and an average number of words/unit of 5.76.

In order to test the hypothesis that each proficiency level presents a different distribution of formula types, they were classified according to whether they were lexical, syntactic or morphological. Lexical formulas in the essays to be analysed

refer to those vocabulary items or expressions which are typically presented, repeated and memorised in language classrooms, as for example "*My name is Maria"*, "*I'm from Spain"* or "*My favourite colour is blue"*. Syntactic formulas refer to those syntactic constructions, normally verb phrases, which are presented and practised in classroom exercises without grammatical explanations, as for example *"Have you got…?", "I've got…", "I like…" or "I don't like…"*. Finally, morphological formulas found in the essays refer to the repeated formulaic and ungrammatical use of verbal morphemes, such as "I'm" or "–ing", as for example *"I'm play the piano" or "I going to school"*.

4. Results

This section presents the percentages of formulaic sequences in each Level group on the basis of the number of units in each essay and in relation to Hypothesis 1 and the percentages of formula types, namely lexical, syntactic and morphological formulaic expressions, in each Level group and in relation to Hypothesis 2 (see section 3.2.). Non-parametric statistical tests were applied to compare the percentages between levels and thus analyse the presence of formulas in developmental terms.

The total number of formulaic sequences found in the 138 essays analysed is 441, 166 of whom correspond to Level 1 and 261 and only 14 correspond to Levels 2 and 3, respectively. Table 1 shows the percentages of formulas that appear in the essays according to their proficiency level and on the basis of the number of units in each essay, and Table 2 displays the descriptive statistics of these percentages:

L2 Proficiency level	Number of units	Number of formulaic sequences	% formulaic sequences
Level 1	422	166	36.51%
Level 2	699	261	38.69%
Level 3	127	14	13.46%

Table 1. *Formulaic expressions in each proficiency level*.

Descriptive Statistics % Formulaic Expressions			
Level		Statistic	Standard Error
1	Mean	36,5163	3,79774
	Standard Deviation	29,41719	
	Minimum	,00	
	Maximum	100,00	
2	Mean	38,6910	3,37621
	Standard Deviation	28,04490	
	Minimum	,00	
	Maximum	100,00	
3	Mean	13,4644	6,49159
	Standard Deviation	19,47478	
	Minimum	,00	
	Maximum	60,00	

Table 2. *Descriptive statistics of the percentages of formulaic expressions*.

Level 1 and Level 2 essays display very similar percentages of formulaic expressions, although as was seen in section 3.2, the two levels do present very different features, which make them be divided into two distinct groups. In fact, Level 2 presents a slightly higher percentage value than Level 1, namely 38.69% and 36.51%, which will be justified and explained in the discussion section below. The percentage of formulas decreases sharply in Level 3, down to 13.46%.

In order to compare formula percentages more accurately and since the data were not normally distributed, the non-parametric statistical U Mann-Whitney test was applied. The results are illustrated in Table 3 below:

	% Formulaic sequences	Mann-Whitney U	Asym. Significance p-value
Level 1	36.51%	1980.000	.669
Level 2	38.69%		
Level 3	13.46%	148.000	.011

Table 3. *Mann-Whitney U test on the percentages of formulaic expressions*.

As expected from the preliminary results, a non-significant difference (U 1980.000; p=.669) results from the comparison between Level 1 and 2, whereas the percentage of formulaic sequences in Level 3 is significantly lower than in Level 2 (U 148.000; p=.011), indicating that the development of formulaic sequences is significant from Level 2 to Level 3 and not between Level 1 and 2. Therefore, the number of formulas decreases as proficiency level increases but only from Level 2 to Level 3, which will be explained in the discussion section.

In relation to Hypothesis 2 and to test whether each Level presents a different distribution of formula types, their percentages were also calculated for each Level. Results are displayed in Table 4:

	Morphological formulas	Syntactic formulas	Lexical formulas
Level 1	31 18.67%	45 27.10%	90 54.21%
Level 2	19 7.27%	104 39.84%	138 52.87%
Level 3	0 0%	8 57.14%	6 42.85%

Table 4. *Percentages of formula types*.

As Table 4 illustrates, lexical formulas represent more than half of the total number of formulas in Level 1 and 2 (i.e. 54.21% and 52.87%, respectively) and almost half of it in Level 3 (i.e. 42.85%). Syntactic formulas represent 27.10% of formulaic expressions in Level 1, 39.84% in Level 2 and are the predominant type of formulas in Level 3 (i.e. 57.14%). Percentages are much lower regarding morphological formulas, these being 18.67% in Level 1, only 7.27% in Level 2 and 0% in Level 3, thus representing the least used type. Results will be explained and discussed in the next section (See the list of formulaic expressions in the students' essays in Appendix A).

5. Discussion

Regarding the percentage of formulas that the students use in each level, the analysis shows that the decrease of formulaic sequences is not significant from

level 1 to level 2, but it is from level 2 to 3. Therefore, it seems that our first hypothesis is not confirmed since the decrease in the number of formulas appears to be outstanding only from level 2 to level 3.

However, when analysing the linguistic characteristics of each level in detail, it can be seen that the results of this analysis show that both level 1 and level 2 coincide in the fact that there is a generalised absence of coordinated and subordinated units. This fact implies that the grammatical complexity of the texts in these two levels is by far more simple than in level 3. Since the number of formulas is also higher in these two levels, this last characteristic confirms the fact that both level 1 and level 2 have a lower level of proficiency than level 3; or to say it in other words, the absence of coordinated and subordinated units, plus a similar and percentage of formulas indicate that the proficiency level of level 1 and level 2 is similar and much lower than in level 3. The fact that there was not a big increase in number of formulas from level 1 and level 2 could then be explained by the fact that both levels of proficiency are similar, at least in text complexity (coordinated and subordinated units). Therefore and confirming Fillmore (in Krashen and Scarcella, 1978) and Myles (2004) and our first hypothesis, the presence of formulaic sequences has a decreasing tendency as proficiency level improves.

Our second hypothesis stated that each proficiency level would have a different distribution of formulaic sequences, and this difference could be an indication of the relationship between types of formulas and second language proficiency level. Three main interesting findings have been observed according to the possible relation formula types-proficiency level.

1. Lexical formulas are the most predominant type in all three levels. They represent more than half the total number in level 1 (54.21%) and level 2 (52.87%), and almost half in level 3 (42.85%).
2. Syntactic formulas are less common in level 1 (27.10%) than in level 2 (39.84%) or level 3 (57.14%) and hence increase regularly according to the level.
3. On the contrary, morphological formulas decrease regularly as the level of proficiency increases. A percentage of 18.6% of them is found in level 1 but only less than half (7.27%) in level 2 and none (0%) in level 3.

We have defined lexical formulas (see section 3.3.) as those vocabulary items which are typically presented and repeated in language classrooms; syntactic formulas refer to those syntactic constructions (usually verb phrases) which are

presented and practised in the classroom exercises without grammatical explanations, and finally morphological formulas refer to the repeated formulaic and ungrammatical use of verbal morphemes.

The results of this research show that the syntactic formulas increase together with the level, and this is probably due to the unconscious need of the student to acquire a syntactic scaffold for his/her second language (Ellis, 2005). We should bear in mind that most theories (Krashen & Scarcella, 1978; May, 1982; Wray, 2000; Schmitt & Carter, 2004; Bishop, 2004; Wood, 2006) state that second language learners learn the sequential position of words by memorising chunks before they can use creative language .

Lexical formulas slightly decrease as proficiency level increases, although as we can see, they are the most predominant type in all three levels. The reason for this is that these type of formulas are the ones which carry more pragmatic load and therefore, and according to Ellis (2005) (section 2) these are the ones that are the most needed by the students for a day-to-day use of the second language, especially when their proficiency level is not very high.

Morphological formulas follow a logical inverted progression according to the increase of proficiency level. We understand that these results can be seen in the light of the normal process that any student undergoes when acquiring a second language: grammatical mistakes tend to be corrected and finally disappear as proficiency level increases. This is particularly clear at lower levels of proficiency. In this case, the morphological formulas which have been classified in this study were basically ungrammatical uses of verbal morphemes which showed texts with a very poor mastery of the language. These types of formulas were the ones that most showed a lack of knowledge of the second language rules. Next section draws the analysis to a conclusion and points out the main findings of the present research.

6. Conclusions and further research

The present study was conducted on 138 essays written by primary school students. The essays were divided into 3 levels according to their proficiency, and although level 3 is clearly higher than level 1, we should not forget that all of them belong to primary school students with an average of 3 hours of tuition per week, in a class of about 30 students and thus their general English level is weak.

We would like to conclude by saying that the final analysis of the results obtained, indicates that both our hypotheses have been confirmed:

1. Hypothesis 1: There is a higher number of formulaic sequences in the texts of the lower levels of proficiency of the students. The difference is higher from level 2 to 3 than from level 1 to 2. This is due to the lack of coordinated and subordinated units in the two first levels, which sets them in a similar level of texts' complexity and also in a significant difference with the level 3's proficiency.
2. Hypothesis 2: Each proficiency level presents a different distribution of formula types: morphological formulas regularly decrease as proficiency level increases, syntactic formulas regularly increase according to the level and lexical formulas present a very similar distribution in all three levels. This is explained by the following observations:

   3.32. Lexical formulas are the ones which carry a bigger load of pragmatic meaning and surely belong to the core language used at school. They are used in the three levels and they represent approximately half of the amount of formulas in all three levels.

   3.33. Syntactic formulas increase as the student increases his/her proficiency level, which means a higher need for well-formed syntactic structures to be able to express more meanings, but the student is still in the need of this scaffold, since he/she is not ready to create his/her own structures yet.

   3.34. Morphological formulas decrease because students improve their proficiency level and therefore start correcting the ungrammatical use of verbal morphemes, which represents a very low grammatical level.

After analysing the data gathered for this paper, we strongly believe that further research should be conducted on the same topic but on higher levels of proficiency and then both results should be compared. It is essential to know whether the conclusions we have reached are only applicable to lower levels of proficiency or can be generalised to all levels. New research is thus being designed, with new compositions from secondary school students.

Acknowledgements

We would like to thank AGAUR (Agència de Gestió d'Ajuts Universitaris i de Recerca), Generalitat de Catalunya, who funded the research project on which this study is based (grant nº. 2006ARIE10065), as well as the other team members in the project *Avaluació de l'Expressió escrita en llengua Anglesa d'alumnes*

participants en projectes Comenius: Ana Mª Coll, Clara Curto, Xavier Chavarria, Stephen Hampshire, Delio del Rincon, M. Mar Rodríguez. We would also like to thank the teachers and children in the schools that took part in the project. We would also like to thank the schools that have accepted to become part of this project: CEIP Mare de Deu del Portal, Batea; CEIP La Roureda, Sabadell; CEIP Santa Anna, Premià de Dalt; CEIP Els Pins, Cornellà de Llobregat; CEIP Riera de Ribes, Sant Pere de Ribes; CEIP Pràctiques, Tarragona; CEIP El Carmel, Barcelona; CEIP Costa i Llobera, Barcelona, spcecially the teachers who guided us through their centers and allowed us to be in their classes.

References

ADOLPHS, S. and DUROW, V. (2004). "Socio-cultural integration and the development of formulaic sequences", in Schmitt, N. (ed). *Formulaic Sequences*. Amsterdam: John Benjamins.

BARDOVI-HARLIG, K. (2002). "A New Starting Point? Investigating Formulaic Use and Input". *Studies in Second Language Acquisition* 24: 189-198.

BISHOP, H. (2004). "Noticing Formulaic Sequences - A Problem of measuring the Subjective". *LSO Working Papers in Linguistics* 4: 15-19.

CELAYA, M. L. et al. (2001). "Matriz de criterios de medición para la determinación del perfil de competencia lingüística escrita en inglés (LE)". *RESLA* 14: 87-98.

COULMAS, F. (ed.). (1981). *Conversational Routines: Explorations in Standarized Communication Situations and Prepatterned Speech*. The Hague: Mouton.

ELLIS, N. C. (1996). "Sequencing in SLA: Phonological memory, chunking and points of order". *Studies in Second Language Acquisition* 18: 91-126

ELLIS, R. (1983). "Formulaic speech in early classroom second language development", in Handscombe, J. *et al*. (eds.) (1983). *On TESOL'83. The Question of Control. Selected papers from the Annual Convention of Teachers of English to Speakers of other languages*. Washington DC: TESOL.

ELLIS, R. (1994). *The Study of Second Language Acquisition*. Oxford: Oxford University Press.

ELLIS, R. (2005). *Analyzing Learner Language*. Oxford: Oxford University Press.

FILLMORE, L. (1978). "Cognitive and social strategies in language acquisition". Ph.D. Dissertation, Stanford Univeristy, in Krashen and Scarcella. 'On routines and patterns in second language acquisition and performance'. *Language Learning* 28: 283-300.

GIRARD, M. and SIONIS, C. (2003). "Formulaic Speech in the L2 classroom: an attempt at identification and classification". *Pragmatics* 13 (2): 231-251.

HAKUTA, K. (1976). "A case study of a Japanese child learning English as a second language". *Language Learning* 26: 321-51

KRASHEN, S. and SCARCELLA, R. (1978). "On routines and patterns in second language acquisition and performance". *Language Learning* 28: 283-300

MAY, M. (1982). *Exceptional Language and Linguistics*. New York: Academic Press.

MYLES, F. (2004). "From data to theory: the over-representation of linguistic knowledge in SLA". *Transactions of the Philological Society* 102 (2): 139-168.

MYLES, F. et al. (1999). "Interrogative chunks in French L2. A basis for creative construction?". *SSLA,* 21.

POLIO, C. (1997). "Measures of Linguistic. Accuracy in Second Language Writing Research". *Language Learning,* 47. 101-143.

SKEHAN, P. (1998). *A Cognitive Approach to Language Learning*. Oxford: Oxford University Press.

SCHMITT, N. and CARTER, R. (2004). "Formulaic sequences in action. An introduction", in Schmitt, N. *Formulaic Sequences*. Amsterdam: John Benjamins.

TODE, T. (2003). "Form unanalysed chunks to rules: The learning of the English copula 'be' by beginning Japanese learners of English". *IRAL* 41: 23-53.

VIHAM, M.M. (1982). "Formulas in First and Second Language Acquisition", in Menn, L. and Obler. L. (eds) *Exceptional Language and Linguisics*. New York: New York Academic Press.

WEINERT, R. (1995). "The Role of formulaic language in second language acquisition: a review". *Applied Linguistics* 16.

WOOD, D. (2002). "Formulaic language in Thought and Word: Vygoskian Perspectives". *Cahiers linguisticques d'Ottawa* 30: 29-48.

—, (2006). "Uses and Functions of Formulaic Sequences in Second Language Speech: An Exploration of the Foundations of Fluency". *The Canadian Modern Language Review* 63 (1): 13-33.

WOLFE-QUINTERO, K. et al. (1998). *Second Language Development in Writing: Measures of fluency, accuracy and complexity*. Honolulu: University of Hawaii, Second Language Teaching and Curriculum Center. (Technical Report, 17).

WRAY, A. (2000). "Formulaic Sequences in Second Language Teaching: Principle and Practice". *Applied Linguistics* 21 (4): 463-489.

—, (2002). *Formulaic language and the Lexicon*. Cambridge: Cambridge University Press.

Appendix

Lexical formulas in the students' essays

Essay	Lexical Formulas
Level 1	
4	**My name is M.**
	I do favourite teacher is M.
6	**(assignatura) favourite is art**
	my favourite friend is G.
64	**MY birthday is on 6th October**
	My favorite friends and scool H., P., D.,
	My favorite friends and playing football is…
66	**My favourite festivity to Roureda is carnival**
	My favourite subject is P.I.
	My favourite teacher is E., O.,
	My favourite color is blue
	My favourite month is January
	My favourite sport is basketball
71	**My faborite equip of futball teen is FC Barcelona**
	My faborite player is Leo Messi
	That is a my faborite school is La Roureda
73	**It's favourite prof is G and E**
	It's favourite is carnival
	My favourite football time and basketball time
	My favourite football player is meyder
	My favourite friends is D.
	My favourite fotball time is Planada
	My favourite subject is P.E.
74	**My name is E.**
	My favourite color is blue
	My favorite musica is [...]
	My favorite fod is spagettis
75	**My birtday is on July.**
	My birtday is in 29th July
	My favorite equip is Barcelona
79	**My favorite country is Finland.**
	My favourite frends is a S. and A.
80	**My favourite countri is Finland.**
	My favourite green is a L. T. S., is a very very green
	My's favourite proff's is a G. (tutor)
82	**My favorit te basketball and futball**
	My favorit basket palyer is [...]

	My favorit futball player is [...]
	My favorit basket tim roureda.
	My favorit selecion (?) en basketball
85	**My favorite school and prof is E.**
	And favourite satject is?
	My favorite clothes is?
	My favorite book is?
	My favorite prof is?
	My favorite prof is E.
	My favourite family is?
127	**My name is E., you age 10 age.**
129	**My favorite animal is dog**
131	**Im from Spain**
133	**My name is R.**
134	**My name is E.**
	My favorit "asignatura" is the naturals
	My favourite book is the [...]
	E. is bery ticher "favorite".
135	**My name is A.**
137	**My name is J.**
	My classtoom favourite is art and P.E.
141	**My feiborit "asignatura" is histori**
	My preferit clas is the my clas
148	**My favorit clas is gimnastic and naturals**
165	**My name is V.**
	My favorite sports: football, [...]
	My favorite food is potatoes.
167	**My name is M.**
	My favorite colour is yellow or orange
171	**My name is J.**
	I my from St.Pere de Ribes
	I my favorit sport is [...]
	I my favorite color is [...]
172	**My name is J.**
	I my from is St Pere de R.
177	**Hello my name is D.**
	I my from is S. P de R.
180	**Hello my name is P.**
183	**My name is T.**
	I from St. P de R
205	**Hi my name is P.**
210	**Hi my name is R.**
244	**my favorite class is the E.F.**

245		**My favori color is pink**
		I from Barcelona
		My favori animal is tiger
250		**May Favorit singers is Xakira**
		May Favorit Color is green and pink
		May Favorit TV program is (?)
		May favorit classrom are is inglish [...]
		May Favorit sport is wsimin-
252		**I favorit colour: green**
254		**Hello! Mynameis B.**
	Level 2	
1		**My favourite signature is Art.**
7		**Hello,my name is Y.**
		My favourite subject is sciences
65		**My name is N.**
		I'm eleven years old.
		My favorit color is pink and red.
		My favourite subject is catalan.
		My favorite teacher is Esther [...]
		My favourite number is 8
		My favourite sport is basketball
68		**My favorite tycher is: Esther [...]**
		My favorit sport is basketball, is very, very, very amazing.
69		**My name is E.**
		I'm eleven yeras old.
		My favorite subject is E.
		My favorite color is green
		My favourite number is 7
		My favourite friend's is N.
70		**My favourite frinds is C. [...]**
		My favorit subject is A. and [...]
		My favourity boock is T. [...]
		My favourit color is green.
		My favourit school is L.R.
		My favourit TV is my
72		**my name is R.**
		My favorite festa is Carnival [...]
		My favorite colour is green dark.
		My favorite project is Comenius Beca.
		My favorite animal is cat.
76		**My favorite subject is maths and [...]**
		my favourite teacher E.
		My favorite country in Comenius Beca is Italy

	My favourite firnds in V.
	I'm 12 years old.
	My favorite football player is Puyol.
77	**My name is A.**
78	**My name is m.**
	My ticher favourite is E.
	My favourite court is red.
	My favourite project is Undertvaser
83	**My name is G.**
84	**Hi! My name is C.**
116	**My name is A.**
	I have eleven years old
118	**My favorite game is [...]**
	My favorite food is pure.
	My favorite sing is [...]
	My neme is J.
	My favorit canal the TV is [...]
	My favorit song is [...]
119	**My back favourite is Septimus**
	I'm 11 years old.
	My music favourite is rap and pop.
120	**My name is S.**
121	**The signatura preferit is M. and [...]**
	The tutora preferid is A.
122	**My favourite school si S. A.**
	My name is F is 11 years.
	I'm from is Catalonia.
	My favourite forever is C.
	The color favourite is red and blou.
125	**My name is N.**
	My favourite food is [...]
	My favourite single is [...]
	My favourite color is green.
128	**It's my name is A.**
	My favorite colour is black and white.
	I am asignatura favourite is E.
132	**Hello! My name is C.**
	And I'm 11 years old.
136	**Hello! My name is M.**
	My age is 11 years.
	My favourite teacher is J.
138	**Hello my name is M.**
	I'm eleven years old.

139	**My favorite subject is English for the teacher M.**
142	**Hello! My name is A.**
	My subject favorite is A. [...]
144	**Hello my name is J**
	My favorite class is PE
164	**My favorite friends is F. [...]**
166	**My favorite sport is football**
	My favorite color is blau
168	**My favorite color is Blau**
	My favorite food is picca
	My favorite sport is tennis.
175	**I am 10 years old.**
178	**Hello!My name is J.**
	I from St. Pere de R.
	I have 10 years old.
182	**Hello my name is A.**
	My faborite color is Blue
	I from is es espain.
185	**My favourite frin is O.**
186	**My name is O**
187	**My name is O.**
	I am 10 years old.
190	**My favorite game is [...]**
202	**Hello My name is V.**
	I am 11 years old.
211	**Hello!!My name is J.**
242	**HelloMy name is R.**
	I'm from F.
	My faborite food is chips
	My faborit color is with
243	**My name is E.**
	I'm 10 years old.
	My favourite colours are [...]
	My favorite number is [...]
	My favorite sport is [...]
	My favourite singer is M.
	I'from is E.
246	**I'm ten years old.**
	My color favourite is blue
253	**Hello My name is B.**
	And ten years old.
	From in Barcelona
	Color favorit and blue

284		**My favorite animal is horse**
		My favorit programme is P. [...]
289		**Hello my name is J.**
		basketball is my favorite sport
		My favourite team is : DKV juventos
		My favorite eat is pizza!
291		**My name is J.**
		My favourite colour is blue.
		My favourite subject is [...]
		My favourite football player is R. [...]
		My favourite pop star is E.
		And my favourite film is Shreck
294		**My name is Monica**
		My favorite animal are the dog
		And my favorite colour are the orange
		My favorite programme are the quiz shows
		My favorite movi are G[...]
298		**I'm 10 uears old**
		My favourite color is blue and
		my favourite animal is cat
	Level 3	
140		**Hello my name is M.**
		I eleven
		My favourite subject is A.
		... but is my favorite teacher.
281		**Hello my name is M**
		I come from B.

Syntactic formulas in the students' essays

Essay		Syntactic Formulas
	Level 1	
4		**I do like subject is P.E., E. and A.**
		I do not like subject is nature and [...]
5		**I like (la meva classroom)**
64		**I like school Roureda**
74		**I can school la Roureda**
		I can carnival [...]
79		**We have got very tree and very very bigs**
80		**we have got a trees [...]**

		we have got: basketball, football, chess [...]
		we have got the progect the Comenius
127		**Have you got a black curly hair.**
		Have you got eyes green
		I love my dog G.
131		**I got a mathe, P.E., G. [...]**
		I can A..
133		**I like play tennis**
171		**I like run**
		I like jump
		I have four cats
		I have one dog
172		**I like class**
		I like Schiping (?)
180		**We have got doing 5è**
183		**I like basket ball**
		I like music an Tokio Hotel
		I like neim: T.
		I like guitar electric
		I don't like school
		I like schliping
		I don't like fut ball
184		**de school has got a 400 boys and girls**
		My clas has got a 27 boys and gils
201		**In the school have got a friends**
245		**I love is tenis**
252		**I lake: Barcelona**
		I love; music
		I lake: english
254		**I love football and basketball**
		I love pizza
282		**Haven't got do you football**
		Have you got some a basketball
		Have you got the bol
		Have do you spell music
287		**But i don't like the party**
		I don't like the history
	Level 2	
65		**My school hav a big garden**
		My school has got more (?)
67		**I like school L.R. And I like he's teacher's.**
		II like Comenius project.

	In the school we are got a: very very by playground basketball court [...]
68	**he have got tree very very very big:**
	have got one library.
69	**I like grut (?) and ice-cream**
70	**I love the film Titanic**
76	**I have got a brother.**
	In my class have got a 24 boys and girls
	I love my father and mother and brother.
77	**I like carnival and**
78	**I like christmas and halloweend**
	I like carnaval
	I like football and basketball.
	I like Comenius projecte
	I like can play basketball
	I like christmas
83	**I like the subjects is m.**
	I don't like geography
	I love christmas and halloween
	I don't like carnival
	in school i like library and class
	I like basketball and bootball
	I love my school
116	**I have got very shoes, T-shirt and trainers.**
	And have got friends from Scotland.
119	**I like me school.**
	I have got hair short and brown.
121	**(I'm) has got a 10 years**
122	**(I'm) have got is big eyes, is tall, [...]**
	There are (?)
	There are our calendar.
	The Bugs Baul there are very well.
	There are is computer is a classroom.
125	**(I'm) like basketball.**
	(I'm) like music.
	(I'm) like English and PE
128	**I love music is Pop**
136	**I like back to school**
149	**I like palyground because [...]**
	I like computer because [...]
185	**I have a notbuck**
	I have a file
186	**I have got a pen.**

	I like school
187	**I like smack down**
193	**I like the bigest of the school**
197	**I lake go of school**
	I lake play bootball
200	**(I'm) like the maths**
207	**I like ed.fisica**
211	**My class there are 26 children, 11 girls and 15 boys**
243	**I'like English**
	I'like school
	I'dont like s. and a.
	I'dont like dancing [...]
	I'like skateboarding [...]
246	**I love skateboardinf**
	I like E.
253	**Ay like Football**
284	**I like ruun**
	I like sports
	I have horse and dog [...]
	I like Montesa and Bultaco
	I have one horse in Vall d'Aran
289	**I don't like worcks**
	I always watch cartoons in the morning
291	**I like eat sushi**
	In my free time I like playing football
292	**I never watch cartoons**
	I never watch wildlife programmes
	I sometime watch horror films
	I usually watch comedy films
	I never watch quiz show
	I sometime watch dream films
	I never watch the news
	I ususally watch fantasy films
	I sometime watch halloween films
	I ususally watch adventure films
	I like the television.
294	**I like play basketball**
	I like look TV
	I like spagety but
	I don't like the eggs
	I really ike the theater
296	**I usually watch cartoons in the evening and sometimes watch films in the Saturday**

		I never watch quizshows and music programmes in the morning
		I always watch quiz shows in the afternoon
		I really like football and [...]
297		**Panallets it can be of chocolate [...]**
		There are sweet but I like it.
300		**He can run**
		He can't flying
		He like tomato
		He don't like strowerrys
		He can't talk
		He can jump
		He like playgraun
		He don't like school
		He live in barcelona
		He like play football
	Level 3	
140		**I like my school**
		I don't like...
290		**I like Halloween**
302		**I like spring**
		I like the traditions of spring
316		**I think**
		I think
		I think so

Morphological formulas in the students' essays

Essay		Morphological Formulas
	Level 1	
2		**I'm starts the school [...]**
3		**I'm (començar) the school 12th of september**
4		**I'm do like is playroom.**
5		**I'm (?) (vaig començar) the school [...]**
6		**aim star play the guitar**
		aim ticher is M.
64		**I'm counties and Europa, España [...]**
		[...] playing de Football [...]
73		**I'm brother name is C.**
75		**I'm scholl is pencils, tables, [...]**
		I'm activitis my scholl is Halloween [...]

		I'm is 24 class
79		**I'm love the garden [...]**
80		**I'm school is a very big.**
		I'm love my school; is a big garden and [...]
82		**I'm brother nom is R. [...]**
123		**I'am is short hair**
		I'am is brilliant
127		**I'am esport a summer**
137		**My school I'm the mountain**
180		**The school doing English**
		Doing 2 brother and one sister
191		**ay am play basketball**
196		**Im school ys ceip de practiques**
		Im classroom is 6è
		Im classroom is big
		Im school studi: spanish, [...]
247		**I'm school is big**
		I'm ticher is Y.
		I'm ticher is guay
		I'm apres colors: black, red [...]
	Level 2	
69		**In christmas playing a lingo**
70		**I'm profof E. is E.**
		I'm profs is G.
116		**I going to school at [...]**
121		**I'm (has got a 10 years)**
122		**I'm (have got is big eyes, is tall)**
125		**I'm love school**
		I'm (like basketball).
		I'm (like music).
		I'm (like E and PE)
128		**I am (asignatura favourite is E)**
190		**I'm live in Tarragona**
193		**I'm miss the old teachers**
199		**I'm play futball**
200		**I'am is the 6è [...]**
		I'm live in catalonia [...]
		I'm to school study Matemàtiques [...]
		I'm (like the maths)
289		**I playing basketball**
	Level 3	
—		**—**