

THIS IS AN EARLIER VERSION OF THE MANUSCRIPT. FOR THE FINAL VERSION, PLEASE CHECK THE JOURNAL WEBSITE: Serrano, R., Stengers, H., & Housen, A. (2015). Acquisition of formulaic sequences in intensive and regular EFL programs. *Language Teaching Research*, 19, 89-106.

Acquisition of formulaic sequences in intensive and regular EFL programs

Raquel Serrano

(Department of English Studies,) Universitat de Barcelona, Spain

Helene Stengers

(Applied Linguistics Department,) Vrije Universiteit Brussel, Belgium

Alex Housen

(Department of Linguistics and Literary Studies,) Vrije Universiteit Brussel, Belgium

Abstract

This paper aims to analyze the role of time concentration of instructional hours on the acquisition of formulaic sequences in English as a foreign language (EFL). Two program types that offer the same amount of hours of instruction are considered: intensive (110 hours/one month) and regular (110 hours/seven months). The EFL learners under study are adults at the beginner (N=35), intermediate (N=44) and advanced levels (N=45). A group of native English speakers (N=12) served as a benchmark. The focus of this study is on the number and range of formulaic sequences the participants used while performing an oral narrative. The results of the statistical analyses show a slight advantage for the learners in the intensive program, especially at the intermediate level, both in terms of frequency and range of formulaic sequences produced. Moreover, results suggest that there are still marked differences between even the advanced EFL learners in our sample and the native speaker benchmarks, again both in terms of number and range of formulaic sequences.

Keywords: formulaic sequences, intensive language courses, EFL, oral production.

Introduction

Studies in both cognitive psychology as well as in first and second language acquisition have suggested that learning is not only influenced by the sheer amount of practice but also by how the time devoted to practice is distributed in time¹. Research in cognitive psychology has demonstrated that repetitions of the same linguistic item or structure foster learning (Pavlik & Anderson, 2005). Moreover, it has been shown that spaced repetitions (including longer time lapses or various intervening items) are more beneficial for learning after a particular treatment/instruction period and especially for long-term retention than massed repetitions (presentations of the same item appear subsequently or with little time/few intervening items in between) (Bahrick & Hall, 2005; Dempster, 1988; Mammarella, Russo, & Avons, 2002; Pavlik & Anderson, 2005; Toppino & Bloom, 2002; Toppino, Hara, & Hackman, 2002). This phenomenon is known in the cognitive psychology literature as the *spacing effect*. This effect is quite robust; however, different variables can modulate the effect. For example, when paraphrased rather than verbatim repetitions of the target feature are included, the spacing effect is either reduced or disappears altogether (Glover & Corkill, 1987; Mammarella et al., 2002). Additionally, at short retention intervals, when testing occurs shortly after the learning sessions, massed items seem to be recalled as well as if not better than spaced items (Bahrick & Hall, 2005; Pavlik & Anderson, 2005; Rohrer & Taylor, 2006). Finally, when the spacing between the items or learning sessions is too wide, detrimental effects tend to be obtained, since participants cannot retrieve the first presentation of the target item. As Bahrick and

Phelps (1987: 349) suggest, “The optimum interval is likely to be the longest interval that avoids retrieval failures.” From this we can conclude that the ideal spacing between repetitions cannot be determined *a priori* and will depend on various factors, including the nature of the target item and the type of repetition. In any case, though, it is important that participants are able to retrieve the cognitive trace of previous presentations in order for subsequent repetitions to have an effect on participants’ memory (Bahrick and Phelps, 1987).

In the case of language acquisition, both first language (L1) and second language (L2), repetitions of patterns, words, sequences, etc. and extensive practice are also crucial for learning. One of the main differences between L1 and L2 acquisition is related to the significantly different amount of time that L1 and L2 speakers engage with the target language. As N. Ellis (2001: 36) remarks, language fluency requires a massive number of hours of practice, something L2 learners rarely dispose of: “Fluent language users have had tens of thousands of hours on task. They have processed many millions of utterances involving tens of thousands of types presented as innumerable tokens.” Apart from the total amount of time on task, another difference between L1 and L2 acquisition is the concentration of such time. In L1 acquisition, children are fully immersed in the language and receive intensive exposure to a gradually increasing range of items and structures adapted to their developing proficiency in the language. With such type of exposure, it is more likely that repetitions of the labels for different objects and actions relevant to the learner will reappear within reasonable intervals; therefore, retrieval of former presentations of those items will be facilitated.

Similarly, second language learners are more likely to notice (Schmidt, 1990) the patterns of the language, when such patterns are repeated constantly and immediately. In the case of the object of the present study, *formulaic sequences* (FSs), or sequences of words that tend to appear together (or collocate), it seems logical to expect that associations between words are established more easily when those sequences are repeated many times and under a relatively concentrated schedule, in which learners do not have time to forget previous presentations of those sequences. Also, the *types* of FSs come in many different guises or *tokens*. However, as alluded to earlier, it seems equally logic to expect that whatever impact the distribution of practice may have on the learning of FSs, this impact will also be mediated by factors such as the (token) frequency of different FSs and other, more structural features as number of constituent words, lexical class to which constituent words belong, degree of morphological variation displayed by the constituent words, and so forth (Stengers, Boers, Housen, & Eyckmans, 2011).

For most classroom L2 learners (which is the context under study here), exposure to the L2 tends to be *distributed*, with few hours of L2 input and output practice per session and widely spaced sessions (with several days in between). Foreign language teaching programs typically consist of two sessions of one to two hours per week. Such limited exposure does not facilitate the proceduralization and automatization of L2 skills (DeKeyser, 2001; Robinson & Ha, 1991; Schmidt, 1992; Schneider & Chein, 2003) or the implicit acquisition of language (DeKeyser, 2000; DeKeyser & Larson-Hall, 2005). We can expect that, under this schedule, it will be relatively harder for learners to establish associations between words and recall previously encountered sequences in the L2.

Apart from the typical classroom schedules for learning languages (regular programs), which include short and widely-spaced sessions over a long period of time (usually several years), also other programs are available for L2 learners that include an extensive amount of hours of contact with the L2 over a short period of time (cf. the English for Academic Purposes programs offered by many English-speaking universities or the intensive English programs in some Canadian primary and secondary schools; Lightbown & Spada 1991; Collins et al. 1999). Even though the total amount of time on task in these programs may still be far from the amount of L1 exposure children receive, time distribution is not so significantly different. There is some variability in the design of intensive courses, but in general these programs tend to provide L2 input and output practice for at least four hours every day, five days a week (Serrano, 2011a). It seems reasonable to expect that such a schedule will be more beneficial for the acquisition of the vocabulary and many FSs of the target language than regular programs, as it may be easier for learners to recall previous presentations of those sequences and thus commit them to long term memory. In other words, the spacing of the presented FSs may be hypothesized to be more conducive to learning in an intensive program than in a regular program, where the spacing intervals are probably too wide to allow recall of former presentations of the items. As Durrant and Schmitt (2010) suggest:

Indeed, since learning a collocation will involve retaining some memory trace of any particular word pair met until that pair is met again, it may be that the relatively sparse nature of most second language input (totaling to perhaps a few hours a week) will mean that the extended time that elapses between two exposures to a collocation

is usually too long and that trace will be lost, with the result that learning of any but the most frequent collocations can never properly get off the ground (p. 169-170).

There is some empirical evidence indirectly supporting this hypothesis. Several studies in the SLA field have suggested that intensive L2 instruction is more effective than regular instruction for the development of a variety of L2 skills. The benefits of intensive L2 instruction have been reported in the case of both children (Collins et al., 1999; Collins & White, 2011; Lightbown & Spada, 1991; Spada & Lightbown, 1989; White & Turner, 2005) as well as adults (Serrano, 2011a) for all the major language skills (listening, oral production, writing, and reading), and also in terms of vocabulary and grammar. Most studies have included learners at the beginning or intermediate levels of proficiency and development. In the cases in which advanced learners were included, the advantages of intensive instruction were not as obvious (Serrano, 2011a).

However, even though the studies of intensive instruction have analyzed many aspects of L2 performance, to the authors' knowledge, there is no study comparing the acquisition of formulaic sequences in intensive vs. regular programs. We are adopting the term formulaic sequence, or FSs, to include all standardized multiword expressions, although in the present study we will only focus on specific types of FSs. Other terms used in the literature for FSs include "chunks", "collocations", "composites", "conventionalized expressions", etc. (Wray, 2000). One of the most commonly cited definitions is Wray's (2002: 9), according to which a FSs is "a sequence, continuous or discontinuous, of words or other elements, which is, or appears to be, prefabricated: that is, stored and retrieved whole from memory at the time of use, rather than being subject to generation or analysis by the language grammar." Whereas the alleged

prefabricatedness of FSs is less arguable in first language production, there is still controversy as to how L2 learners acquire and produce these sequences from a psycholinguistic perspective (Durrant & Schmitt, 2010; Stengers et al., 2011). There is agreement, however, on the claim that the use of FSs contributes to L2 learners' fluency (Granger, 1998; Pawley & Syder, 1983; Skehan, 1998; Wray, 2002). Empirical evidence in support of the claim that appropriate use of FSs can help learners of English reach a higher level of oral proficiency not only in terms of fluency, but also in terms of range of expression and accuracy, has been reported by Boers et al. (2006) and Stengers et al. (2011).

Several studies have examined the acquisition of FSs in *intensive* courses in the target L2 country (i.e., the “study abroad” context). Schmitt, Dörnyei, Adolphs and Durow (2004) examined the development of receptive and productive knowledge of FSs in L2 English over an intensive course in the UK². The learners in this course received explicit instruction on the target formulas. The authors found that the participants improved in both production and receptive knowledge of FSs; nevertheless, they acknowledge that it is not possible to know whether such improvement was due to explicit instruction or to the increased and concentrated exposure that is typical of intensive courses. Dörnyei, Durow and Zahra (2004) analyzed to what extent “acculturation” or “the social and psychological integration of the learner with the target language group” (Schumann, 1986: 379) and language aptitude affect the acquisition of FSs in the same context as the aforementioned study. These authors found that involvement in the L2 community is key for the acquisition of FSs; however, other variables such as high aptitude can compensate for lower degrees of acculturation.

The present study differs from the three previously mentioned studies in three respects. First, the present study includes a benchmark group receiving regular instruction and not just a group in an intensive course. Second, the general language learning context under examination is different; the L2 was the official language in the settings where the Dörnyei et al. and Schmitt et al. studies took place (i.e. an ESL, or English as a Second Language context) but in the present study exposure to the L2 is mostly restricted to the classroom (i.e. an EFL, English as a Foreign Language context). Given that the classroom is typically the dominant or even sole source of exposure to L2 in foreign language contexts, as opposed to in second language contexts where the learners are also exposed to individually variable amounts and concentrations of L2 input in the wider context outside the school, foreign language contexts offer a greater amount of control for studying the effect of factors such as amount, type and concentration of exposure than do second language contexts. Thus, in the context under analysis in the present study, concentrated exposure is unique to the intensive program and it only happens in the classroom, as the target language (English) is not an official language in the setting in which the study was performed (Catalonia, Spain). In this context it is easier to examine the role of time concentration of instructional hours alone.

A third difference between the present study and the studies by Schmitt et al. (2004) and Dörnyei et al. (2004) is that the programs under analysis did not include any instruction targeting FSs (for a comprehensive review of intervention studies see Boers & Lindstromberg, 2012). Given the fact that learners were not given instruction on FSs, their acquisition of FSs relied on learner-autonomous incidental uptake (see Hulstijn, 2001 on the distinction between incidental vs. intentional learning).

Finally, a wider range of L2 proficiency levels are included in the present study than in previous studies in order to examine whether this variable has an effect on the acquisition of FSs in intensive programs. According to the results reported in Serrano (2011a), intermediate learners may benefit more from intensive exposure than advanced learners in terms of listening and reading comprehension, grammar, and lexical complexity in writing. It would be of both theoretical and practical interest to examine whether similar results are obtained for FSs.

Apart from intermediate and advanced learners, this study also includes beginners. Studies on FSs typically involve upper-level learners as FSs are thought to be a late(r) development in SLA (Skehan 1998). However, there are reasons to believe that FS also manifest themselves in the L2 production of lower-level learners as well (Smiskova & Verspoor, in press). We aim to analyze whether intensive exposure is more beneficial for lower-proficiency learners (in which case beginners should obtain the most benefits from this type of instruction), or whether a certain command of the L2 is necessary before learners can benefit from intensive exposure to the language: beginners might feel “overwhelmed” by the amount of novel input to be processed in a relatively short time and selectively allocate their attention to smaller, more easily segmentable elements in the input stream (e.g. single lexical items). Similarly, considering that the acquisition of formulaic sequences is probably incidental (Ellis, 2002; 2012), it is conceivable that intermediate and advanced learners might be more successful at autonomously processing the syntagmatic structure of the input presented to them than beginners.

Advanced learners' output in terms of FSs was compared to that of a group of native English speakers in order to check for a possible ceiling effect for any of the groups in the two programs.

Summarizing, the goal of our exploratory study is to answer the following research questions:

1. Does time distribution (intensive vs. regular) affect the acquisition of FSs in EFL classroom settings?
2. If time concentration has an effect on the acquisition of FSs, is it the same or different for learners of different proficiency levels (beginner, intermediate and advanced)?
3. At the advanced level, how do students in the regular and the intensive program compare to native English speakers?

It is hypothesized that concentrated time distribution in intensive programs will foster the acquisition of FSs due to the increased possibility of repetitions of different types of FSs occurring within an interval in which previous presentations of those sequences are still active in learners' memory: new presentations would enhance such representations. On the other hand, since time intervals between repetitions in the regular program are too widely spaced, learners are perhaps more likely to forget previous presentations of sequences in the input. Regarding proficiency level, as previous studies have suggested that intensive instruction is especially beneficial for less advanced learners (Serrano, 2011a), we hypothesize that, in terms of FSs acquisition, especially beginners and intermediate learners in intensive programs may have a greater learning advantage than

their peers in the regular program. Such advantage may be less apparent for advanced learners. Regarding the difference between beginners and intermediate level students: intermediate students could be hypothesized to be better at autonomously “picking up” FSs than beginners.

Method

Programs and participants

The participants in this study include 124 EFL learners with Spanish/Catalan as their L1³. The participants were adult students, most of them (65%) female, between 18 and 23 years old, who were enrolled in English courses at the language school of a university in Catalonia, Spain. Most of the participants (89%) were undergraduate students, while the remaining 11% were young professionals. The students were all comparable in terms of motivation and previous experiences with English, as they indicated in a background questionnaire.

These learners were enrolled in two program types: intensive (N=58) and regular (N=66). Both programs offered 110 hours of English instruction distributed over four and a half weeks in the summer (five five-hour sessions a week) in the former, and over seven months during the academic year (October-May) in the latter program (two two-hour sessions a week). The methodological approach, textbooks, exams, etc. were the same for the intensive and the regular program, the main difference between the two being time distribution. The approach followed in all these classes was quite traditional, with a

special focus on grammar and vocabulary, although the different language areas were practiced in written as well as in oral communicative activities. One difference between the two program types was that in the intensive courses there were usually slightly more audio-visual activities to make the longer sessions sufficiently engaging for the students.

Three different proficiency levels were considered, as determined by their class level and on the basis of a range of independent proficiency measures in terms of complexity, accuracy and fluency (Serrano, 2011): beginner (N=35), intermediate (N=44), and advanced (N=45) (see Table 1 for details on participants and programs). The equivalent levels as defined by the Common European Framework of Reference for Languages were A1, B1 and B2/C1 respectively.

[Table 1]

Additionally, 12 native English speakers (NES) from the United States were recruited in order to provide baseline data. As mentioned before, the main reason why such data was considered necessary was to check for ceiling effects: the advanced learners in the intensive or in the regular program might not show progress because their performance might be native-like at the beginning of their program for the aspect being investigated, in this case FSs (in Serrano, 2011b a ceiling effect was found for advanced learners in terms of some measures of written production). The literature on FSs, however, suggests that advanced L2 learners might still be far from “native-like” formulaic use, with learners’ phrase production showing patterns of overuse, underuse or misuse (De Cock, 2004; Granger, 1998; Nesselhauf, 2003; Siyanova & Schmitt, 2007; Weinert, 1995;

Wray, 1999; Yorio, 1989). Our purpose is to examine whether this is also the case for the constructions under analysis and for the participants under study. The profile of the NES was comparable to the EFL learners' profile: they were undergraduate students who were also learning another foreign language, in this case, Spanish.

Instrument and procedure

In order to examine learners' use of FSs, we analyzed learners' L2 performance in an oral narrative based on a series of pictures: *The picnic story* (Heaton, 1966). This task was extensively used for research purposes by the "Barcelona Age Factor Project" (see Muñoz, 2006), and in a variety of studies in other contexts (Serrano, 2011; Serrano, Llanes, & Tragant, 2012; Collins & White, 2011; Llanes & Muñoz, 2009; Tavakoli & Foster, 2008). The participants were shown six pictures that represented two children preparing a picnic with their mother. While the children are preparing the picnic their puppy dog gets into their picnic basket and eats their food. When the children are ready to eat their sandwiches they notice that their puppy has eaten everything and they have no food left.

The intermediate and advanced learners in the two program types took the test twice (pretest/posttest), once at the beginning and once at the end of the course in order to gauge the change or progress in the use of FS over the course of the program.

Additionally, analysis of the pretest allowed us to check whether the two groups at each proficiency level were comparable and whether the different proficiency levels were different in terms of FSs at the start of their respective courses. For obvious reasons,

beginners did the test only once at the end of their instructional period. Finally, the NES also took the test once, as their data was only used as benchmark data. In all cases, the students were given around 30 seconds to become familiar with the pictures and when they were ready they narrated the story.

FSs Coding

Before deciding on which FSs to focus on, we read the transcriptions of the oral narratives carefully and observed the type of formulaic language that was produced by both English learners and native speakers, keeping in mind the taxonomies developed by other researchers (Nattinger & DeCarrico, 1992; Granger & Paquot; 2008). Additionally, among the range of FSs that could be included we decided to focus on those that were amenable to objective and systematic coding, taking corpus-based frequency information into account when appropriate. The FSs that were considered for this study could be classified according to their function (a and b) and their lexical make-up (c through f).

- a) Discourse-structuring devices (DSD), such as *first of all, in the next (picture), in conclusion, etc.*
- b) Fluency devices (FD), characterized as follows by Nattinger & DeCarrico (1992:84) “essentially these phrases buy time to help one gather one’s thoughts. They not only promote fluency but also indicate to the hearer that one has not given up, thus serving to rebuff interruptions (turn shifts)”. Some examples include *I don’t know, I think/I suppose, it looks like, or something like that...*

- c) Verb plus particle/preposition (phrasal and prepositional verbs) (VP); e.g. *find out, go away, sit down*, etc.
- d) Verb plus noun (VN); e.g. *climb a hill, fill a bottle, say goodbye*, etc.
- e) Verb plus preposition plus noun (VPN)⁴; e.g. *go up a hill, sit on the grass, go for a walk*, etc.
- f) Verb plus two prepositions (VPP); e.g. *get out of, jump out of, sit down on*, etc.

Two researchers were in charge of the coding. For the discourse-structuring devices and fluency devices the identification of these FSs was mainly guided by Nattinger and DeCarrico's taxonomy (1992), as well as Biber et al. (1999). For the identification of verb FSs (VP, VN, VPN, VPP), the Corpus of Contemporary American English (COCA) was used (Davies, 2008). This corpus includes 425 million words used in different genres (spoken, fiction, popular magazines, newspapers, and academic) between 1990 and 2011. All identified verb collocations with a mutual information score of three or higher were considered as verb FSs (in accordance with Hunston 2002; Stubbs 1995). The Mutual Information (MI) score is a statistical measure expressing the extent to which observed frequency of co-occurrence differs from what could be expected from a statistical point of view. MI provides a "strength of association" between words. MI will compare the frequency of co-occurrence with the overall frequency of the individual (co-occurring) words. Even though we believe the MI score is a reliable and objective measure to analyze FSs, we are aware of its limitations (see Ellis, Simpson-Vlach, & Maynard, 2008 for an analysis of FSs metrics including length, MI, and frequency and their effect for native and non-native speakers' processing). One limitation is that, when the individual

words forming the sequence are very frequent and frequently collocate with other words, MI scores tend to be low. Therefore, when we considered that some verb sequences were formulaic but did not have a high MI score, we decided to check the *Oxford Collocation Dictionary* to verify our intuitions. Additionally, we checked the raw frequency of these sequences in the COCA corpus and they all happened to have a frequency higher than 300, which we defined as the frequency cut-off point to include sequences with low MI score in our analysis. Some examples of FSs in this category include *have lunch, make coffee, make a trip, have fun, or go to school*. As can be seen, most of these FSs included the verb *have* and *make*. These verbs are high frequency verbs that collocate with many other words, which is why the MI score of these sequences was low.

We ignored pauses between words within a FS and included word sequences such as *the mother uhm filled a uhm ... bottle* on our list of FSs⁵. Only target (i.e. native)-like FSs were considered in the count, and those containing errors were discarded (e.g. *in the one hand, how you say...?*).

The two researchers who were in charge of the coding first coded together 5% of the speech samples, in order to make sure they had comparable coding criteria. Then, they coded 25% of the sample separately and their respective codings were correlated to test for consistency. The Pearson correlation coefficient of .82 suggested that the coding was consistent, and therefore only one researcher coded the remaining samples.

Analysis

The CLAN program (MacWhinney, 2000) and the Statistical Package for the Social Sciences (SPSS) were used for coding and analyzing the oral narratives. We decided to focus on the *total* number of formulas that learners used, regardless of their classification (see section FSs coding). The number of sequences of each category was quite low; therefore, differences between groups are hardly noticeable. Additionally, in this particular study we are interested in the degree of formulaicity of learners' language, not in whether, for example, learners produce more VN sequences than VP sequences.

In order to examine learners' use of FSs, we considered both the "tokens", or individual instances formulas, and the "types" (i.e. each FS regardless of how often it occurred and regardless of the morphological variants in which it occurred). For example, *the children go away* and *the boy goes away* were counted as one "type" but two "tokens".

Additionally, as the learners produced narratives of different length, it was considered appropriate to control for text length by analyzing ratios of FSs instead of raw scores. We divided the number of (types and tokens of) FSs by the total number of words produced and multiplied it by 100 (to obtain numbers higher than 1). All the analyses reported in the results section use ratios and not raw scores.

Different statistical analyses were performed for different comparisons. In the analyses involving the participants who did not perform the test twice (comparison between beginners in intensive and regular programs, and between advanced learners and native speakers), independent samples *t*-tests were performed to examine between-groups comparisons. In the case of the learners in the intermediate and advanced levels, it was considered more appropriate to perform a more powerful test, Analysis of Covariance (ANCOVA), with *level* and *program type* as independent variables, the ratio of *FSs types*

and *FSs tokens in the posttest* as the dependent variables, and the ratio of *FSs types* and *FSs tokens in the pretest* as covariates. Before conducting the ANCOVAs we checked that the data did not violate any of the assumptions this type of analysis requires.

Results

This section presents the results of the comparisons between the intensive and the regular program type for each of the three proficiency levels in terms of the number of FSs types and tokens produced.

Beginners: Intensive vs. regular

As mentioned before, the beginners were only tested at the end of their course. Table 2 shows the descriptive statistics for all the learners.

[Table 2]

The learners in the intensive program used more FSs (both in terms of types and tokens) than those in the regular program. This difference is significant in the case of tokens ($t(33)=2.49, p=.018$) (the effect size of this difference being large according to Cohen's d (0.87)), but it was not significant in the case of types ($t(33)=1.73, p=.093$).

Intermediate and advanced learners: Intensive vs. regular

Tables 3 and 4 show the descriptive statistics for the intermediate and the advanced learners respectively. It can be seen that, in the case of the intermediate learners, those in the intensive program seem to use more FSs than those in the regular program at both testing times but the difference between the two program types seem to be especially noticeable in the posttest. It can be also observed that, surprisingly, the learners in the regular program produced slightly fewer types and tokens on the posttest than on the pretest.

[Table 3]

[Table 4]

The descriptive statistics for the advanced learners, however, show the opposite trend, with learners in the regular program outperforming those in the intensive in both number and range of FSs.

The results of the ANCOVAs for the FSs types suggest that, after controlling for pre-test scores, program type did not have any effect on learners' performance in the post-test ($F(1, 84) = .708, p = .402, \text{partial } \eta^2 = .008$). The effect of proficiency was not significant either ($F(1, 84) = 2.06, p = .154, \text{partial } \eta^2 = .024$). Interestingly, however, there was an interaction effect between program type and proficiency level ($F(1, 84) = 4.59, p = .035, \text{partial } \eta^2 = .052$), suggesting that, in terms of types of FSs, and as could be inferred from the descriptive statistics, the intensive program was especially beneficial for the intermediate learners. Regarding FSs tokens, a similar picture is found: there was no effect of program type ($F(1, 84) = .321, p = .573, \text{partial } \eta^2 = .004$) or level ($F(1, 84) = .507, p = .478, \text{partial } \eta^2 = .006$), but there was again an interaction between program

type and proficiency level ($F(1, 84) = 6.70, p = .011, \text{partial } \eta^2 = .074$) in the same direction as the one found for FSs types.

Advanced learners and NES

Finally, the performance of the advanced learners on the pretest was compared to that of a group of NES to control for ceiling effects on the one hand, and on the posttest to examine whether there is a difference in the number of FSs used by advanced EFL students and NES. Table 5 presents the descriptive statistics for pre- and posttest for the intensive learners and 6 for the regular learners.

[Table 5]

[Table 6]

The descriptive statistics indicate that NES use more FSs than the advanced learners in the two program types both at the pretest and the posttest. This difference is always significant for both types and tokens ($p < .001$), at pre- and posttest. This indicates, first, that there were no ceiling effects for these students at the beginning of their program and that the lack of differences between the learners in the intensive and regular programs cannot be attributed to ceiling effects. Additionally, these results indicate that advanced EFL learners' use of FSs at the end of their course is still far from native speakers' use in terms of types and tokens of FSs produced, regardless of program type.

Discussion and conclusion

The results of this study suggest that concentrating time distribution of L2 hours of instruction fosters the acquisition of FSs but only under certain conditions. When considering beginners, differences between program types appear only in tokens, with the learners in the intensive program using a significantly higher number of formulas. In terms of range (types), there were no statistically significant differences between the two groups, even though the learners in the intensive program produced a more varied range of FSs (as seen in the descriptive statistics), suggesting, again, an advantage for concentrating the hours of L2 instruction. However, the benefits of the intensive program can be most clearly seen at the intermediate level. The ANCOVAs performed with the learners at the intermediate and advanced levels indicate that intensity is especially favorable for the former group in both number and range of FSs. It can be claimed that the differences in types are probably more informative of the degree of formulaicity of learners' language, as they show differences in the range of FSs learners use. Differences in tokens could be due to learners' repeating specific types of formulas, which does not necessarily indicate that these learners know more FSs in English than their peers who used fewer tokens.

In view of these results we can conclude that the learners at the intermediate level are the ones that benefit the most from intensive instruction in terms of production of FSs

(especially as compared to those at the advanced level). These results are in line with other studies: Serrano (2011a) found that intermediate learners in intensive programs made significantly more gains than advanced learners in listening skills, grammar, reading comprehension and written lexical richness (in fact, a higher incidence of FSs could be a side-effect of growing lexical resources). Regarding beginners, in the present study certain differences were found between the two program types, but they were significant only with respect to the number of FSs (tokens). Repeating chunks may indicate a strategic competence to enhance fluency. Granger (1998) observed a tendency among L2 learners to overuse familiar and “safe” chunks, which serve as “*islands of reliability*”. Maybe this is especially the case at this proficiency level in the EFL class.

From these results, it can be concluded that intensity is not equally beneficial for the acquisition of FSs at all proficiency levels: learners with an advanced level do not seem to benefit from intensive instruction to the same extent as lower proficiency learners. The results of our analyses for this group indicate that their performance was not native-like in the *pretest* (in terms of the types and tokens of FSs produced); therefore, the lack of differences between the two program types cannot be attributed to ceiling effects for the learners in one particular program. In contrast, in another study, Serrano (2011b) found that the advanced learners in the intensive group were not significantly different from native speakers in written fluency and complexity in the pretest, but that the learners in the regular group were. The fact that in Serrano (2011b) there were no differences in posttest scores between advanced learners in the intensive and regular programs could be attributed to the fact that intensive learners did not have room for improvement and that

might have been one reason why intensity did not have a positive effect (as opposed to what was found for intermediate learners in that same study).

The comparison with NES at the *posttest* in the present study suggests that advanced learners' use of FSs is still quite far from native speakers' use, as other studies have suggested (Granger, 1998; Nesselhauf, 2003; Siyanova & Schmitt, 2007). It must be pointed out, however, that we can only make claims about the difference in frequency but not in the nature of the FSs used, which is probably the aspect that distinguishes the two groups more clearly. The lack of differences at the advanced level can be due to the fact that the acquisition of FSs does not increase linearly, but instead more significant progress is evident at early stages, with the learning curve gradually trailing off as L2 learners approach native speakers' level of proficiency. As happens with complex cognitive skills, and as predicted by the power-law of practice, more improvement takes place at early acquisition stages than at later stages (MacKay, 1982; Newell & Rosenbloom; 1981; Rosenbloom & Newell, 1987)

Additionally, it could be the case that longer programs than the ones under analysis here or immersion in the L2 country would be more beneficial to the acquisition of FSs for advanced level learners. In fact, it is probably easier to significantly develop one's knowledge of FSs in the context analyzed by Schmitt et al. (2004) or Dörnyei et al. (2004): a combination of immersion and classroom instruction, as the amount of exposure to the L2 is higher, more continuous, and more intensive than when input comes uniquely from the L2 class. The study abroad context may indeed be an optimal setting for the investigation of the acquisition of FSs. More studies should be performed in this context, as well as in intensive instruction programs to confirm that the tendencies that

have so far been observed for the benefit of concentrated exposure to the L2 in the acquisition of FSs are generalizable.

The fact that intermediate-proficiency intensive learners show a relatively higher use of FS is in line with the claim that, as exposure to the L2 is concentrated, the students' memory traces of former presentations of FSs are still active when repetitions occurs. These repetitions enhance the active memory representations and thus facilitate learning (Durrant & Schmitt, 2010). Also, implicit acquisition of FSs is probably fostered in intensive programs due to the frequent and concentrated exposure to the L2 in such programs, conditions which are more similar to those of L1 acquisition. It must be mentioned, though, that, since this is a quasi-experimental study, we could not control for the actual exposure to the FSs that the learners produced, and we can only tentatively offer this explanation as a possible reason for the difference between program types.

Our exploratory study focuses on *productive use* (not recognition) of only *some* types of FSs. Learners, however, recognize more words or FSs than they are actually able to produce (De Bot & Stoessel, 1999; Schmitt et al., 2004), and our results might have been different if recognition had been examined. Moreover, the task that we used was an open task in which learners were free to use or avoid FSs. Consequently, the fact that a learner does not produce certain FSs does not necessarily mean that she or he does not know them. More studies are necessary that examine both reception and production of FSs using tasks that are more specific to examine development of certain target FSs (as Schmitt et al., 2004) while also including a comparison group (the progress Schmitt et al., 2004 observed from pre- to posttest could be due to task repetition; therefore, it is always appropriate to include comparison groups to reduce the effect of task repetition).

Additionally, it would be ideal to control for the type of input the learners receive in terms of FSs and analyze how it is reflected in learners' knowledge of those sequences. In summary, more controlled quasi-experimental or experimental studies should be performed in order to examine in more detail the acquisition of FSs under different schedules. As many authors have suggested, the acquisition of FSs is crucial for learners to acquire both fluency and/or accuracy in the L2 (Boers et al., 2006; Granger, 1998; Pawley & Syder, 1983; Skehan, 1998; Wray, 2002; Stengers et al. 2011), and finding out which context or conditions foster the learning of FSs is of high relevance for the SLA field.

Acknowledgements

This research was supported by Grant FFI2010–18006 and Grant 2009SGR137.

Notes

¹ For a more detailed review on the effect of time distribution on learning in general and second language learning in particular, see Serrano (2012).

² Some participants (N=62) followed a two-month course and others (N=32) a three-month course. The two courses were similar in terms of time concentration. There were no statistically significant differences in the development of FS between the two groups.

³ Data from this study come from the corpus of learners analyzed in previous studies for other aspects of L2 performance (Serrano, 2007; 2011a, 2011b)

⁴ In VPN and VPP sequences, the phrasal/prepositional verb was only counted once as VPN or VPP and not twice (one as VP and another as VPN or VPP).

⁵ Although considering FSs with pauses might seem to contradict Wray's definition of FS (2002), it must be pointed out that Wray herself doubts whether adult second language learners can actually acquire FSs holistically. Instead, she suggests that the closest L2 learners can get at later stages is to store lexical chunks as *proceduralized* strings that are assembled from smaller parts (Wray, 2002). Therefore, phrases with hesitation patterns may indicate that these phrases are known by the learner, but perhaps not yet entirely proceduralized. It should be mentioned, though, that we performed an analysis considering FSs with and without pauses and there were no significant differences in the overall results.

References

- Bahrack H. P. & Phelps, E. (1987). Retention of Spanish vocabulary over 8 years. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 13, 344-349.
- Bahrack, H. P. & Hall, L. K. (2005). The importance of retrieval failures to long-term retention: A metacognitive explanation of the spacing effect. *Journal of Memory and Language*, 52, 566-577.
- Biber, D., Johansson, S., Leech, G., Conrad, S., & Finegan E. (1999). *Longman grammar of spoken and written English*. Harlow, Essex: Pearson Education Ltd.
- Boers, F, Eyckmans, J., Kappel, H., Stengers, H., & Demecheleer, M. (2006). Formulaic sequences and perceived oral proficiency: Putting a lexical approach to the test. *Language Teaching Research*, 10, 245-261.

- Boers, F. & Lindstromberg, S. (2012). Experimental and intervention studies on formulaic sequences in a second language. *Annual Review of Applied Linguistics*, 32, 83-110.
- Collins, L. & White, J. (2011). An intensive look at intensity and language learning. *TESOL Quarterly*, 45, 106-133.
- Collins, L., Halter, R. H., Lightbown, P. M. & Spada, N. (1999). Time and the distribution of time in L2 instruction. *TESOL Quarterly*, 33, 655-680.
- Davies, M. (2008-) *The Corpus of Contemporary American English: 425 million words, 1990-present*. Available online at <http://corpus.byu.edu/coca/>.
- De Cock, S. (2004). Preferred sequences of words in NS and NNS speech. *Belgian Journal of English Language and Literatures, New Series*, 2, 225-46.
- DeKeyser, R. M. & Larson-Hall, J. (2005). What does the Critical Period Really Mean? In J. F. Kroll & A. M. B: de Groot (Eds.), *The handbook of bilingualism: Psycholinguistic approaches* (pp. 88-108). Oxford: Oxford University Press.
- DeKeyser, R. M. (2000). The robustness of critical period effects in language acquisition. *Studies in Second Language Acquisition*, 22, 4, 499-533.
- DeKeyser, R. M. (2001). Automaticity and automatization. In P. Robinson (Ed.), *Cognition and second Language Instruction* (pp. 125-151). Cambridge: CUP.
- Dempster, F. N. (1988). The spacing effect: A case study in the failure to apply the results of psychological research. *American Psychologist*, 43, 627-634.
- Dörnyei, Z., Durow, V., & Zahran, K. (2004). Individual differences and their effects on formulaic sequence acquisition. In N. Schmitt (Ed.), *Formulaic sequences* (pp. 87-107). Amsterdam, Philadelphia: John Benjamins.

- Durrant, P., & Schmitt, N. (2010). Adult learners' retention of collocations from exposure. *Second Language Research*, 26, 163-188.
- Ellis, N. C. (2001). Memory for language. In P. Robinson (Ed.), *Cognition and second language instruction* (pp. 33-68). Cambridge: CUP.
- Ellis, N. C. (2002). Frequency effects in language processing: A review with implications for theories of implicit and explicit language acquisition. *Studies in Second Language Acquisition*, 24, 143-188.
- Ellis, N. C. (2012). Formulaic Language and Second Language Acquisition: Zipf and the Phrasal Teddy Bear. *Annual Review of Applied Linguistics*, 32, 17-44.
- Glover, J. A. & Corkill, A. J. (1987). Influence of paraphrased repetitions on the spacing effect. *Journal of Educational Psychology*, 79, 198-199.
- Granger, S. & Paquot, M. (2008). Disentangling the phraseological web. In S. Granger and F. Meunier (Eds.) *Phraseology. An Interdisciplinary Perspective*, (pp. 27-49). Amsterdam: John Benjamins.
- Granger, S. (1998). Prefabricated patterns in advanced EFL writing: Collocations and formulae. In A. P. Cowie (Ed.), *Phraseology, theory, analysis and applications* (pp. 145-60). Oxford: OUP.
- Hunston, S. (2002). *Corpora in Applied Linguistics*. Cambridge: CUP.
- Lightbown, P. M. & Spada, N. (1991). Étude des effets à long terme de l'apprentissage intensif de l'anglais, langue seconde, au primaire. *The Canadian Modern Language Review*, 48, 90-117.
- Llanes, À. & Muñoz, C. (2009). A short stay abroad: Does it make a difference? *System*, 37, 353-365.

- MacKay, D. G. (1982). The problems of flexibility, fluency, and speed-accuracy tradeoff in skilled behavior. *Psychological Review*, 89, 483-506.
- MacWhinney, B. (2000). *The CHILDES project: Tools for analyzing talk. Third Edition*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Mammarella, N., Russo, R., & Avons, S. E. (2002). Spacing effects in cued-memory tasks for unfamiliar faces and nonwords. *Memory & Cognition*, 30, 1238-1251.
- Nattinger, J. R. & DeCarrico, J. S. (1992). *Lexical phrases and language teaching*. Oxford: OUP.
- Nesselhauf, N. (2003). The use of collocations by advanced learners of English and some implications for teaching. *Applied Linguistics*, 24, 223-42.
- Newell, A., & Rosenbloom, P. (1981). Mechanisms of skill acquisition and the law of practice. In J. R. Anderson (Ed.), *Cognitive skills and their acquisition* (pp. 1-55). Hillsdale, NJ: Erlbaum.
- Pavlik, P. I. & Anderson, J. R. (2005). Practice and forgetting effects on vocabulary memory: An activation-based model of the spacing effect. *Cognitive Science*, 29, 559-586.
- Pawley, A. & Syder, F. H. (1983). Two puzzles for linguistic theory: nativelike selection and nativelike fluency. In J. C. Richards, & R. W. Schmidt (Eds.), *Language and communication* (191-225). London: Longman.
- Robinson, P. & Ha, M. (1993). Instance theory and second language rule learning under explicit conditions. *Studies in Second Language Acquisition*, 13, 413-38.

- Rohrer, D. & Taylor, K. (2006). The effects of overlearning and distributed practice on the retention of mathematics knowledge. *Applied Cognitive Psychology, 20*, 1209–1224.
- Rosenbloom, P. & Newell, A. (1987). Learning by chunking: A production system model of practice. In D. Klahr, P. Langley, & R. Neches (Eds.), *Production system models of learning and development* (pp. 221-286). Cambridge, MA: MIT Press.
- Schmidt, R. (1992). Psychological mechanisms underlying second language fluency. *Studies in Second Language Acquisition, 14*, 357-385.
- Schmitt, N., Dörnyei, Z., Adolphs, S. & Durow, V. (2004). Knowledge and acquisition of formulaic sequences: A longitudinal study. In N. Schmitt (Ed.), *Formulaic sequences* (pp. 55-87). Amsterdam, Philadelphia: John Benjamins.
- Schneider, W. & Chein, J. M. (2003). Controlled and automatic processing: From mechanisms to biology. *Cognitive Science, 27*, 525-559.
- Schumann, J. H. (1986). Research on the acculturation model for second language acquisition. *Journal of Multilingual and Multicultural Development, 7*, 379-392.
- Serrano, R. (2007). *Time distribution and the acquisition of English as a foreign language*. Unpublished Ph.D. Dissertation. University of Barcelona.
- Serrano, R. (2011a). The time factor in EFL practice. *Language Learning, 61*, 117-145.
- Serrano, R. (2011b). The effect of program type and proficiency level on learners' written production. *Revista Española de Lingüística Aplicada (RESLA), 24*, 211-226.
- Serrano, R. (2012). Is time concentration good or bad for learning? Reflecting on the results from the cognitive psychology and from the SLA literature. In C. Muñoz

- (Ed.), *Intensive exposure experiences in second language learning* (pp. 3-22). Clevedon: Multilingual Matters.
- Serrano, R; Tragant, E.; Llanes, À. (2012). A longitudinal analysis of the effects of one year abroad. *Canadian Modern Language Review*, 68, 138-163.
- Siyanova, A., & Schmitt, N. (2007). Native and nonnative use of multi-word vs. one-word verbs. *International Review of Applied Linguistics*, 45, 109-139.
- Skehan, P. (1998). *A cognitive approach to language learning*. Oxford: OUP.
- Smiskova, H., & Verspoor, M. (in press). Development of Chunks in Dutch L2 Learners of English. In Evers-Vermeul, J., Rasier, L. & E. Tribushinina (Eds.), *Usage-based approaches to language acquisition and language teaching*. Berlin: Mouton de Gruyter.
- Spada, N. & Lightbown, P. M. (1989). Intensive ESL programs in Quebec primary schools. *TESL Canada Journal*, 7, 11-32.
- SPSS 16.0 for Windows (Computer software) (2007). SPSS Inc.
- Stengers H., Boers F., Housen A., & Eyckmans J. (2011). Formulaic sequences and L2 oral proficiency: Does the type of target language influence the association? *International Review of Applied Linguistics in Language Teaching (IRAL)*, 49, 321-343.
- Stubbs, M. (1995). Collocations and semantic profiles: On the cause of the trouble with quantitative methods. *Functions of Language*, 2, 1–33.
- Tavakoli, P. & Foster, P. (2008). Task design and second language performance. The effect of narrative type on learner output. *Language Learning*, 58, 439-473.

- Toppino T. C. & Bloom, L. C. (2002). The spacing effect, free recall, and two-process theory: A closer look. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 28, 437-444.
- Toppino, T. C., Hara, Y., & Hackman, J. (2002). The spacing effect in the free recall of homogeneous lists: Present and accounted for. *Memory & Cognition*, 30, 601-606.
- Weinert, R. (1995). The role of formulaic language in second language acquisition: A review. *Applied Linguistics*, 16, 185-205,
- White, J. & Turner, C. E. (2005). Comparing children's oral ability in two ESL programs. *The Canadian Modern Language Review*, 61, 491-517.
- Wray, A. (1999). Formulaic language in learners and native speakers'. *Language Teaching*, 32, 213-231.
- Wray, A. (2000). Formulaic sequences in second language teaching: Principle and practice. *Applied Linguistics*, 21, 463-489.
- Wray, A. (2002). *Formulaic language and the lexicon*. Cambridge: CUP.
- Yorio, C. A. (1989). Idiomaticity as an indicator of second language proficiency. In K. Hyltenstam, & L. K. Obler (Eds.), *Bilingualism across the Lifespan* (pp. 55-72). Cambridge: CUP.

Tables

Table 1. EFL Participants

	Elementary	Intermediate	Advanced	TOTAL
	A1	B1	B2/C1	
INTENSIVE	14	22	22	58
REGULAR	21	22	23	66
TOTAL	35	44	45	124

Table 2. Descriptive statistics beginners

	Program	N	Mean	SD	SE
Types Posttest	Intensive	14	5.56	2.88	.77
	Regular	21	3.86	2.81	.61
Tokens Posttest	Intensive	14	6.72	2.77	.74
	Regular	21	4.13	3.14	.68

SD= Standard Deviation; SE=Standard Error

Table 3. Descriptive statistics intermediate learners

	Program	N	Mean	SD	SE
Types Pretest	Intensive	22	5.53	2.18	.46
	Regular	22	4.93	2.08	.44
Types Posttest	Intensive	22	6.22	2.08	.44
	Regular	22	4.57	2.74	.58
Tokens Pretest	Intensive	22	5.94	2.37	.50
	Regular	22	5.61	2.47	.52
Tokens Posttest	Intensive	22	7.19	2.79	.59
	Regular	22	5.27	3.25	.69

SD= Standard Deviation; SE=Standard Error

Table 4. Descriptive statistics advanced learners

	Program	N	Mean	SD	SE
Types Pretest	Intensive	22	5.66	1.83	.39
	Regular	23	6.32	2.46	.51
Types Posttest	Intensive	22	5.96	2.18	.46
	Regular	23	6.89	2.38	.49
Tokens Pretest	Intensive	22	6.44	1.92	.41
	Regular	23	7.19	2.80	.58
Tokens Posttest	Intensive	22	6.40	2.42	.51
	Regular	23	7.90	2.90	.60

SD= Standard Deviation; SE=Standard Error

Table 5. Descriptive statistics advanced EFL intensive vs. NES

	L1	N	Mean	SD	SE
Types	NES	12	8.92	1.70	.49
Tokens	NES	12	10.21	2.35	.67
Types Pretest	Adv. EFL Intensive	22	5.66	1.83	.39
Types Posttest	Adv. EFL Intensive	22	5.96	2.18	.46
Tokens Pretest	Adv. EFL Intensive	22	6.44	1.92	.41
Tokens Posttest	Adv. EFL Intensive	22	6.40	2.42	.51

SD= Standard Deviation; SE=Standard Error

Table 6: Descriptive statistics advanced EFL regular vs. NES

	L1	N	Mean	SD	SE
Types	NES	12	8.92	1.70	.49
Tokens	NES	12	10.21	2.35	.67
Types Pretest	Adv. EFL Regular	23	6.32	2.46	.51
Types Posttest	Adv. EFL Regular	23	6.89	2.38	.49
Tokens Pretest	Adv. EFL Regular	23	7.19	2.80	.58
Tokens Posttest	Adv. EFL Regular	23	7.90	2.90	.60

SD= Standard Deviation; SE=Standard Error