



**HAL**  
open science

## Collocational Development during a Stay Abroad

Amanda Edmonds, Aarnes Gudmestad

► **To cite this version:**

Amanda Edmonds, Aarnes Gudmestad. Collocational Development during a Stay Abroad. *Languages*, 2021, 6 (1), pp.12. <10.3390/languages6010012>. <hal-03129343>

**HAL Id: hal-03129343**

**<https://univ-montpellier3-paul-valery.hal.science/hal-03129343v1>**

Submitted on 25 Jan 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



HAL Authorization

Article

# Collocational Development during a Stay Abroad

Amanda Edmonds <sup>1,\*</sup>  and Aarnes Gudmestad <sup>2</sup>

<sup>1</sup> Département d'Études Anglophones, Université Paul-Valéry Montpellier 3, route de Mende, 34199 Montpellier, France

<sup>2</sup> Department of Modern and Classical Languages and Literatures, Virginia Polytechnic Institute and State University, Blacksburg, VA 24061, USA; agudmest@vt.edu

\* Correspondence: amanda.edmonds@univ-montp3.fr

**Abstract:** The purpose of the current study was to explore if and how additional-language learners may show changes in phraseological patterns over the course of a stay in a target-language environment. In particular, we focused on noun+adjective combinations produced by a group of additional-language speakers of French at three points in time, spanning 21 months and including an academic year in France. We extracted each combination from a longitudinal corpus and determined frequency counts and two strength-of-association measures (Mutual information [MI] score and Log Dice) for each combination. Separate analyses were conducted for frequency and the strength-of-association measures, revealing that phraseological patterns are significantly predicted by adjective position in the case of all three measures, and that MI scores showed significant change over time. We interpret the results in light of past research that has reported contradictory findings concerning change in phraseological patterns following an immersion experience.

**Keywords:** collocation; frequency; MI score; Log Dice; French; stay abroad



**Citation:** Edmonds, Amanda, and Aarnes Gudmestad. 2021. Collocational Development during a Stay Abroad. *Languages* 6: 12. <https://doi.org/10.3390/languages6010012>

Received: 28 October 2020  
Accepted: 6 January 2021  
Published: 12 January 2021

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

The last several decades have seen a wealth of research investigating the impact of a stay abroad on participants' linguistic development (see Kinginger 2009; Llanes 2011, for overviews). Results overwhelmingly point to improved oral fluency (Huensch and Tracy-Ventura 2017), greater pragmatic appropriateness (Shively 2011), and development in the realm of sociolinguistic competence (Howard 2012) after a stay abroad. With respect to vocabulary, several scholars have reported gains in both the number of words known after a stay abroad (e.g., Briggs 2015; Ife et al. 2000) and in the quality of known words. For example, Crossley et al. (2010) demonstrated that stay-abroad learners of English produced more instances of polysemy after four months in the United States, whereas Crossley et al. (2016) showed, among other things, that the same learners used less concrete (and, therefore, more abstract) lexical items after a year abroad.

Certain researchers have also turned their attention to changes at the level of phraseology, meaning changes in the lexical combinations used by additional-language speakers. This interest in the development of phraseology over the course of a stay in a target-language environment has been motivated in part by the rise of usage-based approaches in the field of second language acquisition (SLA). According to such approaches, a large part of language learning involves detecting patterns in language input. If that input is not massive (or varied) enough, the learner may be at a disadvantage in extracting relevant usage patterns: "Language learning is essentially a sampling problem—the learner has to estimate the native norms from a sample of usage experience" (Ellis et al. 2015, p. 364). Ellis et al. go on to suggest that this sampling problem may create particular challenges for the acquisition of phraseological patterns because "[m]any of the forms required for idiomatic use are of relatively low frequency, and the learner thus needs a large input sample just to encounter them." Following this logic, in comparison with the language classroom, the

stay-abroad experience may provide more and better quality input for language learning in general and for phraseological learning in particular. However, the empirical research that has sought to determine how learners' phraseological competence may change as a result of an immersion experience shows a variety of contrasting patterns. In particular, whereas some studies point to development over the course of a stay abroad, with learners producing more word combinations that are typical of the target language at the end of their stay, others show no change or even a move away from phraseological patterns present in the ambient input.

In the current article, we examine if and how phraseological use develops over the course of a stay abroad. In particular, we focus on how the frequency and collocational strength of noun (N) + adjective (Adj) combinations produced by learners evolve over time. Four characteristics of this study allow us to contribute novel insights to this line of inquiry. First, we report on longitudinal data that span a period of 21 months collected at three different points in time: before the participants went abroad, at the end of an academic year spent in the target-language community, and 8 months after their return to their home country. This wide timespan allows us to both explore the impact of a longer stay abroad and investigate potential changes after the return home. Second, the 28 participants we analyzed are learners of additional-language French, making this study one of the few that has investigated a target language other than English. Third, whereas most previous research has focused on the Mutual Information (MI) score to quantify strength of collocational association, recent research by [Gablasova et al. \(2017\)](#) has highlighted the limits of this measure. For this reason, in addition to the MI score, we use a new measure of collocational strength, Log Dice. Finally, following recent calls for methodological change in the field of learner corpus research (see [Gries 2015](#); [Siyanova-Chanturia and Spina 2020](#)), we opted to use mixed-effects linear regression analyses to analyze our written corpus data. Such analytic tools allow us to explore the impact that numerous independent variables may simultaneously exert on changes in N+Adj frequency and strength, all the while accommodating the variability that characterizes the population from which our participants are drawn.

## 2. Literature Review

### 2.1. *Phraseological Development: Frequency and Collocational Strength*

Research on phraseology is characterized by two approaches, which define phraseology in distinct ways (see [Granger and Paquot 2008](#)). On the one hand, researchers adopting what Granger and Paquot call "phraseological" approaches use linguistic criteria in order to identify (more or less semantically and/or syntactically) transparent multi-word units. On the other, scholars working within distributional approaches consider phraseology in a clearly bottom-up manner, generally using theory-independent measures, such as frequency, in order to identify patterns of co-occurrence in large corpora. In the current article, we adopt a distributional approach in our study of N+Adj combinations in a written corpus. As is the case in most distributional approaches, we will rely on both measures of frequency and collocational strength (also referred to as strength-of-association measures) in our analysis. We begin by detailing how these measures are applied and what caveats need to be kept in mind when using them. We end this sub-section by reviewing what the use of these measures has revealed about phraseological competence in an additional language.

Distributional approaches to phraseology are interested in identifying recurrent patterns of word use, which is accomplished using a variety of different measures. The first of these measures is raw frequency, where researchers generally establish a cut-off (e.g., 10 occurrences per 1 million words) and identify all (2-, 3-, 4-, etc.) word combinations that occur more often than the cut-off in question. Approaches relying solely on frequency often claim to be researching "lexical bundles", which focus on such highly frequent combinations as *of the*, *in the*, and *it is* (see [Granger and Bestgen 2014](#), p. 235). Whereas high frequency may reflect phraseological status for a given speech community, researchers

have often turned to measures that tap into the strength of association<sup>1</sup> between words in order to complement a purely frequency-based account. Strength-of-association measures attempt to quantify the level of attraction between two words, thereby determining the exclusivity of the co-occurrence relationship. In other words, such measures reflect “the relationship between the number of times when [two words] are seen together as opposed to the number of times when they are seen separately in the corpus” (Gablasova et al. 2017, p. 160). Of the many different manners in which this relationship can be quantified, the MI score is undoubtedly the most commonly used in SLA research (Manning and Schütze 1999). However, as pointed out by numerous scholars, all co-occurrence measures have their limitations. In the case of the MI score, the combinations that receive high scores tend to involve low frequency words, which led Gablasova et al. to observe that “[h]ighlighting *rare exclusivity* is thus the main practical effect of the mathematical expression of the MI-score” (p. 164, our emphasis). These researchers go on to state that this equation thus “not only measure[s] collocational knowledge (preferences in word combinations), but also lexical knowledge of infrequent lexical items” (p. 172). Whereas most previous research has considered an MI score of 3 or greater as an indication of phraseological status, one important limitation of this measure is that it does not have a theoretical minimum or maximum score. The formula for calculating the MI score is given in (1).

$$\text{MI score} = \log_2 (\text{frequency word}_1 \text{word}_2 / (\text{frequency word}_1 \times \text{frequency word}_2) / \text{number of total words in corpus}) \quad (1)$$

Log Dice, the second strength-of-association measure that we used in the current project, has received little attention in the SLA literature. According to Gablasova et al. (2017, pp. 164–65), both Log Dice and the MI score detect exclusivity, but they argue that Log Dice has several advantages over the MI score: (a) it has a fixed maximum value of 14, (b) it does not overly favor highly infrequent items, which means that it detects exclusivity (and not uniquely rare exclusivity) and (c) the measure does not include expected frequency in its equation, making it more appropriate for very large corpora, like the one that was used in the current study. The equation for Log Dice, taken from Gablasova et al., is presented in (2):

$$\text{Log Dice} = 14 + \log_2 (2 \times \text{frequency word}_1 \text{word}_2) / (\text{frequency word}_1 + \text{frequency word}_2) \quad (2)$$

Numerous studies have investigated how collocational frequency and strength in target-language input may influence how additional-language learners process, judge, and produce such combinations. Ellis et al. (2008) constitutes one widely cited psycholinguistic study. The researchers set out to explore how frequency and strength of association influenced processing patterns among native and additional-language speakers of English. In a series of three online experiments, the investigators explored how participants processed 108 academic formulas that varied according to their length (3, 4 or 5 words), their overall frequency, and their MI score. Taken together, the results pointed to the psychological validity of the phraseological sequences for both groups of speakers, insofar as both groups showed online sensitivity to different distributional profiles. However, if both groups showed significant sensitivity, their processing profiles revealed differences. Ellis et al. report an effect of overall frequency in additional-language processing and an effect of MI score on native processing, such that the additional-language speakers showed facilitated (i.e., faster) processing on high frequency combinations and native speakers showed evidence of facilitated processing on combinations with high MI scores. This finding was interpreted by Ellis et al. in the following way. They suggested that the effect of frequency was less strong in native processing because such speakers had already benefited from large amounts of exposure to their native language, meaning that the frequency differences among the target strings had essentially leveled out. Non-native speakers, however, had had much less exposure to the target language, meaning that this leveling out had not (yet)

<sup>1</sup> Certain researchers have also turned to measures of dispersion and directionality in their investigations of additional-language phraseology (see Ellis et al. 2016; Siyanova-Chanturia and Spina 2020).

occurred, and more frequent combinations continued to be processed more quickly than less frequent ones. As for the finding that higher MI scores were significant predictors of faster processing among native speakers, the authors suggested that “native speakers are attuned to these constructions as packaged wholes” (p. 391). This attunement then was presumably not (yet) in place among additional-language learners.

Since the publication of [Ellis et al. \(2008\)](#), numerous researchers interested in the acquisition of phraseological competence in an additional language have investigated whether the frequency and the collocational strength of a word combination may impact both how learners judge the acceptability of combinations (e.g., [Edmonds and Gudmestad 2014](#)) and which combinations are actually produced by learners. With respect to production, there is evidence pointing to the fact that learners tend to favor frequent combinations over those characterized by a high MI score. For example, in their 2009 study, [Durrant and Schmitt](#) compared academic assignments written by advanced learners and by English native speakers. Word pairs involving a noun and a premodifier (Adj+N or N+N) were extracted from all essays and the researchers then determined how often each pair occurred in a reference corpus. They also used the reference corpus to calculate the MI score for each combination under study. The results revealed that

[a]dvanced non-native phraseology differs from that of natives not because it avoids formulaic language altogether but because it overuses high-frequency collocations and underuses the lower-frequency, but strongly-associated, pairs characterized by high mutual information scores. Since the latter sort appear (intuitively, and on the psycholinguistic evidence presented by [Ellis et al.](#)) to be highly salient for native speakers, their absence may be what creates the feeling that non-native writing lacks ‘idiomaticity’. ([Durrant and Schmitt 2009](#), p. 175)

This finding has subsequently been explored in several other studies, with [Lorenz \(1999\)](#), [Forsberg \(2010\)](#), and [Granger and Bestgen \(2014\)](#) providing corroborating evidence of a greater reliance on high-frequency combinations (as opposed to strongly associated ones) in additional-language production. [Siyanova and Schmitt \(2008\)](#), on the other hand, found no such difference in their comparison of Adj+N collocations produced by native speakers and Russian-speaking learners of English in written texts. As we will see in the following sub-section, the influence of frequency and collocational strength has also been addressed in research that has explored how learners’ phraseological competence develops over the course of an immersion experience.

## 2.2. *Phraseological Development during an Immersion Experience*

According to [Ellis et al. \(2015, p. 364\)](#), research suggests that the learning context—and, in particular, foreign-language learning contexts versus immersion experiences—may significantly influence the learning of phraseological combinations. They explain this with reference to usage-based approaches to language and the (potential) differences in input available to learners in the two contexts: “Learning the usages that are normal or unmarked from those that are unnatural or marked requires a huge amount of immersion in the speech community.” Spending time in a target-language environment presumably provides the learner with a larger and perhaps richer sample from which to extract phraseological patterns. Several researchers have set out to explore if and how phraseological use changes over the course of a stay abroad, and in this article, we limit our review to those studies that have used longitudinal data ([Bestgen and Granger 2014](#); [Crossley and Salsbury 2011](#); [Li and Schmitt 2009, 2010](#); [Siyanova-Chanturia 2015](#); [Siyanova-Chanturia and Spina 2020](#); [Yoon 2016](#)).

The seven studies reviewed present a range of results in terms of change over time. Beginning with the factor of frequency, which was investigated by four studies, we note two distinct interpretations of this notion. In the case study reported by [Li and Schmitt \(2009\)](#), the researchers examined how frequently lexical phrases were produced in one learner’s own writings, revealing a non-linear frequency pattern whereby the greatest density of lexical phrases was found in the second of nine writing assignments. The remaining three

studies sought to determine whether learners use word combinations that are frequent in the target language more often at the end of their stay abroad. Crossley and Salsbury (2011) reported on bigrams produced by six learners of English enrolled in an intensive language program, and they showed that between the first and fourth quarter of study, three of the six learners produced significantly more bigrams that were also attested in a native corpus. In Siyanova-Chanturia's single-authored and collaborative research, the participants under study were Chinese learners of Italian, who were completing an intensive language program in Italy. Both studies examined N+Adj combinations in written assignments completed by either beginner-level learners (Siyanova-Chanturia 2015) or by beginner-, elementary-, and intermediate-level learners (Siyanova-Chanturia and Spina 2020). In the first study, Siyanova-Chanturia found that the learners used more high frequency N+Adj combinations at the end of a 21-week language program than at the beginning. In the second study, which involved a much larger cohort of participants ( $n = 175$ ), a different picture emerged. The frequency of N+Adj combinations was found overall to decrease between the beginning and the end of the intensive Italian class, although a significant interaction between learner proficiency and time nuanced this finding, revealing that the beginner-level learners were more likely to produce less frequent combinations after six months abroad. Overall, this research presents a divergent set of results: In some cases, no change was observed (half of Crossley and Salsbury's learners), in others, learners were seen to produce more frequent combinations at the end of a stay abroad (the other half of Crossley and Salsbury's learners; Siyanova-Chanturia), and in at least one example, learners produced less frequent combinations after their stay abroad (beginner learners in Siyanova-Chanturia and Spina).

Five longitudinal studies examined whether word combinations used by learners evolved in terms of their collocational strength over the course of a stay abroad (Bestgen and Granger 2014; Li and Schmitt 2010; Siyanova-Chanturia 2015; Siyanova-Chanturia and Spina 2020; Yoon 2016). In these studies, the researchers began by extracting certain word combinations from learners' written productions: all bigrams (Bestgen and Granger), N+Adj combinations (Li and Schmitt; Siyanova-Chanturia; Siyanova-Chanturia and Spina), or verb+N combinations (Yoon). The researchers then determined the strength of association for each combination (MI score)<sup>2</sup> using a reference corpus. In group-level analyses, no significant changes in MI scores were found by Bestgen and Granger, Yoon, and Li and Schmitt when comparing written productions from the beginning and the end of a period abroad. However, Li and Schmitt also examined the individual performance of their four Chinese learners of English and reported that one participant showed greater use of strongly associated N+Adj combinations at the end of the academic year abroad. Positive change was also reported by Siyanova-Chanturia, who found that a group of beginner-level learners produced significantly more strongly associated N+Adj combinations at the end of a 21-week intensive Italian course. In their larger project, however, Siyanova-Chanturia and Spina did not find that MI scores associated with N+Adj combinations underwent significant change over time. However, they did find an effect of proficiency, whereby the elementary- and intermediate-level learners had a greater likelihood of producing combinations that were less strongly associated (and, thus, had a lower MI score) than the beginner-level learners.

### 3. The Current Study

Past research has reported that additional-language users tend to hone in on high frequency phraseological patterns, whereas they “produce fewer of those collocations that are less frequent, even though these are strongly linked” (Schmitt et al. 2019, p. 5). Moreover, although Ellis et al. (2015) suggest that the stay-abroad experience may be particularly facilitative for the detection of (phraseological) patterns in the input, attempts to explore

<sup>2</sup> Although two studies (Bestgen and Granger; Li and Schmitt) also reported on *t*-scores (another strength of association measure) for all combinations, this aspect of their analysis will not be reviewed here.

phraseological development as a result of a stay abroad have reported contradictory results. Against this backdrop, the purpose of the current study was to further explore if and how additional-language learners may show changes in phraseological patterns over the course of a stay in a target-language environment, as well as after the return to their home country. The following research questions were addressed: How do the frequency and collocational strength of N+Adj combinations used in written essays evolve both after an academic year spent in a target-language community and 8 months after the return home? What factors significantly predict frequency and collocational strength of N+Adj combinations?

### 3.1. Method

#### 3.1.1. The LANGSNAP Corpus and Participants

Data analyzed for this project came from the LANGSNAP corpus.<sup>3</sup> This freely available corpus is the result of a longitudinal project carried out by a research team based at the University of Southampton (see Mitchell et al. 2017). The LANGSNAP team followed 29 additional-language learners of French enrolled in a French degree program in the United Kingdom between May 2011 and February 2013, before, during, and after an academic year they spent in France (a similar group of learners of Spanish was also followed). Over the course of this 21-month project, the participants met with a researcher on six occasions: once before their stay abroad, three times during the academic year in France, and twice after their return to the United Kingdom. For the current project, we have limited our focus to the first data-collection meeting, which we refer to as pre-stay, the fourth data-collection meeting, which took place at the end of the learners' stay abroad approximately one year after pre-stay data collection (i.e., in-stay), and finally the very last data-collection period, which involved meeting with participants approximately eight months after their return to the United Kingdom (i.e., post-stay). At each data-collection meeting, participants were asked to write an approximately 200-word argumentative essay in French. For this task, the same prompt was used at pre-stay and in-stay (see 3a), whereas a different prompt was used for post-stay (3b):

3. a. *Pensez-vous que les couples homosexuels ont le droit de se marier et d'adopter des enfants?*  
'Do you think that homosexual couples have the right to get married and adopt children?'
- b. *Pensez-vous que, de manière à inciter les gens à manger sainement, on devrait taxer les boissons sucrées et les aliments gras?*  
'Do you think that in order to encourage people to eat in a healthy manner, sugary beverages and fatty foods should be taxed?'

For this project, we analyzed the pre-stay, in-stay, and post-stay written productions from 28 of the 29 additional-language French speakers in the LANGSNAP database; participant 122 was excluded because she did not contribute data at in-stay. Three of the 28 learners were men, and 25 were women. All participants completed an elicited imitation (EI) test at the outset of the project, whose aim was to provide a measure of overall initial proficiency. For this test, participants were asked to repeat out loud 30 sentences in French, ranging in length from 7 to 19 syllables, which they heard pronounced by a French speaker. A score between 0 and 4 points on the basis of accuracy and completeness was awarded for each sentence (see Tracy-Ventura et al. 2014 for details). Out of a possible 120 points, these participants scored between 36 and 97 ( $M = 62.57$ ,  $SD = 18.21$ ). Participants were on average 20.04 years old ( $SD = 0.88$ , range 19-23) and reported having studied French between six and 20 years ( $M = 10.64$ ,  $SD = 3.07$ ). Although the majority of participants noted that English was their first language, the group also contains one first-language speaker of Finnish and one first-language speaker of Spanish. Additionally, three speakers reported having had some access to French during their childhood (one of these speakers reported that he had been learning French since birth, that is, for 20 years). Finally, participants provided information as to their principal occupation during their stay abroad: These

<sup>3</sup> <https://slabank.talkbank.org/access/>.

individuals were either teaching assistants in French schools ( $n = 14$ ), exchange students enrolled at a French university ( $n = 8$ ), or workplace interns ( $n = 6$ ).

### 3.1.2. The Dataset under Study

The dataset contains 84 essays, for a total of 17,292 words. Each essay was read in order to identify all instances in which a learner used an adjective to modify a noun within the same noun phrase. In this study, we adopted a broad understanding of adjective, which means that we included many of what [Durrant and Schmitt \(2009, p. 166\)](#) called semi-determiners. For example, *même* ‘same’, *tout* ‘all’, *certain* ‘certain’, *tel* ‘such’, and both cardinal and ordinal numbers were included when they were used adjectivally. Examples from our dataset are provided in (4).

4. *les mêmes droits* ‘the same rights’  
*tout le monde* ‘everyone’  
*certains produits* ‘certain products’  
*une telle mesure* ‘such a measure’  
*deux parents* ‘two parents’  
*vingtième siècle* ‘twentieth century’

Whereas most attributive adjectives in French follow the noun they modify, a subset of adjectives tends to precede the noun. These include the examples given in (4), but also adjectives such as *bon* ‘good’, *beau* ‘beautiful’, *grand* ‘big’, and so forth. Unlike [Siyanova-Chanturia \(2015\)](#) and [Siyanova-Chanturia and Spina \(2020\)](#), we did not limit our investigation to one of these two orders, and in what follows, we use the N+Adj notation to cover both. A total of 1094 N+Adj combinations were identified in this corpus. Two hundred and thirty-seven tokens were removed, as they corresponded to combinations contained in the essay prompts (see 3a-b: *couples homosexuels*, *boissons sucrées*, *aliments gras*). After removing these tokens, our final corpus contains 857 N+Adj occurrences, of which 506 showed the adjective in postnominal and 351 in prenominal position. Details are provided in [Table 1](#).

**Table 1.** Details on the N+Adj dataset.

Data Collection	Total Words	N+Adj Occurrences	
		Tokens	Types
Pre-stay	5898	278	194
In-stay	5672	296	206
Post-stay	5722	284	204
TOTAL	17,292	857	531

### 3.1.3. The Reference Corpus and Data Coding

Once all N+Adj combinations had been extracted, each occurrence was coded with respect to its frequency and collocational strength (dependent variables in our analysis), as well as with respect to several independent factors that were hypothesized to impact the use of N+Adj combinations. Beginning with the dependent variables, a reference corpus was used to determine the frequency and collocational strength of each combination. The frTenTen12 WaCky corpus ([Baroni et al. 2009](#)) is a large (9,889,689,889 words), web-crawled corpus compiled in 2012, which corresponds to the period during which the LANGSNAP learners were in France. As such, the corpus should provide a relatively close approximation to written online input to which the participants may have been exposed. We searched the frTenTen12 WaCky corpus for the lemmatized frequency for each noun, adjective, and N+Adj combination. In conducting these searches, we always specified part of speech. For the N+Adj combination searches, two additional details are relevant. First, we searched the combination only in the order produced by the learner. Although certain adjectives may appear in both pre- and postnominal positions, a change in position generally involves a change in meaning (see [Anderson 2008](#)), meaning that the two orders cannot necessarily

be considered two versions of the same collocation. This decision means that the same adjective and noun used in different orders were treated as different combinations. In our dataset, there is just one example of this, involving the noun *opinion* “opinion” and the adjective *fort* “strong” (see 5).

5. a. *mais d’un autre côté il existe les gens avec les fortes opinions* (Participant 102, in-stay)  
 ‘but on the other hand there exist people with strong opinions’
- b. *les autres livres pour avoir un opinion plus forte* (Participant 127, in-stay)  
 ‘the other books to have an opinion stronger’

Second, given that adjectives can be separated from the noun they modify by other words (e.g., 5b), we did not require strict adjacency in this analysis. After exploring the results returned when search windows of various sizes were used, we opted for allowing up to one word to separate the noun from the adjective, as larger search windows returned a high proportion of inappropriate hits. Inspection of the search results revealed that even with this small search window, certain inappropriate combinations were returned. For this reason, we further restricted our searches such that the intervening element could not be a verb (when the order was noun followed by adjective) or a preposition (when the order was adjective followed by noun).

Lemmatized frequency counts from the frTenTen12 WaCky corpus were used to calculate two measures of collocational strength for each N+Adj combination. Whereas the MI score gives greater weight to rare lexical items and, for this reason, has been claimed to reflect rare exclusivity, Log Dice has been argued to reflect collocational exclusivity without favoring low frequency words. Table 2 provides the 10 highest scoring combinations according to the three dependent variables. This table shows that although approximately half of each list overlaps with one of the others, the remaining combinations (those highlighted in grey) are unique to one measure.

Table 2. Highest scoring N+Adj combinations.

	Frequency		MI Score		Log Dice	
	Combination	Frequency	Combination	Score	Combination	Score
1	<i>tout monde</i> “all world”	1,700,987	<i>société contemporaine</i> “contemporary society”	17.95	<i>dessin animé</i> “animated drawing”	11.17
2	<i>même temps</i> “same time”	902,134	<i>boisson gazeuse</i> “carbonated drink”	14.37	<i>être humain</i> “human being”	10.64
3	<i>tout autre</i> “all other”	750,705	<i>vingt-et-unième siècle</i> “twenty-first century”	13.02	<i>tout monde</i> “all world”	10.54
4	<i>deux choses</i> “two things”	635,420	<i>dessin animé</i> “animated drawing”	12.42	<i>parti socialiste</i> “socialist party”	10.53
5	<i>tout jour</i> “all day”	482,863	<i>vingtième siècle</i> “twentieth century”	11.53	<i>même temps</i> “same time”	10.51
6	<i>toute façon</i> “all manner”	482,402	<i>tout autre</i> “all other”	11.31	<i>court terme</i> “short term”	10.17
7	<i>autre part</i> “other part”	468,992	<i>famille monoparentale</i> “monoparental family”	10.94	<i>premier ministre</i> “prime minister”	10.11
8	<i>tout deux</i> “all two”	427,727	<i>mission civilisatrice</i> “civilizing mission”	10.86	<i>chose importante</i> “important thing”	10.1
9	<i>dernières années</i> “last years”	410,868	<i>couple hétérosexuel</i> “heterosexual couple”	10.74	<i>boisson gazeuse</i> “carbonated drink”	10.00
10	<i>premier ministre</i> “prime minister”	374,419	<i>parti socialiste</i> “socialist party”	10.73	<i>haut niveau</i> “high level”	9.85

Note. Combinations in grey are unique to one measure.

The 857 occurrences were also coded with respect to six independent variables. These factors are presented in (6). We analyzed participant as a random effect in the analysis; the other five variables were fixed effects.

6.
  - a. Time: the occurrence was produced at pre-stay, in-stay, or post-stay
  - b. Years of French study: the number of years that the speaker reported having studied French
  - c. EI score: the EI score obtained by the speaker at pre-stay
  - d. Placement: the main activity (teaching assistant, workplace intern, student) in which the speaker was engaged while abroad
  - e. Adjective position: the adjective occurred either in prenominal or postnominal position
  - f. Participant: which participant produced the occurrence

Exploring whether time significantly impacted the frequency and/or strength of association of N+Adj combinations produced by additional-language speakers revealed whether there were changes both after the stay abroad and after the learners' return home. The variables years of French study and EI score both constituted ways to gauge the proficiency of the participants. As shown by [Siyanova-Chanturia and Spina \(2020\)](#), proficiency level may significantly influence phraseological choices and development. We have also included the variable of placement in this analysis. As the three different occupations in which the learners were engaged may result in different opportunities and types of input, we wanted to determine whether this variable significantly impacted the frequency or collocational strength of N+Adj combinations produced. One linguistic variable was explored in this study. Unlike previous work on N+Adj combinations, we chose to examine both prenominal and postnominal adjectives. We thus included the variable of adjective position in our analysis in order to explore if and how collocational frequency and strength varied as a function of adjective position. Finally, the variable of participant was included as a random effect, in order to account for variability among the participants.

#### 3.1.4. Data Analysis

We report on three linear mixed-effects models, one for lemmatized frequency, one for MI score, and one for Log Dice. To reduce skew in the frequency variable, the values were logarithmically transformed. For the three discrete independent variables investigated, one category of the variable was selected as a reference category. For example, for the variable time, the category pre-stay was the reference category, and our analysis looked for differences when comparing data produced at in-stay and at post-stay with data produced at pre-stay. The reference categories for the other two discrete variables were postnominal (adjective position) and teaching assistant (placement). EI score and years of French were continuous variables and thus had no reference category. Finally, participant was included in each model as a random effect in the form of a random intercept<sup>4</sup>. The three models were built using the package lme4 in RStudio ([RStudio Team 2020](#)). For each of the three models, the six independent factors presented in (6) were investigated. The creators of the lme4 package in R decided not to include *p* values for mixed-effects models because there is trouble estimating *p* values with these types of model structures. For this reason, we used a model comparison to arrive at the best-fit model, which means that we began with a model that contained only the random effect and then progressively added each independent factor. When the inclusion of a given factor significantly improved the fit of the model according to an ANOVA comparison, this factor was retained in the final model. Moreover, we checked for potential significant interactions between the factor time and any other significant factors, in order to examine how factors impacting N+Adj may have evolved over time. Finally, we assessed the fit of our models in two ways. First, marginal and conditional  $R^2$  were computed and, second, we calculated the Bayesian Information Criterion for each model. Marginal  $R^2$  for a model reflects how much of the overall variance is accounted for by the fixed effects, whereas conditional  $R^2$  provides an indication of the amount of variance explained by the whole model (fixed and random effects) The Bayesian

<sup>4</sup> We explored the possibility of including both varying intercepts and varying slopes for the random effect, but because of convergence problems, only varying intercepts were included in our final models.

Information Criterion compares the log-likelihoods of the final model with a null model (containing only the dependent variable). The model obtaining the lower score shows the better fit, with a difference of at least 10 reflecting strong evidence in favor of that model. In what follows, we first report descriptive statistics for the three dependent variables at pre-stay, in-stay, and post-stay. We then present the three linear mixed effects models.

### 3.2. Results

Table 3 provides an overview of the frequency and collocational strength for the N+Adj combinations produced by the 28 learners of French at pre-stay, in-stay, and post-stay.

**Table 3.** Descriptive statistics.

Time	Frequency		MI Score		Log Dice	
	<i>M</i> ( <i>SD</i> )	Range	<i>M</i> ( <i>SD</i> )	Range	<i>M</i> ( <i>SD</i> )	Range
Pre-stay	86,744 (299,814)	0–1,700,987	4.05 (3.16)	−3.57–17.95	4.71 (3.1)	−5.63–10.65
In-stay	121,660 (367,934)	0–1,700,987	4.39 (3.18)	−4.38–13.02	5.17 (3.07)	−5.63–10.65
Post-stay	99,993 (325,428)	0–1,700,987	4.83 (2.77)	−2.35–14.37	5.19 (3.28)	−4.67–11.17

The details concerning the final models for frequency, MI score, and Log Dice are presented in Table 4a,b, Table 5a,b, and Table 6a,b, respectively. These three models showed several similarities. First, neither of the two proficiency measures (years spent studying French or EI score) significantly predicted the frequency or the strength of association of N+Adj combinations used. The same can be said about the placement type while abroad, which was not found to be significant in any of the models. Furthermore, the inclusion of the factor time was found to significantly improve only one model: MI score. The details for this model are provided in Table 5a, where we see positive parameter estimates for both in-stay and post-stay essays. This indicates that, when compared to the combinations produced at pre-stay, this group of learners was more likely to make use of combinations that enjoyed a higher level of co-occurrence strength as indicated by the MI score at the end of their year abroad and/or after their return to the United Kingdom. Because this variable has three categories, the model results in Table 5a do not specify whether this significant change concerns only the pre- versus in-stay comparison, only the pre- versus post-stay comparison or both. To more precisely pinpoint where significant development occurred, we carried out three pairwise *t*-tests using the observations from each time point and the residual standard error associated with the fit model. Before conducting these tests, we checked to make sure that the scores at each time point were normally distributed. We moreover applied the Bonferroni correction to our significance level, which resulted in an alpha level of 0.0033 (0.01/3). The three comparisons revealed that MI scores did not significantly change between pre- and in-stay ( $p = 0.1656$ ) or between in- and post-stay ( $p = 0.0749$ ). A significant change was however revealed when comparing MI scores for combinations produced at pre- versus those produced as post-stay ( $p = 0.0018$ ). Figure 1 provides a visualization of the influence of time on co-occurrence strength, by mapping out the MI scores for each N+Adj combination at pre-, in- and post-stay.

**Table 4.** (a) Summary of final model for frequency. (b) Random effects for participant in the frequency model.

<b>(a) Summary of Final Model for Frequency.</b>				
<b>Factor</b>	<b>Estimate</b>	<b>SE</b>	<b>t Value</b>	<b>Confidence Intervals</b>
(Intercept)	3.121	0.051	61.78	[3.021, 3.222]
Adjective position				
Pre	1.338	0.076	17.462	[1.189, 1.487]
<b>(b) Random Effects for Participant in the Frequency Model.</b>				
<b>Participant</b>	<b>Random Intercept</b>	<b>Participant</b>	<b>Random Intercept</b>	
100	0.0061051	115	0.0047082	
101	0.0192619	116	−0.0026766	
102	−0.0345696	117	−0.0122701	
104	0.0343793	118	−0.0041386	
105	0.0318979	119	0.0156721	
106	−0.003412712	120	−0.0011016	
107	0.0146950	121	−0.0040971	
108	0.0045674	123	0.0052533	
109	−0.0297249	124	−0.0175949	
110	−0.0213394	125	0.0112086	
111	−0.0116328	126	−0.0211113	
112	0.0049613	127	−0.0620854	
113	0.0331438	128	0.0099693	
114	−0.0281526	129	0.0498071	

**Table 5.** (a) Summary of final model for MI score. (b) Random effects for participant in the MI score model.

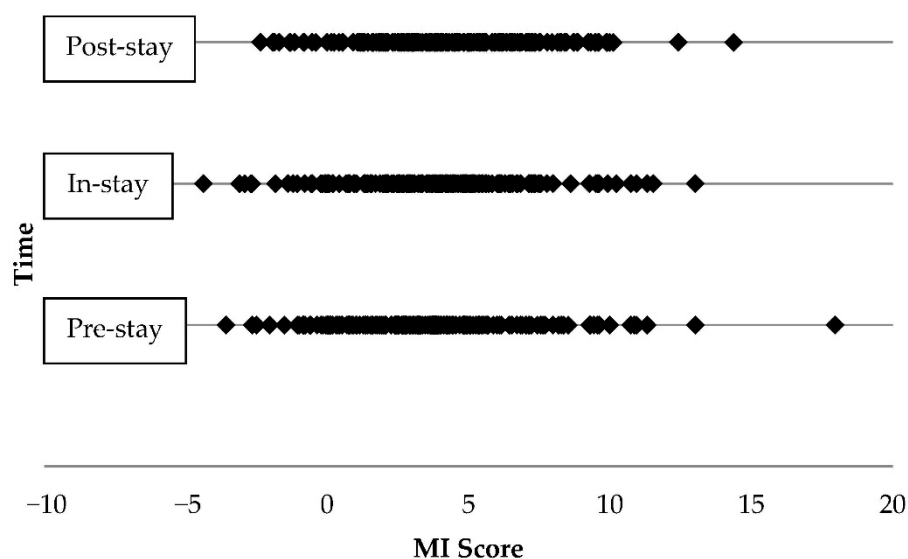
<b>(a) Summary of Final Model for MI Score.</b>				
<b>Factor</b>	<b>Estimate</b>	<b>SE</b>	<b>t Value</b>	<b>Confidence Intervals</b>
(Intercept)	4.5251	0.2045	22.13	[4.120, 4.926]
Time				
in-stay	0.3790	0.2492	1.521	[−0.111, 0.868]
post-stay	0.7433	0.2515	2.955	[0.250, 1.237]
Adjective position				
pre	−1.1936	0.2072	−5.760	[−1.601, −0.786]
<b>(b) Random Effects for Participant in the MI Score Model.</b>				
<b>Participant</b>	<b>Random Intercept</b>	<b>Participant</b>	<b>Random Intercept</b>	
100	0.0313194	115	−0.1334736	
101	0.1522636	116	−0.1056354	
102	−0.0303156	117	−0.0671213	
104	0.1499458	118	−0.0290557	
105	0.1185338	119	−0.0496603	
106	0.0085123	120	−0.1435863	
107	−0.1694525	121	−0.0488590	
108	0.2631510	123	−0.0233297	
109	−0.0674191	124	−0.0159254	
110	−0.0154777	125	−0.0248067	
111	0.1222740	126	−0.0924364	
112	−0.0474726	127	−0.0673566	
113	−0.0287572	128	0.0417407	
114	−0.1482262	129	0.4206269	

**Table 6.** (a) Summary of final model for Log Dice. (b) Random effects for participant in the Log Dice model.

(a) Summary of Final Model for Log Dice.				
Factor	Estimate	SE	t Value	Confidence Intervals
(Intercept)	4.2805	0.1391	30.774	[4.005, 4.553]
Adjective position pre	1.8191	0.2103	8.651	[1.405, 2.230]

(b) Random Effects for Participant in the Log Dice Model.			
Participant	Random Intercept	Participant	Random Intercept
100	0.0173211	115	-0.0123437
101	0.0582294	116	-0.0267567
102	-0.0776156	117	-0.0189802
104	0.1115772	118	0.0181944
105	0.0802951	119	0.0162897
106	-0.0061611	120	-0.0307761
107	-0.0166828	121	-0.0031515
108	0.0363880	123	0.0126978
109	-0.0404584	124	-0.0374099
110	-0.0533816	125	-0.0068696
111	-0.0220314	126	-0.0583006
112	-0.0174288	127	-0.0995125
113	0.0538182	128	0.0156374
114	-0.0453277	129	0.1527400



**Figure 1.** Fixed effect plot for the factor time (MI score model).

The final variable that we investigated was adjective position, and this variable proved to be significant in each of the three analyses, but with different patterns. In the case of frequency (Table 4a) and Log Dice (Table 6a), prenominal adjectives were significantly associated with higher frequency and with higher Log Dice scores, as is visible in the positive parameter estimates. The opposite was the case for the MI score analysis (Table 5a): Prenominal adjectives had a greater likelihood to result in combinations with lower MI scores than postnominal adjectives.

Finally, we explored the quality of our models. First, the calculation of  $R^2$  for the three models revealed that the inclusion of the random effect for participant always improved the fit of the model (i.e., conditional  $R^2$  is always greater than marginal  $R^2$ ): overall frequency of N+Adj combinations (marginal  $R^2 = 0.266039$ , conditional  $R^2 = 0.269053$ ), MI score (marginal  $R^2 = 0.047712$ , conditional  $R^2 = 0.056071$ ), and Log Dice (marginal  $R^2 = 0.080451$ ,

conditional  $R^2 = 0.082617$ ). Second, the Bayesian Information Criterion indicated that all three final models offered a better fit of the data than the null model: frequency (final model: 2598.3, null model: 2899.1), MI score (final model: 4343.7, null model: 4360.5), and Log Dice (final model: 4355.6, null model: 4413.8).

#### 4. Discussion

In commenting on research focused on language production, [Arnon and Snider \(2010, p. 67\)](#) observe that “findings show that language users are sensitive to detailed distributional information on many levels of linguistic analysis.” Although their article is focused on production in one’s native language, similar claims have been made for additional-language speakers (e.g., [Ellis et al. 2015, 2016](#)). However, the sensibility of additional-language speakers appears to differ in certain respects from that of native speakers. In particular, research has reported that additional-language learners show greater sensitivity to overall frequency in the input than to collocational strength. This sensitivity has been argued to be reflected in their overuse of high frequency word combinations, to the detriment of strongly associated ones. How a stay in a target-language community may reinforce or alter these tendencies has been explored by several researchers, who have tracked word combinations produced by learners over time to explore how they change (or not) in frequency and collocational strength. Overall, these studies have reported contradictory results. As we will see in what follows, the results from the current study contribute to these findings.

How do the frequency and collocational strength of N+Adj combinations used in written essays evolve both after an academic year spent in a target-language community and 8 months after the return home? Following [Ellis et al. \(2015\)](#), it is reasonable to expect that an immersion experience may facilitate phraseological learning, given that large amounts of input are necessary in order to encounter many phraseological patterns and that a stay abroad has the potential to provide such input. However, despite having spent an academic year in France, we noted little significant change with respect to the N+Adj combinations produced by the LANGSNAP learners. Beginning with overall frequency, the results from our linear mixed-effects analysis show that frequency was not significantly influenced by the factor time. Expressed differently, the 28 additional-language learners of French did not tend to use N+Adj combinations in the in-stay and the post-stay essays that were significantly more (or less) frequent than those used at pre-stay. While this result may reflect lack of sensitivity to overall frequency in the input, it may also be the case that these learners were already using high frequency N+Adj combinations and that this tendency was simply maintained during the stay abroad. Indeed, what little past research has shown change in frequency of phraseological units during a stay abroad has generally focused on lower-level learners (i.e., [Siyanova-Chanturia 2015](#)).

Turning to collocational strength, the two measures used—MI score and Log Dice—returned different results with respect to change over time, meaning that evolution with respect to rare exclusivity (as detected by MI score) and by general exclusivity (reflected in Log Dice) were distinct in this dataset. Beginning with MI score, time was a significant and positive predictor, reflecting the fact that the N+Adj combinations used at post-stay were more likely to have higher collocational association scores than those used at pre-stay. Among previous research, only [Siyanova-Chanturia \(2015\)](#) reported significant changes in collocational strength at the group level, the other four studies having reported that MI scores remained static between the beginning and end of a stay abroad ([Bestgen and Granger 2014](#); [Li and Schmitt 2010](#); [Siyanova-Chanturia and Spina 2020](#); [Yoon 2016](#)). Our own results reinforce this general trend, insofar as we found no significant change in MI score between N+Adj combinations produced before and at the end of a stay abroad. Interestingly, however, our analysis reveals change when the MI scores for combinations produced at pre-stay were compared with those produced 8 months after the participants’ return to the United Kingdom. In other words, although no change was detected immediately at the end of the stay abroad, it may be the case that the stay in France initiated the

significant development that is visible after the participants' return to their home university. Little research currently exists that has included data that aim to examine evolution after the immersion experience. In the case of the LANGSNAP corpus, the final data-collection meeting took place about eight months after the learners' return to the United Kingdom, allowing researchers to investigate change or maintenance after a return home (see, for example, [Huensch and Tracy-Ventura 2017](#), for an analysis of change in fluency in the Spanish portion of this corpus during and after the stay abroad). Importantly, the LANGSNAP research team has collected additional data three years after the end of the original project from a subset of the original participants.<sup>5</sup> Exploring how phraseological patterns may have changed in the years that followed the stay abroad offers an exciting avenue for future research.

The significant finding with respect to MI score must be nuanced, however, by the results from the Log Dice analysis. Indeed, although both Log Dice and MI score are strength-of-association measures, for the Log Dice analysis, time was not a significant predictor. From a methodological point of view, the results from the two collocational strength measures provide support for the observations made by [Gablasova et al. \(2017\)](#) concerning the fact that the two measures reveal distinct information about phraseological patterns, because each favors different types of associations. With respect to the LANGSNAP data, the fact that change was observed in the MI score analysis but not the Log Dice analysis suggests that the observed evolution concerned items that enjoy rare exclusivity (i.e., word combinations involving infrequent words), as opposed to greater use of strongly associated combinations involving words from across the frequency spectrum. In combination, the MI score and Log Dice analyses thus allow us to more specifically pinpoint the types of combinations that underwent change. Moreover, when considered with respect to the results from the overall frequency analysis, we note that the LANGSNAP learners seem to present results that are distinct from what has been reported for additional-language learners: Instead of showing sensitivity to overall frequency, they show sensitivity (in the form of development) to strongly associated collocations. It may thus be the case that these learners have reached a point in their acquisition where frequency effects are not visible, at least with respect to the production of N+Adj combinations.

The second research question that guided this project focused on the factors that significantly influence the use of more or less frequent and more or less strongly associated N+Adj combinations. As the factor of time has already been addressed, we turn now to the four remaining independent variables that were explored. Of these, three were not significant in any of the analyses: EI score, years spent studying French, and placement type. The first two variables—EI score and years spent studying French—were both intended to reflect general proficiency in additional-language French. Given recent research by [Siyanova-Chanturia and Spina \(2020\)](#), in which proficiency level significantly influenced collocation use in additional-language Italian, we sought to explore whether one or both of these variables may influence N+Adj combination use in this dataset. The fact that neither was significant may indicate that the role of overall proficiency in collocation use is stronger among lower-level learners, like those who participated in Siyanova-Chanturia and Spina's study. As for placement type, our results detected no difference with respect to this variable. This finding contributes to evidence reported by [Mitchell et al. \(2015\)](#) on the oral development by these same learners. These researchers analyzed changes in EI score and in lexical diversity over time and found no significant impact of placement type. On the basis of their findings, they observed that "every placement type offers in principle a rich exposure to French and interactional opportunities" (p. 133). In the context of the current study, it thus appears that although teaching assistants, interns, and students may receive different kinds of input, these differences did not influence the frequency or collocational strength of N+Adj combinations that they produced.

<sup>5</sup> French: <https://slabank.talkbank.org/access/French/LANGSNAP3.html>; Spanish: <https://slabank.talkbank.org/access/Spanish/LANGSNAP3.html>.

The final variable that was investigated explored the importance of adjective position (prenominal vs. postnominal). This variable was found to be significant in all three analyses, indicating that adjective position influenced both the frequency and collocational strength of N+Adj combinations in this dataset, although two contrasting patterns were observed. More specifically, when the adjective preceded the noun, there was a greater likelihood that the resulting N+Adj combination would enjoy both higher frequency and a higher Log Dice score, whereas its MI score was more likely to be low. These different patterns held over time, given that adjective position was not found to interact with time in any of the analyses. An examination of the dataset reveals that the findings for adjective position reflect at least in part the difference in the type of adjectives that were used in these two positions. In particular, we note that the adjectives found in prenominal position tended to be more frequent (on average, they appear in the frTenTen12 WaCky corpus 4,752,332 times, or 481 times per 1 million words) than the set of adjectives that appeared in postnominal position (here, the average frequency is 662,642, or 67 occurrences per 1 million words). This difference helps explain why overall frequency and Log Dice are higher for combinations involving prenominal adjectives (which, on average, have higher overall frequency), whereas MI score—which favors rare exclusivity—tends to be higher for combinations including the generally less frequent postnominal adjectives. Whereas previous research, such as that of [Granger and Bestgen \(2014\)](#), has reported different collocational strength patterns as a function of type of word combination examined (e.g., N+N, Adj+N, adverb+Adj), the current investigation extends that finding by showing that syntactic patterns within a single category (in this case, word-order differences for N+Adj combinations) may also influence collocational profiles. Thus, when analyzing phraseological categories with syntactic variation, it is important to not assume that this variation has no impact on collocational patterns. In order to be able to account for the potential influence of such variation while also exploring other factors, researchers would do well to follow [Gries' \(2015\)](#) advice and consider the use of multivariate analytical approaches. As we have shown in the current analysis, multivariate analyses provide the researcher with a way to explore the influence of multiple factors simultaneously.

## 5. Conclusions

In this paper, we explored if and how one part of the phraseological spectrum changes over time, including an academic year spent in France. Documenting the influence of a stay abroad has attracted attention from numerous researchers, and the current analysis offers additional insight into changes (but also into stasis) as regards N+Adj combinations used in written essays. Whereas our results revealed no changes in frequency or in one of the collocational strength measures, we did observe significant change in the use of combinations receiving a high MI score. When the results from the Log Dice and MI score analyses are taken together, they allow us to affirm that the stay abroad ultimately led to the increased use of N+Adj combinations that involved less frequent words, as seen in the post-stay data. Although the findings provide another example of the learning trajectory during a stay abroad, when interpreted against the backdrop of past research, several avenues for future research remain open. These include, among others, exploring the role of length of stay and proficiency in the development of phraseological competence, as well as the potential variation in developmental paths as a function of category of phraseological unit (e.g., verb+N vs. N+Adj). Moreover, one of the main findings from research on phraseological competence in an additional language suggests a clear preference for high-frequency collocations over strongly associated ones. The results from the current study are unusual insofar as we find evidence of change in MI score, but not in overall frequency (or Log Dice), suggesting particular sensitivity to N+Adj combinations enjoying rare exclusivity on the part of the LANGSNAP learners. Presuming that this finding can be replicated in future research, this result may indicate that the learners who contributed to the LANGSNAP project were no longer sensitive to frequency effects for N+Adj combinations.

**Author Contributions:** Conceptualization, A.E. and A.G.; methodology, A.E. and A.G.; formal analysis, A.E. and A.G.; writing—original draft preparation, A.E. and A.G.; writing—review and editing, A.E. and A.G. Both authors have read and agreed to the published version of the manuscript.

**Funding:** This research received no external funding.

**Institutional Review Board Statement:** Ethical review and approval were not necessary for this study, as it was conducted on data drawn from a publicly available corpus.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** Publicly available datasets were analyzed in this study. These data can be found here: [<https://slabank.talkbank.org/access/>].

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

- Anderson, Bruce. 2008. Forms of Evidence and Grammatical Development in the Acquisition of Adjective Position in L2 French. *Studies in Second Language Acquisition* 30: 1–29. [[CrossRef](#)]
- Arnon, Inbal, and Neal Snider. 2010. More than Words: Frequency Effects for Multi-word Phrases. *Journal of Memory and Language* 62: 67–82. [[CrossRef](#)]
- Baroni, Marco, Silvia Bernardini, Adriano Ferraresi, and Eros Zanchetta. 2009. The WaCky Wide Web: A Collection of Very Large Linguistically Processed Web-crawled Corpora. *Language Resources and Evaluation* 43: 209–26. [[CrossRef](#)]
- Bestgen, Yves, and Sylviane Granger. 2014. Quantifying the Development of Phraseological Competence in L2 English Writing: An Automated Approach. *Journal of Second Language Writing* 26: 28–41. [[CrossRef](#)]
- Briggs, Jessica. G. 2015. Out-of-class Language Contact and Vocabulary Gain in a Study Abroad Context. *System* 53: 129–40. [[CrossRef](#)]
- Crossley, Scott, and Thomas L. Salsbury. 2011. The Development of Lexical Bundle Accuracy and Production in English Second Language Speakers. *International Review of Applied Linguistics and Language Teaching* 49: 1–26. [[CrossRef](#)]
- Crossley, Scott, Thomas Salsbury, and Danielle McNamara. 2010. The Development of Polysemy and Frequency Use in English Second Language Speakers. *Language Learning* 60: 573–605. [[CrossRef](#)]
- Crossley, Scott, Kristopher Kyle, and Thomas Salsbury. 2016. A Usage-based Investigation of L2 Lexical Acquisition: The Role of Input and Output. *The Modern Language Journal* 100: 702–15. [[CrossRef](#)]
- Durrant, Philip, and Norbert Schmitt. 2009. To What Extent do Native and Non-native Writers Make Use of Collocations? *International Review of Applied Linguistics in Language Teaching* 47: 157–77. [[CrossRef](#)]
- Edmonds, Amanda, and Aarnes Gudmestad. 2014. Your Participation is Greatly/Highly Appreciated: Amplifier Collocations in L2 English. *Canadian Modern Language Review* 70: 76–102. [[CrossRef](#)]
- Ellis, Nick C., Rita Simpson-Vlach, and Carson Maynard. 2008. Formulaic Language in Native and Second Language Speakers: Psycholinguistics, Corpus Linguistics, and TESOL. *TESOL Quarterly* 42: 375–96. [[CrossRef](#)]
- Ellis, Nick C., Rita Simpson-Vlach, Ute Römer, Matthew B. O'Donnell, and Stefanie Wulff. 2015. Learner Corpora and Formulaic Language in SLA. In *Cambridge Handbook of Learner Corpus Research*. Edited by Sylviane Granger, Gaëtanelle Gilquin and Fanny Meunier. Cambridge: Cambridge University Press, pp. 357–78.
- Ellis, Nick C., Ute Römer, and Matthew B. O'Donnell. 2016. *Usage-based Approaches to Language Acquisition and Processing: Cognitive and Corpus Investigations of Construction Grammar*. New York: Wiley.
- Forsberg, Fanny. 2010. Using Conventional Expressions in French. *International Review of Applied Linguistics in Language Teaching* 48: 25–51. [[CrossRef](#)]
- Gablasova, Dana, Vaclav Brezina, and Tony McEnery. 2017. Collocations in Corpus-based Language Learning Research: Identifying, Comparing, and Interpreting the Evidence. *Language Learning* 67: 155–79. [[CrossRef](#)]
- Granger, Sylviane, and Yves Bestgen. 2014. The Use of Collocations by Intermediate vs. Advanced Non-native Writers: A Bigram-based Study. *International Review of Applied Linguistics and Language Teaching* 52: 229–52. [[CrossRef](#)]
- Granger, Sylviane, and Magali Paquot. 2008. Disentangling the Phraseological Web. In *Phraseology: An Interdisciplinary Perspective*. Edited by Sylviane Granger and Fanny Meunier. Amsterdam: Benjamins, pp. 27–50.
- Gries, Stefan Th. 2015. Statistics for Learner Corpus Research. In *The Cambridge Handbook of Learner Corpus Research*. Edited by Sylviane Granger, Gatanaëlle Gilquin and Fanny Meunier. Cambridge: Cambridge University Press, pp. 159–81.
- Howard, Martin. 2012. The Advanced Learner's Sociolinguistic Profile. On Issues of Individual Differences, L2 Exposure Conditions and Type of Sociolinguistic Variable. *The Modern Language Journal* 96: 20–33. [[CrossRef](#)]
- Huensch, Amanda, and Nicole Tracy-Ventura. 2017. L2 Utterance Fluency Development Before, During, and After Residence Abroad: A Multidimensional Investigation. *The Modern Language Journal* 101: 275–93. [[CrossRef](#)]
- Ife, Anne, Gemma Vives Boix, and Paul Meara. 2000. The Impact of Study Abroad on the Vocabulary Development of Different Proficiency Groups. *Spanish Applied Linguistics* 4: 55–84.
- Kinginger, Celeste. 2009. *Language Learning and Study Abroad: A Critical Reading of Research*. New York: Palgrave Macmillan.

- Li, Jie, and Norbert Schmitt. 2009. The Acquisition of Lexical Phrases in Academic Writing: A Longitudinal Case Study. *Journal of Second Language Writing* 18: 85–102. [CrossRef]
- Li, Jie, and Norbert Schmitt. 2010. The Development of Collocation Use in Academic Texts by Advanced L2 Learners: A Multiple Case Study Approach. In *Perspectives on Formulaic Language: Acquisition and Communication*. Edited by David Wood. New York: Bloomsbury, pp. 22–46.
- Llanes, Àngels. 2011. The Many Faces of Study Abroad: An Update on the Research on L2 Gains Emerged During a Study Abroad Experience. *International Journal of Multilingualism* 8: 189–215. [CrossRef]
- Lorenz, Gunter. 1999. *Adjective Intensification—Learners versus Native Speakers*. Amsterdam: Rodopi.
- Manning, Christopher D., and Hinrich Schütze. 1999. *Foundations of Statistical Natural Language Processing*. Cambridge: MIT Press.
- Mitchell, Rosamond, Kevin McManus, and Nicole Tracy-Ventura. 2015. Placement Type and Language Learning during Residence Abroad. In *Social Interaction, Identity and Language Learning During Residence Abroad*. Edited by Rosamond Mitchell, Nicole Tracy-Ventura and Kevin McManus. Eurosla Monographs Series; Amsterdam: Eurosla Association.
- Mitchell, Rosamond, Nicole Tracy-Ventura, and Kevin McManus. 2017. *Anglophone Students Abroad: Identity, Social Relationships, and Language Learning*. London: Routledge.
- RStudio Team. 2020. *RStudio: Integrated Development for R. RStudio*. Boston: PBC, Available online: <http://www.rstudio.com/> (accessed on 21 September 2020).
- Schmitt, Norbert, Suhad Sonbul, Laura Vilkaitė-Lozdienė, and Marijana Macis. 2019. Formulaic Language and Collocation. In *The Encyclopedia of Applied Linguistics*. Edited by Carol A. Chapelle. New York: John Wiley and Sons, pp. 1–10.
- Shively, Rachel L. 2011. L2 Pragmatic Development in Study Abroad: A Longitudinal Study of Spanish Service Encounters. *Journal of Pragmatics* 43: 1818–35. [CrossRef]
- Siyanova, Anna, and Norbert Schmitt. 2008. L2 Learner Production and Processing of Collocation: A Multi-study Perspective. *The Canadian Modern Language Review* 64: 429–58. [CrossRef]
- Siyanova-Chanturia, Anna. 2015. Collocation in Beginner Learner Writing: A Longitudinal Study. *System* 53: 148–60. [CrossRef]
- Siyanova-Chanturia, Anna, and Stefania Spina. 2020. Multi-word Expressions in Second Language Writing: A Large-scale Longitudinal Learner Corpus Study. *Language Learning* 70: 420–63. [CrossRef]
- Tracy-Ventura, Nicole, Kevin McManus, John M. Norris, and Lourdes Ortega. 2014. ‘Repeat as Much as you Can’: Elicited Imitation as a Measure of Oral Proficiency in L2 French. In *Measuring L2 Proficiency: Perspectives from SLA*. Edited by Pascale Leclercq, Amanda Edmonds and Heather Hilton. Clevedon: Multilingual Matters, pp. 143–66.
- Yoon, Hyung-Jo. 2016. Association Strength of Verb-noun Combinations in Experiences NS and Less Experiences NNS Writing: Longitudinal and Cross-sectional Findings. *Journal of Second Language Writing* 34: 42–57. [CrossRef]