EDICIONES COMPLUTENSE

Círculo de Lingüística Aplicada a la Comunicación
ISSN-e: 1988-2696

ARTÍCULOS

A frequency-based exploratory study of word combinations by EFL learners towards adjective-noun collocations

Erdem Akbaş
Erciyes University, Turkey

Ceyhun Yükselir
Osmaniye Korkut Ata University, Turkey

https://dx.doi.org/10.5209/clac.81885 Received: 21 September 2021 • Accepted: 13 September 2022

Abstract: The mastery of formulaic sequences by language learners is considered to be crucial since it enables them to accomplish the basic communicative requirements (Wray, 2002) of target language use. Of all the formulaic sequences, the mastery of collocations is known to have an enormous impact on learners' ability to move beyond the intermediate level. This data-driven study explores the word combinations (both conventional and unconventional L2 strings) of L2 users of English. Specifically, the adjective-noun combinations in an L2 corpus (Turkish Undergraduate Writers of English) were explored to find the potential L2 strings functioning as unconventional collocations. Then, these L2 strings were compared with the adjective-noun collocations in the LOCNESS corpus to determine whether the choices made by L2 learners differed from those of L1 users. The findings show that the spectrum of the L2 word combinations was broader in comparison with LOCNESS. The results suggest that the L2 learners produced a higher number of unconventional collocations than conventional collocations matching with the L1. The L2 strings support that the use of collocations can be rather problematic for learners of English however advanced they are. Some pedagogical implications of the findings for explicit collocation and vocabulary teaching are discussed.
Keywords: L2 word combinations; adjective-noun collocations; corpus linguistics; foreign language learning; argumentative writing.

Contents: 1. Introduction. 2. Collocational performance. 3. Overview of the study. 4. Methodology. 4.1. Data. 4.2. Procedures, Tools and Data Analysis. 5. Findings and discussion. 5.1. Overall results. 5.2. Conventional as opposed to unconventional collocations in L2. 5.3. Overall discussion. 5.4. Implications. Acknowledgments. CREDiT Authorship Contribution. References.

How to cite: Akbaş E., Yükselir C. (2024). A frequency-based exploratory study of word combinations by EFL learners towards adjective-noun collocations. Círculo de Lingüística Aplicada a la Comunicación 98 (2024), 193-208. https://dx.doi.org/10.5209/clac.81885

1. Introduction

It has long been known that even advanced learners can be regarded as disadvantaged compared with L1 users (Kjellmer, 1990; Wray, 2013) who have their naturally processed formulaic sequences within their ready and easy reach. This has led to a prominent field of research being developed to explore how to achieve a better understanding of the difficulty of L2 learners in producing multi-word units (Siyanova-Chanturia & Spina, 2020). Recent studies (Bardovi-Harlig & Stringer, 2017; Nesselhauf, 2005; Wood, 2010;) have generally focused on the choices of L2 learners to investigate whether learners have fluency in their native-like selections. This is because the «lack of this knowledge may impede the comprehensibility of learners' expression» (Laufer & Waldman, 2011, p. 648).

Wray (2013) pointed out that formulaic language can be defined as multi-word units with a focus on a particular meaning and function when the whole expression is successful with respect to its morphological structure. Despite various labels found in the literature to signal these components (ready-made utterances, formulaic sequences, and multi-word units), there is consensus over the definition of such units in that there is «some kind of syntagmatic relation of words» (Nesselhauf, 2005, p. 11). To understand the nature of collocations in language use, it is important to clarify two concepts: «restricted co-occurrence» and «relative semantic transparency» introduced by Laufer and Waldman (2011). These terms essentially promote

the understanding of how collocations should be positioned between free combinations and idioms with a fixed order and meaning. Laufer and Waldman (2011) underlined the ease of replaceability of the individual units for free combinations by following the grammatical constructions, such as replacing «second» in «second-hand» with «first» to create another meaning as in «first-hand» experience. In contrast, restricted co-occurrence, which might seem a problematic issue for L2 learners, clearly limits the potential elements to be used in the collocation. Relative semantic transparency means that the further the meaning of the collocations is away from the separate meanings of the individual words, the less transparent it becomes, such as «catch somebody red-handed», making it more idiomatic. More transparent terms, e.g., «make an application», could provide a better opportunity to draw the meanings from the separate parts.

In our study, word combinations are labelled as «collocations» considering a frequency-based view compared to traditional phraseology, which is called the study of structure and word combinations. Paquot and Granger (2012) highlighted that such combinations could be seen as «co-selection of lexical items» (p. 136) determined via statistical methods rather than looking at combinations with a lexically constrained nature, such as «make an effort». With these issues in mind, this frequency-based exploratory study is designed to scrutinize the collocational performance of Turkish EFL learners by focusing on three sets of adjectives and compare it with that of L1 users to identify conventional and unconventional combinations in L2.

2. Collocational performance

The mastery of formulaic sequences by language learners is considered to be crucial since it enables them to accomplish the basic communicative requirements (Millar, 2011; Wray, 2002) of target language use. Of all the formulaic sequences, the collocations are known to have an enormous impact on learners' ability to move beyond the intermediate level, even in oral production (Eguchi & Kyle, 2020; Saito, 2020). Previous studies emphasized how critical collocational knowledge and performance in L2 are with respect to achieving a more understandable and fluent communication in spoken and written forms (Hunston & Francis, 2000). Nesselhauf (2005) commented that collocation is «some kind of syntagmatic relations of words» (p.11) in reference to the semantic and lexical aspects of language. The innate ability of L1 users to acquire language idiomaticity and use it appropriately in the required context is, unfortunately, what many L2 learners strive for so that they might grasp the way to produce such combinations as closely as possible to native conventions. What makes it possible for L1 users to retrieve collocations more or less automatically (Ädel & Erman, 2012; Jiang & Nekrasova, 2007) is linked to their available cognitive resources to be utilized and processed. L2 learners, in contrast, appear to be disadvantaged in attaining such an automaticity and accuracy. Not only does it require much time for L2 learners to achieve «nativelike selection» (Pawley & Syder, 1983, p. 192) or idiomaticity, but it also calls for enough practice to pile them up en bloc as if they were single units. Otherwise, as Siyanova and Schmitt (2008) argued, learners are likely to end up with unconventional sequences. Pawley and Syder (1983) referred to this as «foreignisms» in the target language, giving rise to less effective and sensible communication. This could lead to an L1 influence over L2 collocation use (Yamashita & Jiang, 2010) because of the flexibility and cross-linguistic nature of achieving appropriate collocational performance.

Some studies reported that L2 learners have problems with collocations in their written and spoken L2 (Granger, 1998a; Howarth, 1998; Nesselhauf, 2003; Siyanova & Schmitt, 2008). Using appropriate collocation can even be problematic for advanced L2 learners; they «often underuse or misuse native-like expressions and use atypical word combinations instead» (Siyanova-Chanturia, 2015, p. 149). It is noted that L1 users have a tendency to use conventional phraseology to convey their ideas whilst L1 users choose unconventional word combinations (Siyanova & Schmitt, 2008). Regarding the phraseological nature of language, many researchers (Hoey, 2005; Wray & Perkins, 2000) have remarked that phraseology occupies a significant role in being fluent and accomplishing idiomatic language use. In line with this, some combinations could make learners less effective in their ability to become fluent and communicative. As for the difficulty of achieving high performance in collocations, Bahns and Eldaw (1993) stated that the L2 collocational competence does not run parallel with their general vocabulary knowledge. In a study of verb-noun collocations, Nesselhauf (2003) found that advanced German-speaking learners of English made errors in collocations (79%) in contrast with free combinations (23%) and idioms (23%). Farghal and Obiedat (1995) investigated adjective-noun and noun-noun collocations by Arabic learners of English and concluded that the L2 learners were not aware of collocations and not proficient enough to use them in their L2 properly.

3. Overview of the study

Earlier studies have already shown that L2 learners use unconventional word combinations compared to L1 users, even at advanced-levels of proficiency (e.g., Altenberg & Granger, 2001; Howarth, 1998; Nesselhauf, 2005; Wang & Shaw, 2008). However, it is also widely documented that more research from different contexts is needed to explore L2 collocation development in English for a more universal understanding of the case. In line with this, not much is known as to how Turkish EFL learners use word combinations including (un)conventional collocations. Motivated by the importance of learner corpus as a valuable resource for understanding the nature of inter-language performance (Gilquin, 2015) in SLA research, the present study was designed specifically to explore L2 word combinations of Turkish learners of English by examining a relatively large learner corpus and comparing their performance with that of L1 users by focusing on a selected set of adjectives and their collocating nouns. The reason for investigating adjective-noun collocations here is that the syntactic distribution of adjective-noun collocations in both Turkish and English follows the same

syntactic order (as in *yakın* (adjective) gelecek (noun) meaning immediate/near future), enabling us to avoid having to focus on any syntactic problems of Turkish EFL learners. The following questions were therefore devised:

1. How do Turkish EFL learners perform with adjective-noun word combinations in their English texts?
1.1. What are the most common adjective-noun collocations used by L2 users, specifically with three groups of adjectives, namely, good/bad; positive/negative and important/significant/crucial?
1.2. Do the L2 strings match those of L1 users?

While exploring L2 users' extent of producing collocations, the immediate comparison of the collocations from the L2 corpus with the L1 corpus is significant since the comparison could enable us to see (un)conventional use by L2 learners.

4. Methodology

4.1. Data

The learner corpus of our study is the corpus of Turkish Undergraduate Writers of English (TUWE hereafter, Akbaş, 2022) and the native-speaker corpus is the Louvain Corpus of Native English Essays (LOCNESS hereafter). The TUWE corpus contained 341 essays written by Turkish EFL learners. The corpus is a medium-sized specialized learner corpus with approximately 270,000 words. The writers were first-and second-year undergraduates with English proficiency level of intermediate to upper intermediate. The essay topics ranged from «population explosion», «migration» to «parenting» and «job selection». The reason why learners were asked to write on broad topics is linked to making them express their opinions without much research, at least not as much as is required by academic writing on a topic. The LOCNESS corpus compiled by Granger (1998b) contains 325,000 words from essays written by British and American students on different topics. Considering the corpora sizes, we believe that TUWE represents a reliable sample of L2 and LOCNESS as a reference corpus gives us an opportunity to compare L2 performance with L1 use.

Table 1. Adjectives & nouns in corpora

Adjectives	***Simple***	***per 10,000***	***Comparative***	***per 10,000***	***Superlative***	***per 10,000***	***TOTAL***	***per 10,000***
TUWE	21,400	803.95	1,174	44.11	585	21.98	23,159	870.04
LOCNESS	23,915	741.59	1,459	45.24	601	18.63	25,975	805.46
Nouns	***Singular***	***per 10,000***	***Plural***	***per 10,000***	***Proper***	***per 10,000***	***TOTAL***	***per 10,000***
TUWE	42,719	1604.87	24,918	936.12	3668	137.80	71,305	2678.79
LOCNESS	51,604	1600.21	21,161	656.19	13,013	403.52	85,778	2659.92

The reason why TUWE was compared specifically with LOCNESS is because of the status of equivalence (Moreno, 2008) of the texts. All texts in both corpora were single-authored; the level of expertise of the writers was similar and both corpora were compiled from the same genre, that is, argumentative essays, in educational settings. Moreover, the writers in the corpora were assumed to have a similar communicative purpose because they produced their argumentative writing mainly for their teachers/markers who can therefore be classified as the intended readership. Since our focus is on comparing the adjective-noun collocations in the TUWE with those in the LOCNESS, it is important to see how similar the selected corpora are in terms of the general use of adjectives and nouns (cf. Table 1). Thus, we believe that the two corpora were comparable to explore to what extent the collocational performance of the L2 learners exhibited features of their L1.

4.2. Procedures, Tools and Data Analysis

The principal focus of this study is only adjective-noun collocations but to make the analysis and the comparisons more manageable, we shall zoom in on the twenty most common adjectives in both corpora with a synonymous or antonymous semantic relationship between them. Here, we shall explain why we started with adjectives for adjective-noun collocations and how we ended up with three sets of adjectives. Since this is an exploratory study, we first identified the most common adjectives in TUWE and LOCNESS using the UAM Corpus Tool 3.3, software for the annotation of texts (O'Donnell, 2008) for part of speech. Then, we removed very context- and topic-specific adjectives (i.e., European, nuclear, British, Turkish, genetic) from the list to find congruent ones.

Listing the adjectives, we identified the twenty most common adjectives between the two corpora from the top 200 adjectives as in Table 2. The reason why we started by looking at the top 200 adjectives from each corpus is to narrow down the possibilities for the matching most common adjectives with semantically related nature. It should be noted that the selection of the adjectives was made in reference to the TUWE

corpus. Thus, the adjectives listed from TUWE were checked individually in the LOCNESS adjective list. As an example, the raw frequency of «essential» in TUWE was eighty-seven and it ranked as the 58th most commonly used adjective, but it ranked 213th with only eight occurrences in LOCNESS. Therefore, «essential» was excluded from the top-twenty-adjectives list.

Table 2. Most common adjectives matching in the two corpora within the top 200 adjectives

	TUWE			***LOCNESS***		
	Freq	***per 10,000***	***Noun Collocates***	***Freq***	***per 10,000***	***Noun Collocates***
important	462	17.35	12	213	6.60	4
good	408	15.32	14	315	9.76	9
better	343	12.88	8	182	5.64	2
different	285	10.70	12	201	6.23	3
new	273	10.25	8	303	9.39	8
same	265	9.95	6	205	6.35	6
crucial	192	7.21	7	9	0.27	0
significant	175	6.57	8	52	1.61	0
negative	171	6.42	5	79	2.44	2
great	168	6.31	4	175	5.42	2
bad	163	6.12	6	153	4.74	2
main	154	5.78	5	133	4.12	7
high	151	5.67	8	116	3.59	1
political	149	5.59	6	150	4.65	4
effective	142	5.33	4	60	1.86	1
positive	138	5.18	5	75	2.32	2
right	136	5.10	5	284	8.80	1
personal	113	4.24	4	92	2.88	3
easy	109	4.09	1	62	1.92	3
big	108	4.05	5	61	1.89	2

Note 1: We checked systematically the most frequent TUWE adjectives against the most frequent LOCNESS adjectives, retaining for the analysis the first 20 appearing within the 200 most frequent adjectives in both corpora.
Note 2: The threshold for the statistic value was set to 6.0 and for the collocation frequency was set to 5.0 to find the collocating nouns of these adjectives.

We used LancsBox 4.5, a corpus tool that helps researchers visualize the collocational networks (Brezina et al. 2015), to retrieve the adjective-noun collocations. Table 2 shows the raw frequency of the matching adjectives in both corpora and the noun collocates. We note that the collocates were determined by looking only at the adjacent nouns only coming immediately after the adjectives. Otherwise, the collocation profiles for each adjective changed considerably depending on the parts of speech coming after an adjective, e.g., the preposition «to» which could be identified as a collocate to the adjective «important» as in «important thing to know». The span was therefore adjusted only one word to the right while searching for the noun collocates.

When the collocates were retrieved from TUWE (cf. Table 2), the total number of collocations (Freq>5) was 1708. Owing to our aim of comparing L2 adjective-noun combinations with L1, the collocates we found were too many to carry out a robust analysis and identify the (un)conventional L2 strings. Yet, we found that particular adjectives were semantically related in one way or another, providing an opportunity to deal with a «more manageable number of adjective-noun collocations». We therefore focused on the antonymous adjectives «(1) good, (2) bad» and «(3) positive, (4) negative» and synonymous adjectives «(5) important», «(6) crucial», and «(7) significant».

Following the argument by Daille (1995) and Garner et al. (2018) about the significance of MI2 score over MI, we relied on MI2 scores to decide the word combinations and fine-grained understanding of the frequency and selection of words, which is also suggested by Gablasova et al. (2017). Some noun collocates for the adjectives were excluded by measuring the MI2 scores and frequency. An MI2 score of a statistical value over 6.0 and a collocation frequency over 5.0 were considered the cut-off point to determine the strongest associations between adjectives and collocating nouns. For example, in LOCNESS, the collocation of «negative» with the noun «aspect» (MI2: 10.90) shows that it has a slightly stronger association compared with the collocation of «negative» with «effect» (MI2: 9.80). Although the cut-off point was arbitrary, there was nevertheless a large number of collocations to deal with. To summarize, to carry out a detailed analysis of L2 word combinations, we performed a comparison between the learner corpus and the native-speaker corpus with the same set of adjectives by setting the threshold of an MI2 score of 6.0 and the collocations occurring at least five times in the corpora. Other than the MI2 score, we also relied on chi-squared tests to check the difference between two corpora with respect to word combinations produced by the same adjectives (as in Table 3) and the individual adjective-noun collocations (see Table 4, 5, and 6).

5. Findings and discussion

5.1. Overall results

In order to answer the major research question with sub-questions, we addressed the adjective-noun collocations produced by L2 users and checked them against the L1 corpus. Table 3 gives a general overview of the adjective-noun collocations in each corpus with examples of collocating nouns, if there were any.

Table 3. Adjective-Noun Collocations in TUWE and LOCNESS

	TUWE			***LOCNESS***			***P***
Adjectives	***Nouns***	***Examples***	***Freq***	***Nouns***	***Examples***	***Freq***	
Good	14	good role good job good effect	170	9	good cause good quality good job	80	0.00
Bad	6	bad habit bad side	103	2	bad faith bad thing	59	0.00
Positive	5	positive side positive effect	63	2	positive aspect positive effect	13	0.57
Negative	5	negative effect negative thing	115	2	negative aspect negative effect	14	0.01
Important	12	important role important decision	146	4	important aspect important role	25	0.09
Significant	8	significant role significant effect significant thing	74	0	N/A	0	0.00
Crucial	7	crucial role crucial matter	70	0	N/A	0	0.00
Total	57		742	19		191	

Note 1: The threshold for the statistic value was set to 6.0 and for the collocation frequency was set to five to find the collocating nouns of these adjectives.
Note 2: Freq represents the frequency of an adjective with collocating nouns in TUWE or LOCNESS and $p<0.05$ indicates a statistically significant difference between TUWE and LOCNESS

Table 3 shows that there is a broader range and higher number of L2 word combinations with most of these adjectives than the L1 speakers as far as the p-values indicate. There are several significant points in this finding. Preliminary findings showed that the L2 users produced more collocations with respect to the selected adjectives (742 in TUWE; 191 in LOCNESS). Specifically, L2 users tended to attach adjectives to nouns which were not modified by the same adjectives in comparison with L1 users of English. For instance, in Figures 1 and 2, the collocation formed by the adjective «good» modifying the noun «job» is regarded as a conventional collocation since the same collocation is also found in LOCNESS. However, the combinations formed by attaching the same adjective (good) to other nouns, i.e. «side», «way» and «effect», suggest that the L2 users produced unconventional collocations as there were no instances of such L2 strings in LOCNESS (cf. Figure 2).

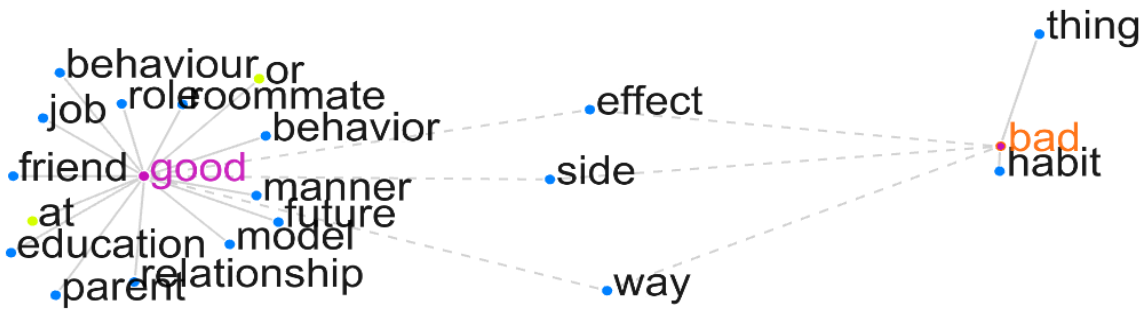

Figure 1. Collocation network of «good»-«bad» in TUWE

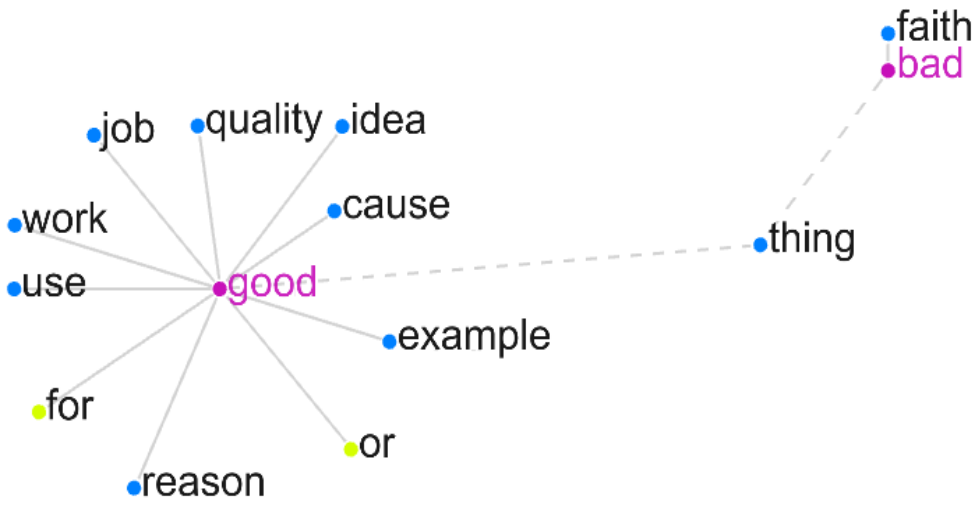

Figure 2. Collocation network of «good»-«bad» in LOCNESS

Second, Figure 3 shows that L1 users reserved the gradable adjective pairs «positive» and «negative» almost exclusively for «aspect» and «effect». In contrast, interestingly, the L2 users produced adjective-noun strings not present in the L1 corpus (compare Figure 3 and 4). As an illustration, «positive side» (MI2:11.54) and «negative side» (MI2:11.42) in TUWE did not seem to be conventional since L1 speakers opted for attaching these adjectives to modify particular nouns only (i.e., «aspect» and «effect»).

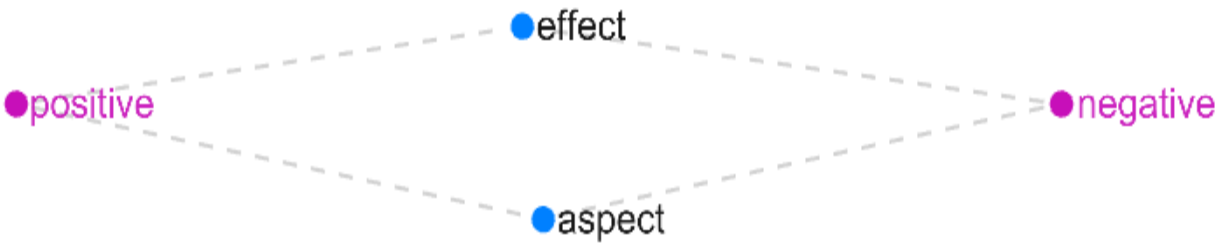

Figure 3. Shared nouns in LOCNESS between «positive» -«negative»

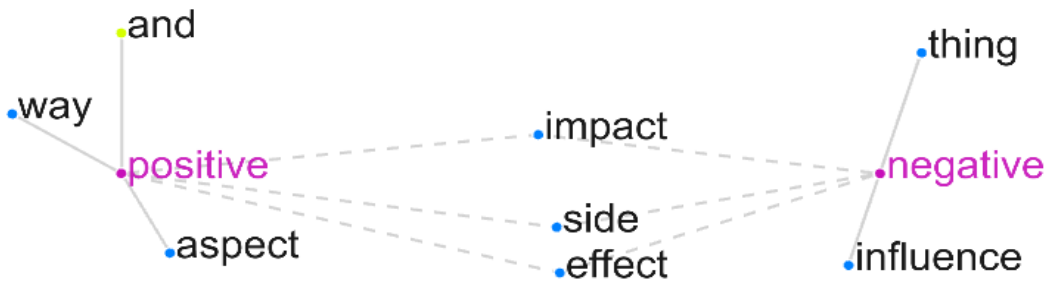

Figure 4. Shared nouns in TUWE between «positive» -«negative»

Third, although the adjectives «important», significant», and «crucial» were thought to be synonymous, there was a clear tendency in L1 speakers to use «important» to create collocations with particular nouns («aspect», «role», «decision», and «part») but not the others. Conversely, L2 users modified almost the same nouns, «role», «point», «problem», and «factor», with all of these synonymous adjectives interchangeably (Figure 5). This becomes more apparent when Figures 5 and 6 are compared:

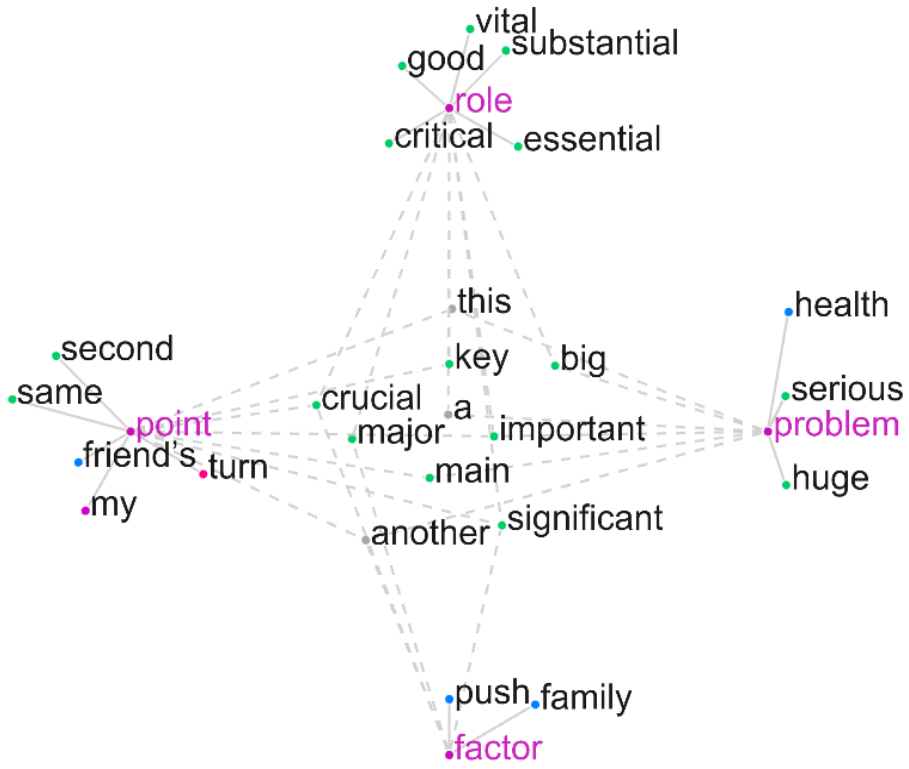

Figure 5. Common nouns modified by «important», «significant», and «crucial» in L2

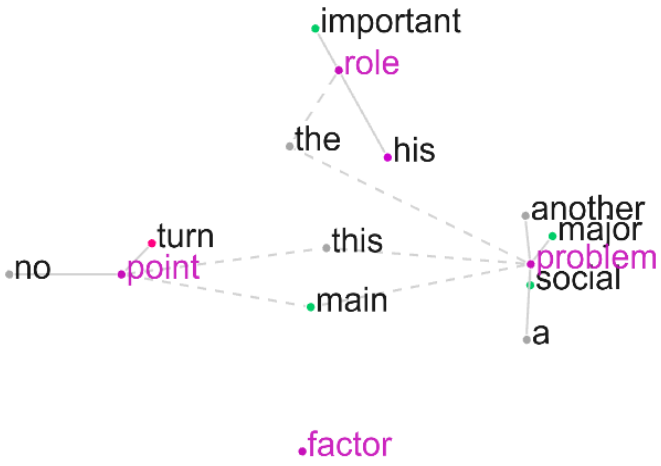

Figure 6. Noun collocates of «important», «significant», and «crucial» in L2 checked against L1

Interestingly, despite a relatively close raw frequency of these adjectives, with 462 for «important», 175 for «significant», and 192 for «crucial» in TUWE, L1 speakers did not employ these adjectives as much as L2 users did (213, 52, and 9 in LOCNESS, respectively). Furthermore, L1 speakers did not produce collocates in which the nouns matched with these three adjectives (cf. Figure 6). This shows that of all three adjectives, L1 speakers gave preference to an overwhelming use of «important» and a scarce use of «crucial» compared with L2 users with relatively equal frequency among them. This might stem from potential L2 input, knowledge with respect to the use of synonyms or to L1 interference since these three adjectives have an exact counterpart in Turkish: «önemli».

With these findings in mind, we then sought to categorize the performance of L2 users with conventional as opposed to unconventional collocations to create a map to answer the second sub-question.

5.2. Conventional as opposed to unconventional collocations in L2

This section shall consider whether the collocational performance of the L2 users was close to that of L1 speakers or deviated from L1 for potential reasons. Before reporting conventional or unconventional collocations, it is important to discuss what is meant by them. To be more specific, when a collocation produced by L2 users is in harmony with the native-speaker use, we categorize it as a conventional collocation, but if there is no exact match between groups or there is a gap between the groups when the L2 string is checked against L1 via MI2, then the word combination production is coded as unconventional. Although conventional

collocations can be assumed to signal a higher level of collocational performance from the inter-language perspective, unconventional collocations could underline the assumption that the inter-language users have difficulty in «denoting a type of word combination» (Nesselhauf, 2003, p. 224) owing to either L1 influence or the lack of L2 input. Tables 4, 5, and 6 clearly illustrate how such an analysis contributes to our understanding of L2 use compared with L1.

Table 4. Collocational performance of L2: Good/Bad+noun

	Collocations in L2	***MI2 in L2***	***MI2 in L1***	***p***
Good + Noun	good role	11.01	0.00	0.00
	good deal	9.46	0.00	0.00
	good side	9.33	0.00	0.00
	good manner	9.12	0.00	0.00
	good model	8.92	0.00	0.00
	good relationship	8.71	0.00	0.00
	good way	8.49	0.00	0.00
	good job	8.16	7.93	0.38
	good friend	8.10	0.00	0.00
	good behaviour	7.82	0.00	0.00
	good effect	7.30	0.00	0.00
	good future	6.91	0.00	0.00
	good opportunity	6.23	0.00	0.00
	good parent	6.19	0.00	0.00
Bad + Noun	bad habit	14.81	0.00	0.00
	bad side	11.30	6.20*	0.05
	bad effect	10.35	0.00	0.00
	bad way	8.81	0.00	0.00
	bad thing	7.46	7.68	0.95
	bad friend	6.09	0.00	0.00

Note 1: The frequency threshold of L1 collocations was set to 1 (previously 5) to see whether particular collocations in L2 ever existed in LOCNESS.
Note 2: For Table 4, the status of the collocations in L1 was set to conventional when the same collocation achieved MI2>6.0; F>5 in L1; unconventional when the same collocation failed to achieve MI2>6.0; F>5 in L1. The asterisk reveals that the collocation appears in L1 with an expected MI2 value but cannot meet the frequency threshold of 5.
Note 3: $p<0.05$ means a statistically significant difference. In case of $p>0.05$, it can be regarded as conventional, indicating both corpora include them.

The most interesting aspect of this table is that the L2 users seem to have produced various unconventional collocations. For the adjective pair «good-bad», there was only one conventional collocation for each; «good job» (MI2:8.16 vs. 7.93) and «bad thing» (MI2:7.46 vs. 7.68). The rest of the word combinations by L2 users (i.e., «good future», «good way», «bad effect», and «bad friend») were assumed to be unconventional and thus can be categorized as deviant or «atypical word combinations» (Siyanova-Chanturia, 2015, p. 149).

Although there could be many types of mistakes made by L2 users in producing target-language collocations, it is highly possible to mention transfer issues from learners' L1 into their L2 for some particular word combination sets in the present study. This most probably results in collocational faux amis (Klégr & Šaldová, 2006) also known as false collocations. One of the unconventional collocations, «good future» (MI2:6.91), can be regarded as interference by the Turkish language (L1) because a direct translation of «good future» can be «*iyi gelecek*» in Turkish, but that could be back-translated into English as «bright future» as the closest meaning. Figure 7 shows that some L2 users did use «bright» to modify «future» (MI2:11.50, Freq< 5) rather than «good», so «good future» appears to be, in fact, a representation of an atypical collocation because of unconventional sets in L2 not showing up in L1 at all.

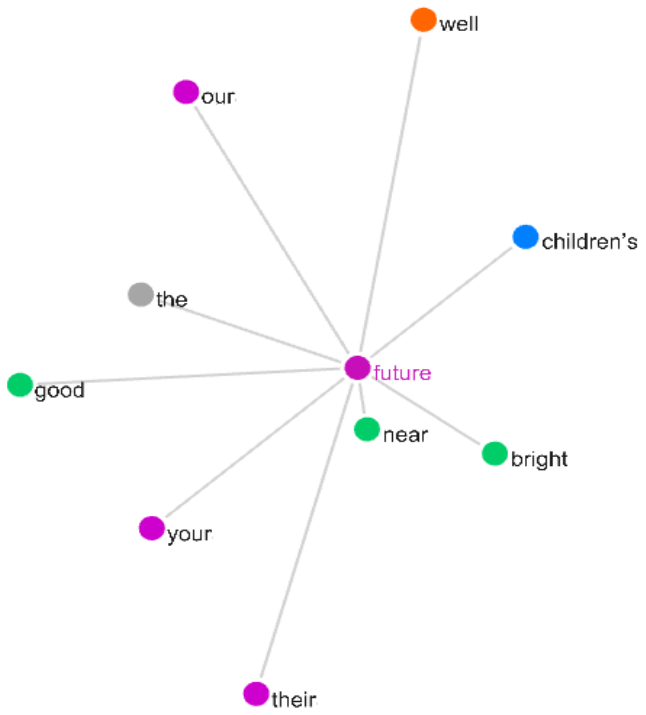

Figure 7. Collocates of «future» in TUWE

Another instance of a potential L1 influence on the collocational performance of L2 users was the case of «good» and «bad» combined with the noun «effect». The MI2 scores of «good effect» (7.30) and «bad effect» (10.35) signaled a strong association in L2. Nevertheless, the L1 corpus did not reveal instances of such combinations. Stubbs (1995) suggested that the combination of pairs with a higher score simply reveals that the components of the combinations do not seem to break up but stand out as a formulaic sequence. So with no apparent occurrence in the L1 corpus, the L2 users appeared to access this formulaic language through their L1 since the strong association measures of these collocations could also be a sign of a transfer issue.

Regarding collocations with «positive-negative», there was a discrepancy between what L2 users modified by using them and what L1 speakers did. L1 speakers appeared to modify only «effect» and «aspect» and create related collocations (cf. Figure 3 and 4) even though the range of noun collocates in L2 was five for each with three of them matching between positive-negative in L2: «impact», «side», and «effect».

Table 5. Collocational performance of L2: «positive»/«negative»+noun

	Collocations in L2	***MI2 in L2***	***MI2 in L1***	***p***
Positive + Noun	positive side	11.54	0.00	0.00
	positive effect	11.11	9.43	0.94
	positive impact	10.13	0.00	0.00
	positive aspect	9.68	10.98	0.17
	positive way	8.22	6.10*	0.79
Negative+ Noun	negative effect	14.05	9.80	0.77
	negative side	11.42	8.33*	0.80
	negative impact	10.99	7.08*	1.00
	negative influence	9.34	9.67*	0.09
	negative thing	6.95	7.16*	0.14

Note 1: For Table 5, the status of the collocations in L1 was set to conventional when the same collocation achieved MI2>6.0; F>5 in L1; unconventional when the same collocation failed to achieve MI2>6.0; F>5 in L1. The asterisk reveals that the collocation appears in L1 with an expected MI2 value but cannot meet the frequency threshold of 5.
Note 2: The asterisk reveals that the collocation appears in L1 with an expected MI2 value but cannot meet the frequency threshold of 5.

Also, as shown earlier in Table 4, Figures 3 and 4, it is apparent that the L2 users employed more combinations with a wider range of noun collocates whereas L1 speakers tended to use a restricted range of collocating words modified by «positive» and «negative». Based on MI2 scores, «aspect» was the only noun collocate matching between the two corpora modified by «positive», but «effect» was the only noun collocate shared by L1 and L2 users modified by «negative». Interestingly, the noun «impact» was the common collocate for «positive» and «negative» for L2 users, but that does not seem to be the case for L1. L1 users also used «negative» to modify the word «impact» in reference to unfavourable issues whereas they preferred to modify the word «impact» by «significant», «considerable» or «great» rather than «positive» to convey a positive meaning.

With respect to the synonymous adjectives «important-significant-crucial», there were also striking results over the L2 collocational performances in comparison with the L1 users. Table 6 presents clear evidence that the L2 users again employed a wider range of noun collocates. To be more specific, even though there were twenty-seven different adjective-noun combinations in L2 with the synonymous words, there were only four adjective-noun collocations found in L1 (MI2>6.0; F>5).

Table 6. Collocational performance of «important», «significant», and «crucial»+noun

	Collocations in L2	***MI2 in L2***	***MI2 in L1***	***p***
Important + Noun	important role	10.48	8.95	0.35
	important issue	9.32	6.77*	0.83
	important criterion	9.13	0.00	0.00
	important decision	8.80	7.95	0.12
	important thing	8.68	0.00	0.00
	important point	7.84	0.00	0.00
	important factor	7.73	8.70*	0.38
	important step	7.39	0.00	0.00
	important part	7.29	7.33	0.00
	important one	6.72	0.00	0.00
	important problem	6.23	0.00	0.00
	important reason	6.06	0.00	0.00
Significant + Noun	significant role	10.23	0.00	0.00
	significant factor	9.87	8.74*	0.18
	significant criterion	9.53	0.00	0.00
	significant effect	8.24	0.00	0.00
	significant issue	8.16	0.00	0.00
	significant cause	8.13	0.00	0.00
	significant thing	7.74	0.00	0.00
	significant point	6.72	0.00	0.00
Crucial + Noun	crucial criterion	10.22	0.00	0.00
	crucial role	9.87	0.00	0.00
	crucial part	9.28	7.25*	0.58
	crucial factor	9.04	0.00	0.00
	crucial matter	8.72	0.00	0.00
	crucial point	8.28	0.00	0.00
	crucial step	8.27	0.00	0.00

Note 1: For Table 6, the status of the collocations in L1 was set to conventional when the same collocation achieved MI2>6.0; F>5 in L1; unconventional when the same collocation failed to achieve MI2>6.0; F>5 in L1. The asterisk reveals that the collocation appears in L1 with an expected MI2 value but cannot meet the frequency threshold of 5.

Note 2: The asterisk reveals that the collocation appears in L1 with an expected MI2 value but cannot meet the frequency threshold of 5.

Figure 8 illustrates the fact that L2 users used these synonymous adjectives in place of one another with almost the same nouns; eight nouns were essentially the shared ones modified by these adjectives.

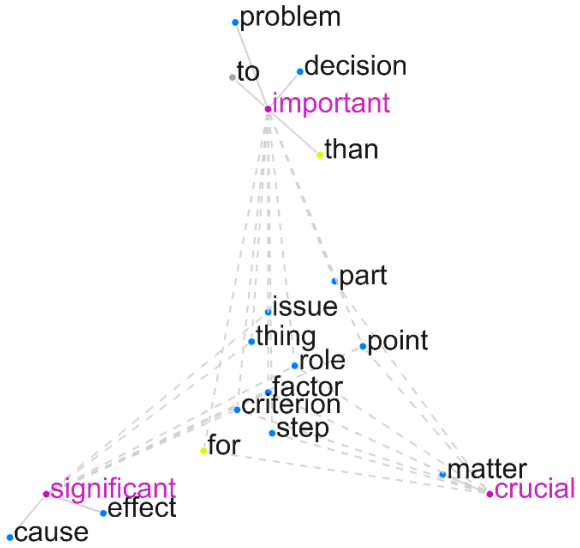

Figure 8. Shared noun collocates: «important», «significant», and «crucial» in L2.

With respect to the case of L1 speakers» use of these synonymous adjectives, a major difference between L1 and L2 users becomes more apparent when Figure 8 is compared with Figure 9, indicating that these adjectives were used to modify more specific nouns and create a discrete collocational network in L1.

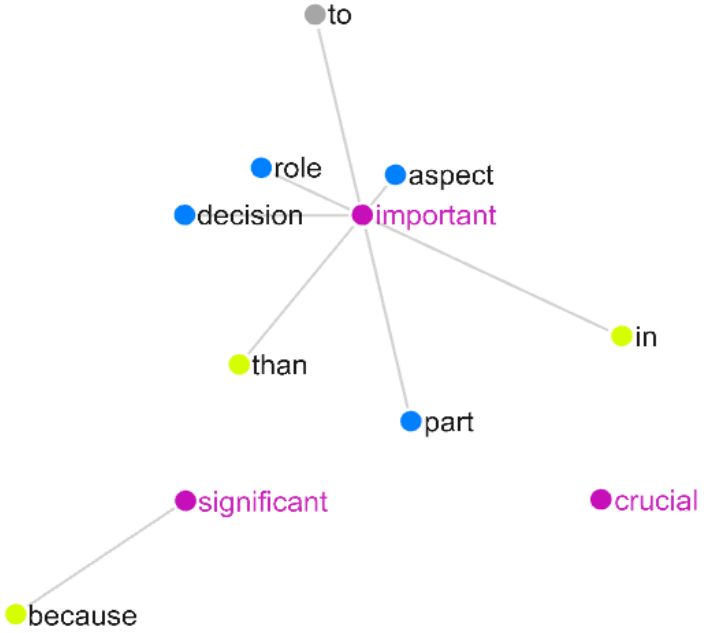

Figure 9. Shared noun collocates: «important», «significant», and «crucial» in L1

The Tables 4, 5, and 6 lead to the conclusion that of the 57 L2 strings, there were only eight (less than 14% of all) matching with those of L1 users, which can be considered conventional combinations:

- ☐ «good job»
- ☐ «bad thing»
- ☐ «positive effect», «positive aspect»
- ☐ «negative effect»
- ☐ «important role»; «important decision», «important part»

However, the 49 L2 strings either did not show up in the L1 corpus (n=39, e.g., «good future», «bad friend», «significant cause») or had a weaker association due to their infrequent nature in L1 (n=10, e.g., «negative side», «positive side», «important issue»). When the MI2 scores were calculated by decreasing the frequency criterion to 1 for infrequent combinations in L1, we found ten more combinations in L2 matching with L1 speakers. Yet, L1 speakers are less likely to use them, based on the relatively low frequency (Freq< 5), i.e., «bad side», «positive way», and «crucial part».

5.3. Overall discussion

The present study investigated the collocational performance of Turkish EFL learners using a frequency-based view to identify the most common adjective-noun combinations focusing on three sets of adjectives (i.e., good-bad; positive-negative as antonymous and important-significant-crucial as synonymous). Three principal conclusions were drawn after the detailed analyses. First, the L2 users produced more varied combinations with respect to the selected adjectives in comparison with the results from L1 users on the basis of the criteria to identify combinations. Specifically, L2 users had a tendency to attach adjectives to nouns which were not modified by the same adjectives in the L1 corpus. In other words, the results show that many adjective-noun combinations that were identified as collocations in L2 differed from those identified as collocations for L1 speakers. Second, out of 57 L2 adjective-noun strings, only eight matched the L1 collocational uses, which can be considered conventional collocations. It is nevertheless noteworthy to state that some L2 strings of Turkish EFL learners did not appear to have strong associations (n=39) i.e., «good future», «bad friend», and «significant thing» or higher frequency (n=10) in L1. Third, the L2 users seemed to produce various unconventional collocations than conventional collocations in comparison with those produced by L1 speakers, due partly to potential L1 interference, incorrect choice or a lack of awareness. This result is in parallel with the study findings of Nesselhauf (2005) stating that almost half of collocational errors could be stemming from the L1 interference. According to Wang and Halenko (2022), instruction for L2 learners can help them reduce negative transfers and contribute to their pragmatic competence and communicative tasks. For «important», «significant», and «crucial», there seem to be interesting results in the collocational performances of L2 compared with L1 users since the L2 users employed a wider range of noun collocates modified by these adjectives.

These results are not in line with those of previous studies exploring collocational uses. Regarding both adjective-noun and verb-noun combinations, earlier research (see Altenberg & Granger, 2001; Laufer & Waldman, 2011) found that L2 users produce considerably fewer collocations compared to native speakers of English. In addition, Siyanova and Schmitt (2008) found that L2 learners could produce more appropriate adjective-noun collocations (45%) than deviant ones. Yet, our analysis indicated that L2 learners in the current study tended to produce more deviant collocations rather than conventional ones that appear in L1 corpus, such as «good future» rather than «bright future», suggesting a possible influence of their L1. This confirms the findings of Demirel and Kazazoglu (2015) exploring collocational use by Turkish learners of English; they concluded that their L1 had a greater influence on the production of appropriate target-language collocations. In particular, Koç (2006) also concluded that Turkish speakers made commonly collocational errors. Regarding the source of collocational errors, Hasselgren (1994) introduced the concept of the element of familiarity, suggesting that the choices made by L2 learners mostly stem from their L1 which they use for the literal translation of structures or phrases, resulting in «lexical teddy bears» (p.238) unique to a particular group of L2 users. Some lexical teddy bears for Turkish EFL learners can be represented by «good future», «crucial matter», and «bad effect».

The collocational performance of L2 learners of English is important for acquiring language fluency to produce natural and authentic discourse (Basal, 2019; Laufer, 2011; Lewis, 2000; Wray, 2002) Some studies (Durrant & Schmitt, 2009; Siyanova & Schmitt, 2008) concluded that L2 learners can produce a high proportion of acceptable collocations which match the standard norms of L1 users. However, such results are not generalizable for different L2 users simply because L2 learners have also been found to have difficulties in using collocations properly no matter how advanced they are (Hill, 2000; Howarth, 1998; Laufer & Waldman, 2011; Nesselhauf, 2003; Shin & Nation, 2008). This seems to be the case for the present study as well when the particular findings are taken into consideration. As an example, the L2 users produced 57 different adjective-noun word combinations, only eight of which appeared in the L1 corpus and could therefore be regarded as well-formed collocations. The findings of the current study therefore also contribute evidence that L2 learners can devise word combinations not matching the native norms/conventions, possibly stemming from L1 influence, lack of L2 exposure and awareness, or the absence of explicit instructions over collocations. However, this does not necessarily mean that L2 competence in general can be treated as a deviant performance of L1-English speakers. Rather, the findings need to be evaluated under the light of a broader conceptualization known as interlanguage. In addition, due to the exploratory nature of the current study focusing on three sets of adjectives with a particular semantic relationship, we are well aware that our study could only help us gain a certain understanding in relation to the collocational performance of Turkish EFL learners. Thus, a larger number of L2 strings produced by Turkish learners and even data from the same L2 with typologically different L1 groups might be explored to delve into the L2 collocational performance to see more evidence of consistent behavior among the learners of English. The findings of the study could also be treated as a departure point for a further qualitative study exploring why the use of collocations poses a challenge for EFL learners, particularly for Turkish EFL learners.

5.4. Implications

The findings of this data-driven study show that Turkish EFL learners had some issues regarding adjective-noun collocations. The results demonstrate that they produced more varied collocations and even unconventional ones with the selected adjective-noun combinations compared with L1 speakers. This is most probably the consequence of the arbitrary and problematic nature of collocations (Fan, 2009) for L2 users of English.

Durrant and Schmitt (2010) suggested that L1 users have the opportunity to acquire a language ability in sequences and collocational chunks. However, the collocational competence and performance of L2 users are strongly linked to their L2 vocabulary repertoire (Fan, 2009). It is therefore highly possible that the failure to use conventional forms of collocations could stem from insufficient exposure to the target language vocabulary in practical use. Wray (2002) stated that L2 learners are likely to use individual words rather than collocational chunks in their learning process, which gives them a sense of authority over the target language.

In the light of the problematic nature of collocations and our findings, there are several significant pedagogical implications for teaching and learning collocations other than the explicit teaching of collocations and formulaic sequences (see Liou & Chen, 2018) for explicit teaching.

First, from the teaching perspective, integrating various online tools and designing multi-faceted tasks rather than using traditional activities will provide effective and practical ways of teaching collocations. As an example, Basal (2019) suggested that incorporating a combination of various online tools, i.e., Google Docs, in well-designed activities for collocation teaching could enable learners to collaborate with one another in their own learning process. By integrating an automatic collocation writing assistant, Chang et al. (2008) studied the issue of deviant verb-noun collocations and found that such an automatic writing assistant could make learners competent in attaining better collocational knowledge and accordingly reducing the possibility of unconventional choices in collocation. In other words, the design and implementation of technology-assisted context-rich collocation activities could enhance both learners' competence and performance with respect to conventional language use. To help learners having difficulty in using well-formed collocations, it is therefore better for teachers to benefit from the power of technological tools in language classes for learning to write in L2 (Pérez-Paredes, 2022), especially for teaching collocations together with tasks to develop learners' accessible collocational knowledge. However, according to Henriksen (2013), most of language teachers do not have chance to access to suitable materials for teaching collocations or which word co-occur with other words.

Teachers could also prepare materials and concordancing activities through available L1 corpora to increase learners' awareness, especially in vocabulary teaching (see, Girgin, 2019; Ucar & Yükselir, 2015). For instance, Ordem and Paker (2016) found that a period of treatment and exposure to lexical items through explicit teaching could have an effect on the learning of collocations and producing fewer unconventional collocations. It will work better when words can be taught in collocational chunks rather than as individual units. Shin and Kim (2017) showed that lexical bundles help to teach the correct use of articles in discourse and context for L2 learners of English of different levels, which shows that vocabulary and grammar teaching meet at some point. In addition, teachers can also enable their students to benefit from miscellaneous types of target language corpora. The integration of freely available (e.g., COCA) and related corpora (e.g., learner, native, academic corpora) into the curriculum can emphasize the importance and role of authentic language use in language learning. This will provide L2 learners with many potential learning opportunities because they could make use of the corpus at any time, making them aware of the standard norms and language usages. By accessing real language use through corpus, learners can simply view language as a whole where vocabulary choices and grammar work together to make meaning. Finally, it is recommended to equip L2 learners with the ability to search for the appropriate language to use through research. This is feasible when L2 learners can act more like researchers by adopting a data-driven learning (DDL) approach, making them more autonomous. The DDL approach promotes language learning by filling the gaps in learners' minds and enables them to interact with various data (Flowerdew, 2015). Basal (2019) stated that learners could undertake the role of «explorers» while treating the collocations in comply with DDL principles. Koosha and Jafarpour (2006) showed that the collocational performance of L2 learners could be greatly enhanced by DDL because the approach contributes to their awareness significantly. DDL-oriented activities could eventually encourage autonomous learning since the learning becomes more student-led (Crosthwaite, 2017). In addition, we highly recommend that the L2 learners need to be supported in such activities by some visuals generated via corpus tools (i.e., GraphColl) showing collocational networks and shared collocations of target items to increase L2 learners' chances of discovering patterns in language.

To conclude, our findings show that the use of collocations could be a difficult task for L2 English learners, even for upper-intermediate ones and for the ones who have experienced great exposure to incongruent L2 collocations (Khantiwong & Thienthong, 2022). Although there could be many reasons for this difficulty, (over)-dependence on the familiarity (Hasselgren, 1994) might clarify the issue for L2 learners in this context. So what L2 users feel sufficiently comfortable with might lead to them producing «lexical teddy bears». The results clearly support Hasselgren"s (1994) argument that the choices made by L2 users appear to «function like the L1 equivalent» (p. 256).

In order to minimize L1 interference on L2 collocation, the steps proposed by Yamashita and Jiang (2010, p. 663) can be taken into consideration for L2 classes: (1) understanding the lexical combination through its meaning en bloc; (2) frequent exposure to specific lexical combinations over time in the long-term memory; and (3) generating an association between a particular collocation and its conceptual representation. This suggests that collocational performance is not a component of language to be acquired quickly, so practitioners can accordingly create various opportunities using web-tools and designing tasks in L2 pedagogy, particularly in vocabulary teaching, to offer opportunities for L2 learners to follow these steps. This will ultimately make learners more aware of the conventional language use with respect to collocations.

We therefore support the idea that corpora-related activities seem to be the way forward for learners to be able to experience how particular words can «go hand-in-hand» with others to make meaning.

Acknowledgments

An earlier version of this paper was presented at the International Conference on Applied Linguistics (ICRAL 2018), Inönü University. We would like to thank Dr. Doğuş Can Öksüz, University of Cambridge for his valuable advice and constructive comments on the earlier version of the paper. We are also grateful to Roger Witts (UK) for proofreading our text and his language-related suggestions to the earlier version of the paper. The corpus of the paper compiled by the first author can be accessed via Akbaş, Erdem (2022). *Corpus of Turkish Undergraduate Writers of English (TUWE)* (Version 1) [Data set]. Mendeley Data. https://doi.org/10.17632/jhv9tgx2jy.1

CREDiT Authorship Contribution

Erdem Akbaş (First and corresponding author) Conceptualization, Resources, Methodology, Research, Formal Analysis, Writing-Original draft, Writing-Review & Editing; **Ceyhun Yükselir** (Co-author), Conceptualization, Research, Writing-Original draft, Writing-Review & Editing.

References

Ädel, Annelie, & Erman, Britt. (2012). Recurrent word combinations in academic writing by native and non-native speakers of English: A lexical bundles approach. *English for Specific Purposes,* 31(2), 81–92. https://doi.org/10.1016/j.esp.2011.08.004

Akbaş, Erdem (2022). *Corpus of Turkish Undergraduate Writers of English (TUWE)* (Version 1) [Data set]. Mendeley Data. https://doi.org/10.17632/jhv9tgx2jy.1

Altenberg, Bengt, & Granger, Sylviane. (2001). The grammatical and lexical patterning of MAKE in native and non-native student writing. *Applied Linguistics, 22*(2), 173– 195. https://doi.org/10.1093/applin/22.2.173

Bahns, Jens, & Eldaw, Moira. (1993). Should we teach EFL students collocations? *System*, *21*, 101–114. https://doi.org/10.1016/0346-251X(93)90010-E

Bardovi-Harlig, Kathleen., & Stringer, David. (2017). Unconventional expressions: Productive syntax in the L2 acquisition of formulaic language. *Second Language Research, 33*(1), 61–90. https://doi.org/10.1177/0267658316641725

Basal, Ahmet. (2019). Learning collocations: Effects of online tools on teaching English adjective-noun collocations. *British Journal of Educational Technology, 50*(1), 342-356. https://doi.org/10.1111/bjet.12562

Brezina, Vaclav; McEnery, Tony, & Wattam, Stephen. (2015). Collocations in context: A new perspective on collocation networks. *International Journal of Corpus Linguistics*, *20*(2), 139-173.

Chang, Yu-Chia; Chang, Jason S.; Chen, Hao-Jan, & Liou, Hsien-Chin. (2008). An automatic collocation writing assistant for Taiwanese EFL learners: A case of corpus-based NLP technology. *Computer Assisted Language Learning, 21*(3), 283-299. https://doi.org/10.1080/09588220802090337

Crosthwaite, Peter. (2017). Retesting the limits of data-driven learning: feedback and error correction. *Computer Assisted Language Learning, 30*(6), 447-473. https://doi.org/10.1080/09588221.2017.1312462

Daille, Beatrice. (1995). *Combined approach for terminology extraction: Lexical statistics and linguistic filtering*. Technical Report 5, UCREL, University of Lancaster.

Demirel, Elif Tokdemir, & Kazazoglu, Semin. (2015). The comparison of collocation use by Turkish and Asian learners of English: The case of TCSE corpus and ICNALE corpus. *Procedia-Social and Behavioral Sciences, 174,* 2278-2284. https://doi.org/10.1016/j.sbspro.2015.01.887

Durrant, Philip, & Schmitt, Norbert. (2009). To what extent do native and non-native writers make use of collocations? *International Review of Applied Linguistics in Language Teaching, 47*(2), 157-177. https://doi.org/10.1515/iral.2009.007

Durrant, Philip, & Schmitt, Norbert. (2010). Adult learners' retention of collocations from exposure. *Second Language Research, 26*(2), 163-188. https://doi.org/10.1177/0267658309349431

Eguchi, Masaki, & Kyle, Kristopher. (2020). Continuing to explore the multidimensional nature of lexical sophistication: The case of oral proficiency interviews. *The Modern Language Journal, 104*(2), 381-400. https://doi.org/10.1111/modl.12637

Fan, May (2009). An exploratory study of collocational use by ESL students – A task based approach. *System, 37,* 110-123. https://doi.org/10.1016/j.system.2008.06.004

Farghal, Mohammed, & Obiedat, Hussein. (1995). Collocations: A neglected variable in EFL. *International Review of Applied Linguistics, 33*, 315–331. https://doi.org/10.1515/iral.1995.33.4.315

Flowerdew, Lynne. (2015). Data-driven learning and language learning theories: Wither the twain shall meet. In A. Lenko-Szymanska & A. Boulton (Eds.), *Multiple affordances of language corpora for data-driven learning*, (pp. 15–36). John Benjamins. https://doi.org/10.1075/scl.69.02flo

Gablasova, Dana; Brezina, Vaclav, & McEnery, T. (2017). Collocations in corpus-based language learning research: Identifying, comparing, and interpreting the evidence. *Language Learning*, *67*(1), 155-179. https://doi.org/10.1111/lang.12225

Garner, Jamie.; Scott, Crosslet., & Kyle, Kristopher. (2018). Beginning and intermediate L2 writer's use of N-grams: an association measures study. *International Review of Applied Linguistics in Language Teaching, 58*(1), 51-74. https://doi.org/10.1515/iral-2017-0089

Gilquin, Gaëtanelle. (2015). From design to collection of learner corpora». In S. Granger, G. Gilquin, & F. Meunier (Eds.), *The Cambridge Handbook of Learner Corpus Research* (pp. 9-34). Cambridge University Press. https://doi.org/10.1017/CBO9781139649414.002

Girgin, Ufuk. (2019). The effectiveness of using corpus-based activities on the learning of some phrasal-prepositional verbs. *Turkish Online Journal of Educational Technology, 18*(1), 118-125.
Granger, Sylviane. (1998a). Prefabricated patterns in advanced EFL writing: Collocations and formulae. In A. P. Cowie (Ed.), *Phraseology: Theory, analysis, and application* (pp. 79–100). Oxford University Press.
Granger, Sylviane. (1998b). The computer learner corpus: A versatile new source of data for SLA research. In S. Granger (Ed.), *Learner English on Computer*, 3-18. Longman. https://doi.org/10.4324/9781315841342-1
Hasselgren, Angela. (1994). Lexical teddy bears and advanced learners: A study into the ways Norwegian students cope with English vocabulary. *International Journal of Applied Linguistics, 4*, 237-258. https://doi.org/10.1111/j.1473-4192.1994.tb00065.x
Henriksen, Birgit. (2013). Research on L2 learners' collocational competence and development: A progress report. In C. Bardel, C. Lindqvist, & B. Laufer (Eds.), *L2 vocabulary acquisition, knowledge and use: New perspectives on assessment and corpus analysis* (pp. 29-56). Eurosla
Hill, Jimmie. (2000). Revising priorities: From grammatical failure to collocational success. In M. Lewis (Ed.), *Teaching collocation. Further developments in the lexical approach* (pp. 47–69). Language Teaching Publications.
Howarth, Peter. (1998). The phraseology of learners' academic writing. In A. P. Cowie, (Ed.), *Phraseology: Theory, analysis, and applications* (pp. 161–186). Oxford University Press.
Hunston, Susan, & Francis, Gil. (2000). *Pattern grammar: A corpus-driven approach to the lexical Grammar of English*. John Benjamins. https://doi.org/10.1075/scl.4
Hoey, Michael. (2005). *Lexical priming: A new theory of words and language*. Routledge.
Jiang, Nan, & Nekrasova, Tatiana M. (2007). The processing of formulaic sequences by second language speakers. *Modern Language Journal, 91,* 433–445. https://doi.org/10.1111/j.1540-4781.2007.00589.x
Khantiwong, Waraphorn, & Thienthong, Atikhom. (2022). Thai EFL Learners' Knowledge of Congruent and Incongruent Academic L2 Collocations. *LEARN Journal: Language Education and Acquisition Research Network*, *15*(1), 809-835.
Kjellmer, Goran. (1990). A mint of phrases. In K. Aijmer, & B. Altenberg (Eds.), *English corpus linguistics: Studies in honour of Jan Svartvik* (pp. 111–127). Longman.
Klégr, Aleš, & Šaldová, Pavlína. (2006). Collocational Faux Amis. In F. Čermák & M. Šulc (Eds), *Kolokace* (pp. 168-177). NLN.
Koç, Gülay. (2006). *Developing collocational awareness* (Unpublished master»s thesis). Bilkent University.
Koosha, Mansour, & Jafarpour, Ali Akbar. (2006). Data-driven learning and teaching collocation of prepositions: The case of Iranian EFL adult learners. *Asian EFL Journal Quarterly, 8*(4), 192–209.
Laufer, Batia. (2011). The contribution of dictionary use to the production and retention of collocations in a second language. *International Journal of Lexicography, 24*, 29–49. https://doi.org/10.1093/ijl/ecq039
Laufer, Batia, & Waldman, Tina. (2011). Verb-noun collocations in second language writing: A corpus analysis of learners' English. *Language Learning, 61*(2), 647–672. https://doi.org/10.1111/j.1467-9922.2010.00621.x
Lewis, Michael. (2000). *Teaching collocation. Further developments in the lexical approach*. Language Teaching Publications.
Liou, Hsien-Chin, & Chen, Wen-Feng. (2018). Effects of explicit instruction on learning academic formulaic sequences for EFL college learners' writing. *Taiwan Journal of TESOL*, *15*(1), 61-100.
Millar, Neil. (2011). The processing of malformed formulaic language. *Applied Linguistics*, *32*(2), 129-148.
https://doi.org/10.1093/applin/amq035
Moreno, Ana. I. (2008). The importance of comparable corpora in cross-cultural studies. In U. Connor, E. Nagelhout, & W. Rozycki (Eds.), *Contrastive rhetoric: Reaching to intercultural rhetoric* (pp. 25-41). John Benjamins. https://doi.org/10.1075/pbns.169.04mor
Nesselhauf, Nadja. (2003). The use of collocations by advanced learners of English and some implications for teaching. *Applied Linguistics, 24,* 223–242. https://doi.org/10.1093/applin/24.2.223
Nesselhauf, Nadja. (2005). *Collocations in a learner corpus*. Benjamins. https://doi.org/10.1075/scl.14
O'Donnell, Michael. (2008, April). The UAM CorpusTool: Software for corpus annotation and exploration. In *Proceedings of the XXVI Congreso de AESLA*, Almeria, Spain.
Ordem, Eser, & Paker, Turan. (2016). Retention and use of lexical collocations (verb+ noun and adjective+ noun) by applying lexical approach in a reading course. *European Journal of Foreign Language Teaching, 1*(1), 144-160.
Paquot, Magali, & Granger, Sylviane. (2012). Formulaic language in learner corpora. *Annual Review of Applied Linguistics, 32*, 130–149. https://doi.org/10.1017/S0267190512000098
Pawley, Andrew, & Syder, Frances Hodgetts. (1983). Two puzzles for linguistics theory: Nativelike selection and nativelike fluency. In J. C. Richards, & R. W. Schmidt (Eds.), *Language and communication* (pp. 191-226). Longman.
Pérez-Paredes, Pascual. (2022). A systematic review of the uses and spread of corpora and data-driven learning in CALL research during 2011–2015. *Computer Assisted Language Learning, 35*(1-2), 36-61. https://doi.org/10.1080/09588221.2019.1667832
Saito, Kazuya. (2020). Mult-or single-word units? The role of collocation use in comprehensible and contextually appropriate second language speech. *Language Learning*, *70*(2), 548-588. https://doi.org/10.1111/lang.12387
Shin, Dongkwang, & Nation, Paul. (2008). Beyond single words: the most frequent collocations in spoken English. *ELT Journal, 62*(4), 339–348. https://doi.org/10.1093/elt/ccm091

Shin, Yu Kyoung, & Kim, YouJin. (2017). Using lexical bundles to teach articles to L2 English learners of different proficiencies. *System, 69*, 79-91. https://doi.org/10.1016/j.system.2017.08.002

Siyanova-Chanturia, Anna. (2015). Collocation in beginner learner writing: A longitudinal study. *System, 53,* 148-160. https://doi.org/10.1016/j.system.2015.07.003

Siyanova, Anna, & Schmitt, Norbert. (2008). L2 learner production and processing of collocation: A multi-study perspective. *Canadian Modern Language Review, 64*(3), 429-458. https://doi.org/10.3138/cmlr.64.3.429

Siyanova-Chanturia, Anna, & Spina, Stefania. (2020). Multi-word expressions in second language writing: A large-scale longitudinal learner corpus study. *Language Learning*, *70*(2), 420-463. https://doi.org/10.1111/lang.12383

Stubbs, Michael. (1995). Collocations and semantic profiles: On the cause of the trouble with quantitative methods. *Functions of Language, 2*(1), 1–33. https://doi.org/10.1075/fol.2.1.03stu

Ucar, Serpil, & Yükselir, Ceyhun. (2015). The effect of corpus-based activities on verb-noun collocations in EFL classes. *Turkish Online Journal of Educational Technology, 14*(2), 195-205.

Wang, Jiaji, & Halenko, Nicola. (2022). Developing the use of formulaic language for study abroad: a targeted instructional intervention. *The Language Learning Journal*, *50*(4), 409-426. https://doi.org/10.1080/09571736.2022.2088446

Wang, Ying, & Shaw, Philip. (2008). Transfer and universality: Collocation use in advanced Chinese and Swedish learner English. *ICAME Journal, 32*, 201-232.

Wood, David. (2010). *Formulaic language and second language speech fluency: Background, evidence and classroom applications*. Continuum.

Wray, Alison. (2002). *Formulaic Language and the Lexicon*. Cambridge University Press. https://doi.org/10.1017/CBO9780511519772

Wray, Alison. (2013). Formulaic language. *Language Teaching, 46*, 316–334. https://doi.org/10.1017/S0261444813000013

Wray, Alison, & Perkins, Michael. (2000). The functions of formulaic language: An integrated model. *Language and Communication, 20,* 1-28. https://doi.org/10.1016/S0271-5309(99)00015-4

Yamashita, Junko, & Jiang, Nan. (2010). L1 influence on the acquisition of L2 collocations: Japanese ESL users and EFL learners acquiring English collocations. *TESOL Quarterly, 44,* 647-668. https://doi.org/10.5054/tq.2010.235998