*ATLANTIS*
*Journal of the Spanish Association of Anglo-American Studies*
46.1 (June 2024): 243-264
e-ISSN 1989-6840
DOI: http://doi.org/10.28914/Atlantis-2024-46.1.13


# Constructional Emergence in A1-C1: A Bird's Eye View and Alignment

TAN ARDA GEDIK
FAU Erlangen-Nürnberg
tan.gedik@fau.de

This study aims to take a usage-based constructionist approach to the observation of the emergence of constructions in L2 speakers of English from a bird's eye perspective and align certain of them with Common European Framework of Reference for Languages (CEFR) levels. To do this, five equally balanced subparts from the EFCAMDAT corpus were compiled and analyzed using TAASSC and SPSS. The findings confirm previous studies in that they show strong evidence that speakers with lower proficiency levels use fixed or prototypical expressions and do not deviate as much as those with higher levels from conventional ways of combining constructions, i.e., collexemes. The ten highest and lowest frequency constructions do not show a developmental trajectory. Finally, the study presents a set of constructions aligned to each CEFR level which can serve as a rudimentary table of alignment.

Keywords: construction grammar; constructional learning; L2 writing; CEFR levels

...

# La emergencia constructiva en A1-C1. Panorámica global y alineación

Este estudio tiene como objetivo adoptar un enfoque construccionista basado en el uso a fin de observar la aparición de construcciones en hablantes de inglés como L2 desde una perspectiva global y alinear algunas de ellas con los niveles del Marco Común Europeo de Referencia para las lenguas (MCER). Para ello, se compilaron y analizaron cinco subpartes igualmente equilibradas del corpus EFCAMDAT utilizando TAASSC y SPSS. Los hallazgos confirman estudios anteriores en el sentido de que aportan pruebas sólidas de que las y los hablantes con niveles más bajos de competencia utilizan expresiones fijas o prototípicas y no se desvían tanto de las formas convencionales de combinar construcciones, es decir los colexemas, como quienes tienen niveles más altos de competencia. Las diez construcciones de mayor y menor frecuencia no muestran una trayectoria de desarrollo. Finalmente, el estudio presenta un conjunto de construcciones alineadas con cada nivel del MCER que pueden servir como tabla de alineación rudimentaria.

Palabras clave: gramática de la construcción; aprendizaje constructivo; escritura L2; niveles del MCER

## 1. Introduction

One of the central objectives of Second Language Acquisition (SLA) is understanding how language and increasing proficiency develops over time in a target language. As such, there have been many studies that have considered second language development from a variety of different perspectives and using various approaches. Recently, with more linguists subscribing to approaches that do not separate lexis from grammar, SLA has seen many studies on the lexicogrammatical development of L2 speakers of many different languages.

This study seeks to observe the emergence of certain constructions in L2 speakers of English across different proficiency levels and aligns them with the Common European Framework of Reference for Languages (CEFR; Council of Europe 2020) levels. As such, the current research study deepens our understanding of constructional learning and presents further evidence for constructional development in L2 learners. Subscribing to a usage-based constructionist approach, this study is similar in nature to those of Ellis (2008), Ellis and Ferreira-Junior (2009), Römer et al. (2014), Römer et al. (2018) and Römer and Yilmaz (2019), to name a few. However, one difference is that this study takes a more global perspective to those mentioned above, which focus on a specific set of verb-argument constructions in L2 speakers and their development across various CEFR levels. This paper is intended to be a complementary study to assist in our understanding of how constructions emerge across proficiency levels and what the implications of this are for SLA. The research hypotheses and research questions that are investigated in this study are:

H1: There will be correlations between syntactic indices and CEFR levels.

H2: Highly frequent constructions will have their roots in earlier proficiency levels.

H3: Low frequency constructions will show a developmental path as proficiency increases.

H4: As proficiency increases, attested constructions and lemma-construction combinations in the corpus will decrease.

RQ1: Are there clear-cut CEFR levels at which specific constructions start to emerge? If so, when and which constructions are they?

## 2. Usage-Based Theories

Within usage-based approaches, language is regarded as an emergent structure that arises in and through usage-events, i.e., repetition of sequences of words. By subscribing to an emergent understanding of language, a priori rules for language are no longer needed and thus language becomes an adaptive system, that is, a system which changes on the basis of ambient input (e.g., Hopper 1987; LarsenFreeman 1997; Ellis and LarsenFreeman 2006). Usage-based approaches have, in fact, almost become synonymous with cognitive linguistics, a result of the overlapping findings in the two fields. Another reason is that counting frequency of linguistic structures and investigating these frequency effects have been carried out by cognitive linguists, whose results have confirmed usage-based assumptions (see for instance Divjak 2019, 40-95). These assumptions are, namely, that a) language is learned through general domain cognitive abilities, for instance hearing, perception, pattern recognition and joint attention, to name a few (Tomasello 2003, 282-320); and that b) language arises from usage-events (e.g., Diessel 2016).

Construction grammar has been a successful theory in that there is ample evidence to support its assumptions and claims. While it is important to acknowledge that construction grammar is actually a family of theories, most of them agree on several central tenets (see Hoffmann and Trousdale 2013, 1-14 for a discussion). These are that: a) language consists of form-meaning pairings, i.e., constructions; b) constructions are learned; c) different deep and surface structures do not exist, in other words derivation is not applicable in construction grammar; d) constructions vary in shape, size and abstraction; and finally, e) constructions are usage-driven. In this study, we subscribe to a (usage-based) construction grammar which embodies the assumptions outlined above.

### 2.1. Construction Grammar

In a nutshell, construction grammar is a symbolic view of language that unifies form and meaning, where constructions are learned without the separation of lexis from grammar and there is no innate grammar regulatory system, i.e., Universal Grammar.

Figure 1. Lexicogrammatical Continuum (adapted from Gedik 2021, 30)

| Prefixes, suffixes | Words | Fixed expressions | Partially fixed expressions | Fully abstract schemas |
|---|---|---|---|---|
| *un-, re-, de-…* | *cat, dog, bird…* | *Here you are* | *The Xer, The Yer* | *Subj Verb Obj Oblique* |

Constructionists believe that there is no dividing line between lexis and grammar, rather that they merge to create what is called the lexicogrammatical continuum. This continuum (figure 1) signifies the gradience of linguistic items in that some items behave more like lexis and some display properties that would be regarded as *grammar*. What this continuum also shows is that any construction presented on the continuum is a pairing of form and meaning. For instance, *the Xer the Yer* construction is a partially filled construction, that is, it has fixed elements and slots, and this form sequence is paired with the meaning of correlationality. In other words, the intensity or probability of one event described in one clause is dependent on the other.

Many linguists believe that speakers start their language learning journey with idiomatic phrases or verb-islands (Tomasello 2003, 117-21) and keep detailed records of constructions: the items that occur with them and their lexicosemantic features as well as extra-linguistic conditions (Bybee 2010, 14-32). Frequency of exposure to language helps learners distinguish the conventionalized forms of constructions from their unconventionalized forms. Herbst (2020, 84) makes this explicit by saying "layers of usage events [...] become linked on the basis of recognized similarities between them." Focusing on entrenchment, Divjak (2019, 51; italics in the original) defines it as being "repeated presentations of a verb in particular constructions (e.g., *The rabbit disappeared*) [which] cause a child [learner] to infer probabilistically that the verb cannot be used in non-attested constructions (e.g., **The magician disappeared the rabbit*)." In Goldberg's (2019, 77) account, this type of entrenchment, where frequency is "simply a proxy for familiarity," is called simple entrenchment. Another, perhaps more important, type of entrenchment that needs attention is what Goldberg (2019, 77) calls conservatism via entrenchment, i.e., statistical preemption. This ability is activated when "the more frequently a verb has been witnessed in a language in any other construction, the more resistant it should be to being used in any new way" (Goldberg 2019, 77). In other words, speakers will calculate how many times an item and a construction should occur together based on the frequency information of the item and the construction and based on this information they arrive at a conclusion about the generalizability of an item. This, however, does not mean we retain all the item-specific information for a construction, since memory is lossy (Goldberg 2019, 73), but whenever we experience a construction it "can form a lossy structured representation that prioritizes what the word designates and includes various contextual aspects of the encounter" (Goldberg 2019, 16). For Goldberg (2019, 94), entrenchment also explains how "better-covered constructions are easier to access, which results in more conventional language being used more often, which further strengthens the association between conventional forms and particular messages-in-context." This is the reason why a positive correlation between increasing proficiency, that is mastery of target language constructions, and a greater accuracy of idiomatic speech is expected of L2 speakers of any language. Furthermore, it would be plausible to assume that there exists an inverse correlation between construction frequency, as in high and low frequency constructions, and proficiency. This stems from frequency effects, because arguably a speaker will only experience less common constructions as proficiency increases. In other words, a learner with growing proficiency is more likely to encounter more low frequency constructions, especially if the learner has written language as well as spoken language exposure in the target language. There is experimental evidence that L2 speakers with more print exposure outperformed those with less print exposure on tasks requiring vocabulary, collocation and grammatical knowledge, all of which require varying levels of constructional knowledge (Dabrowska 2019). The connection between print exposure and low frequency constructions stems from the fact that written language harbors more complex and subsequently rarer constructions (see Roland 2007 for English).

### 2.2. Constructional Knowledge in L1 and L2 Speakers

Various studies have shown that L2 learners of different languages do not differ in terms of their constructional knowledge (e.g., Römer et al. 2014) of constructions. In this sense, constructions have an ontological status for both L1 and L2 speakers and this indicates that grammar is just as meaningful as lexical items are. However, although there are many studies that analyze the constructional knowledge of L1 speakers (Lieven et al. 1997; Goldberg et al. 2004; Behrens 2009; Goldberg 2014; Ambridge and Lieven 2015, to name a few), studies that analyze L2 constructional knowledge are far fewer in comparison (Eskildsen 2012, 2014; Roehr-Brackin 2014; Tode and Sakai 2016). This difference can, however, be explained by the lack of reliable L2 corpora until recently (see Meunier 2015 on this).

Previous research demonstrates that L2 speakers of English have constructional knowledge, differ in their verb-VAC (verb-verb argument construction) associations depending on their proficiency and L1 background, and that there are systematic differences in their usage of certain constructions (Gries and Wulff 2005; Römer et al. 2014; Römer et al. 2018). However, research of a general nature, i.e., that does not subscribe to a particular L1 background, is relatively scarce. Römer (2019) is one exception to this. She investigates the constructional development of Mexican and German speakers of English from A1 through C1 CEFR levels, using the same corpus as used in the present study. Her analyses provide a detailed observation of how constructions develop at certain levels with specific items in two different learning groups. The analyses conclude that learners differ in their item-specific usage of certain constructions but become more productive over time. A similar study was conducted by Römer and Berger (2019) that compared the same learner groups sampled from the same corpus whereby they specifically focus on a list of prepositional constructions (i.e., V about N, V across N and others, see Römer and Berger 2019, 1095). Their analyses also show that there is a correlation between growing proficiency and productivity as well as an inverse correlation between growing proficiency and using fewer fixed expressions. This productivity may partially lead to unconventional attestations of a given construction, even at advanced levels. For instance, both Goschler and Stefanowitsch (2023) and Gedik and Uslu (2023) provide evidence for advanced German and Turkish speakers, respectively, showing that at times advanced L2 speakers' choice for the verbal slot in a ditransitive construction may be affected by strongly entrenched items in the corresponding construction in the target language. Although this may not always be the case, it is still possible and therefore it is plausible to hypothesize that with growing proficiency the number of corpus-attested item-construction combinations decreases.

It is, however, important to note that so far these studies have compared specific learner groups and have not involved a sample with learners from different linguistic backgrounds. Furthermore, there have not been many studies that systematically attempt to align CEFR levels to specific constructions. This is understandable as it is quite a tedious task to do. McCarthy (2016) is the only study that partially aligns the ditransitive construction, i.e., *she gave him a book*, with CEFR levels, and this is done by analyzing error rates and data from learner corpora. He notes that the ditransitive construction is mastered by C1, i.e., it is produced with no errors. However, he does not necessarily claim that the ditransitive construction belongs to a specific CEFR level, and by that token, that construction would be an A1 level construction.

Scholars have demonstrated that L2 speakers start their language learning journey with a set of fixed and highly repetitive constructions, just as L1 speakers do, which then grows in complexity, productivity and becomes less fixed (Eskildsen and Cadierno 2007; Eskildsen 2009; Li et al. 2014). Studies also suggest that with increasing proficiency, the accuracy of constructional knowledge also increases (Crossley and Salsbury 2011; Bestgen and Granger 2014). In addition to this, there is evidence that L2 speakers' knowledge of constructions is also influenced by their L1 (Li et al. 2014; Eskildsen et al. 2015; Römer and Yilmaz 2019; Gedik and Uslu 2023; Goschler and Stefanowitsch 2023). There is strong evidence that advanced L2 speakers are also influenced by strongly entrenched verb-VAC combinations in their L1 (Gedik and Uslu 2023; Goschler and Stefanowitsch 2023).[1]

In an experiment, Lee and Kim (2011) tested Korean speakers' knowledge of the English intransitive construction, the ditransitive and resultative constructions, developmentally. They explain that Korean speakers of English did not show a developmental understanding of the intransitives. Put simply, the speakers did not start from the bottom of a taxonomical constructional family and construct the superordinate intransitive construction. Their performance on the ditransitive and the resultatives also varied, with most participants finding them difficult. For instance, as the authors also argue, constructions that are similar across the two languages may be learned faster in a specific learner group in comparison to others because the L1 in that case may act as a training crutch. This arguably shows that both L1 and also other

[1] VAC is another name used to refer to constructions in natural language processing and L2 constructional acquisition studies, i.e., the caused-motion or the ditransitive constructions. To keep terminological confusion to a minimum, we also use VAC to refer to constructions in this study.

personal factors can contribute to these variations. Personal factors include but are not limited to working memory capacity, attention and motivation, to name a few (see Sparks 2022 for a detailed review).

FIGURE 2. Two VACs from TAASSC

| VACs | Examples |
|---|---|
| Nsubj_verb_dobj (the transitive construction) | *I cook dinner* |
| Modal_nsubj_verb_xcomp (the modal-question construction) | *Would you consider applying?* |

## 3. Methodology and Association Measures

The study was carried out using the EFCAMDAT corpus in late 2021. In what follows, information about the corpus, the subcorpora created for this study and the association measures used are introduced. The EFCAMDAT corpus is based on texts that were submitted by users of a website, which was previously known as Englishtown (now englishlive.com), where learners of English had to take a placement test to enroll in the courses offered by Education First online language school. The placement test would place students in one of the sixteen proficiency levels available, all of which were aligned with CEFR levels. Students were regularly given writing tasks, such as producing an email, a movie review and introducing themselves, to name just three (Alexopoulou et al. 2015). In all, 128 such tasks are examined in the corpus.

In the texts compiled here, as there was no data available for C2, the level had to be discarded. Each subcorpus was roughly a combination of 10,000 writing samples, all of which were balanced out across different writing tasks and a variety of topics to improve the reliability of the results. However, keeping frequency counts equal was difficult, as with increasing proficiency there were more words per sample. Nevertheless, for each level, there was data from four different, randomly selected and equally balanced tasks.

In this study, we compiled five well-balanced subcorpora from the Education First-Cambridge Open Language Database (EFCAMDAT; Geertzen et al. 2013; Alexopoulou et al. 2015), which is a large corpus of written texts by L2 speakers of English from different linguistic backgrounds. Ranging from A1 through C1, each subcorpus had roughly one million words (see table 1 for a detailed overview) and included writing samples for different tasks that were given to the students. The subcorpora were created as .txt files and were split up into 500 individual .txt files and processed in batches of 50. This was necessary as TAASSC seems to have a limit on how many words it can process in one file (Kristopher Kyle, personal communication, February 8, 2022). The text files were analyzed using the syntactic sophistication setting of the tool, with minimum VAC frequency set to 5. When selecting indices to compare against the reference corpora, ALL_COCA was selected, as in the case of this dataset speakers did not have to use specialized language, i.e., academic or journalistic. As such, a combination of all of the subsections of COCA makes language representative of all of its special uses, and it is assumed that speakers learn all of them in a piecemeal fashion. The indices were then imported into SPSS for a multiple regression analysis. Indices that did not meet the assumptions of a multiple regression analysis, e.g., due to collinearity and violations of normality, were discarded (Tabachnik and Fidell 2014).[2] In the end, there were 15 corpus-based indices used, which were calculated based on the subcorpus under analysis and against-reference-corpus, which is the COCA.

Table 1. Subcorpus Word Count

| CEFR Level | Word count |
|---|---|
| A1 | 1142862 |
| A2 | 1217936 |
| B1 | 1202052 |
| B2 | 1264006 |
| C1 | 1305818 |

### 3.1. Association Measures and Automatized Tools

It is possible to measure constructional frequency in finer detail. To date, three main approaches have been employed in studies: a) faith scores; b) delta p scores; and c) collostructional strength. However, due to space related issues, only c) will be explained here.

Collostructional analysis (Gries and Stefanowitsch 2003) predicts the likelihood of two items from the corpus appearing next to one another. Kyle and Crossley (2017, 525) employ the following formula in TAASSC, which is slightly different to the original formula, in order to calculate collostructions (Gries and Stefanowitsch 2003) as it is computationally lighter, and the authors claim that it is perfectly compatible with the original formula: $\left(\left(\frac{a}{a+b}\right) - \left(\frac{c}{c+d}\right)\right) * \ (a+b)$. This formula gives the output $a{+}b{+}c{+}d$ for "approximate collexeme strength" (Kyle and Crossley 2017, 525).

The tool used in this study is TAASSC, developed by Kristopher Kyle (2016). The tool automatically analyzes given texts in relation to several syntactic measures. However, the measure used in this study is syntactic sophistication, which calculates usage-based indices and automatically detects VACs. Alongside the above-mentioned association measures,

[2] To detect collinearity, coefficients and Variance Inflation Factor (VIF) scores were used; Shapiro-Wilk test (the p values) was used to determine normality.

the tool also calculates the approximate percentage of constructional and lemma coverage in texts against a reference corpus, namely the COCA (Davies 2010), and produces type-token ratios for constructions, lemmas and lemma-construction combinations, among many other syntactic indices.[3] The tool has been successfully used in quite a few studies (Kyle and Crossley 2017; Kyle and Crossley 2018; Gedik 2021 to name a few) and so was selected as a means of automatically detecting and calculating constructional data.

## 4. Results and Analysis

Out of 35 indices initially considered, 20 had to be discarded as they violated the assumptions of the analysis, i.e., normality and collinearity. In tables 2 and 3, the descriptive statistics for the remaining indices are reported. See the supplementary material for detailed descriptive statistics for each CEFR level and index. See the appendix for the descriptive statistics of corpus-based and against-corpus indices.

Table 2. Correlations for Corpus-Based Indices

| **Correlations (corpus-based)** | | | | | |
|---|---|---|---|---|---|
| | | LF | CF | LFC_per_mil | Collexeme |
| Pearson Correlation | CEFR | -.394 | .108 | -.115 | -.200 |
| Sig. (1-tailed) | CEFR | .000 | .000 | .000 | .000 |

Lemma-frequency (LF) is calculated based on the items in the verbal slot of the constructions. The verbs that appear in the slot are then compared against their actual corpus frequencies from the reference corpus of TAASSC. Thus, LF shows the usage of high- or low frequency verbs in the verbal slot of the constructions. Construction-frequency (CF) works in the exact same way as LF, but for VACs identified in the corpus and the reference corpus. LFC_permil (lemma-construction frequency) is the calculation of verb-VAC combination frequency per million, i.e., normalized. In other words, it calculates how often, for example, *give* occurs in the ditransitive in the corpus, and then compares it against the reference corpus. The end result shows whether these verb-VAC combinations are highly frequent (common) or not (rare), e.g., *I gave him a book* vs *I begrudged him his affluence*, respectively. Finally, collexeme, also known as collostructional analysis (Gries and Stefanowitsch 2003), calculates the joint probability that two items in a corpus will co-occur. Collexeme can give insight into the prototypicality or novelty of verb-VAC combinations in a corpus.

[3] If the link breaks, the same file is hosted at TAASSC - NLP TOOLS FOR THE SOCIAL SCIENCES (linguisticanalysistools.org).

TABLE 3. Correlations for Against-Reference-Corpus Indices

| Correlations (against-reference-corpus) | | all_av_lemma_freq | all_av_construction_freq | all_av_lemma_construction_freq | all_av_approx_collexeme | all_lemma_ttr | all_construction_ttr | all_lemma_construction_ttr | all_lemma_attested | all_construction_attested | all_lemma_construction_attested |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Pearson Correlation | CEFR | -.693 | -.332 | -.706 | -.279 | .853 | .829 | .850 | .163 | -.541 | -.725 |
| Sig. (1-tailed) | CEFR | .000 | .000 | .000 | .000 | .000 | .000 | .000 | .000 | .000 | .000 |

All_av_lemma_freq compares lemma frequencies against the lemma frequencies in the reference corpus, while all_av_construction_freq works in the same way, but for VAC frequencies. All_av_lemma_construction_freq also works in the same way as the previous two indices, but it calculates frequencies for lemma-construction combinations. All_av_approx_collexeme calculates the average collostructional strength in the corpus against the reference corpus. All_lemma_ttr calculates the main verb lemma type-token ratio against the reference corpus for the corpus itself. All_construction_ttr does the exact same calculation as the previous index, but for constructions. Similarly, all_lemma_construction_ttr does the same calculation but for lemma-construction combinations. All_lemma_attested is the percentage of lemmas in the corpus that are in the reference corpus. All_construction_attested is the same ratio calculated for constructions. All_lemma_construction_attested is the same ratio but this time it is calculated for lemma-construction combinations.

### 4.1. Hypothesis 1: Correlations between Indices and CEFR Levels

TAASSC first identifies part-of-speech (POS) tags, and then establishes dependencies to identify constructions (see Kyle 2016, 35-43 for a detailed discussion). As reported in previous studies, e.g., Römer and Berger (2019), there were correlations between CEFR levels and lemma-frequency (r= -.394), construction-frequency (r= .108), lemma-construction frequency per million (r= -.115) and collexemes (r= -.200). The statistically significant findings (*p= .00*) suggest that the correlation between proficiency and the various indices is real. The weak correlation shows that the contribution of proficiency among other indices is relatively smaller. To interpret the practical effects of proficiency on these indices we can examine the $r^2$ scores (LF $r^2$ = .155, CF $r^2$ = .012, LFC_permillion $r^2$ = .013, collexemes $r^2$ = .040). As such, an increasing proficiency would account for 15.5% of the variance in lemma-frequency, 1.2% in construction-frequency, 1.3% in lemma-construction frequency per million and 4% in collexemes in learner corpora. Although the effect sizes are small, these correlations can be interpreted as follows. As proficiency increases, students' use of common or highly frequent verbs decreases. This has been reported as being a predictor of being a proficient writer, in that it signifies that lexical diversity effectively increases (e.g., McCarthy and Jarvis 2010). CF also shows a positive correlation and this indicates that learners use a greater variety of constructions as proficiency increases. This is not surprising, as previous studies also report similar findings of low proficiency students using a set of fixed constructions (see Römer 2019). LFC per million supports the findings here regarding CF and LF, and suggests that on average, as proficiency increases, students gradually decrease their use of highly frequent lemma-construction combinations. Per million in this index is computed based on the corpus analyzed in the study and not against a reference corpus. Finally, there seems to be an inverse correlation between increasing proficiency and lemma-construction combinations, i.e., collexemes. This indicates that speakers use verbs that are less attracted to the constructions and as such they move away from formulaic and fixed expressions towards a more varied writing vocabulary. To see if this increase or decrease for the above-mentioned indices holds across all CEFR levels, a multivariate analysis with the contrast option (K Matrix) was run. The contrast option helps with contrasting the findings for each variable across independent variables, i.e., CEFR levels. Table 4 shows CEFR level comparisons, all of which show statistical significance except for LFC_permillion from A2 onward. The p-values become clearer in light of descriptive statistics for each index (see the appendix). LF drastically decreases until B2, where it increases again, but then nosedives once more at C1 and hits an all-time low. The increase at B2 may be due to task requirements that encouraged learners use more high frequency words. CF increases until A2 and then decreases until B2 and reaches its all-time high at C1.

TABLE 4. K Matrix Contrast Results

| | A1 vs. A2 | A2 vs. B1 | B1 vs. B2 | B2 vs. C1 |
|---|---|---|---|---|
| LF (sig. 1-tailed) | .000-.693 | .000-.332 | .000-.706 | .000-.279 |
| CF (sig. 1-tailed) | .000 | .000 | .000 | .000 |
| LFC_permillion (sig. 1-tailed) | .000 | .354 | .877 | .016 |
| Collexeme (sig. 1-tailed) | .000 | .000 | .000 | .000 |

Turning our attention to the indices in table 4, which compare the findings in the corpus against the COCA, first a description of each is necessary. Those that end in _freq are frequency-based indices, those with _ttr indicating the type-token ratio and those with _attested referring to how many items in the corpus analyzed here exist in the reference corpus. All indices are statistically significant (*p*= *.00*).

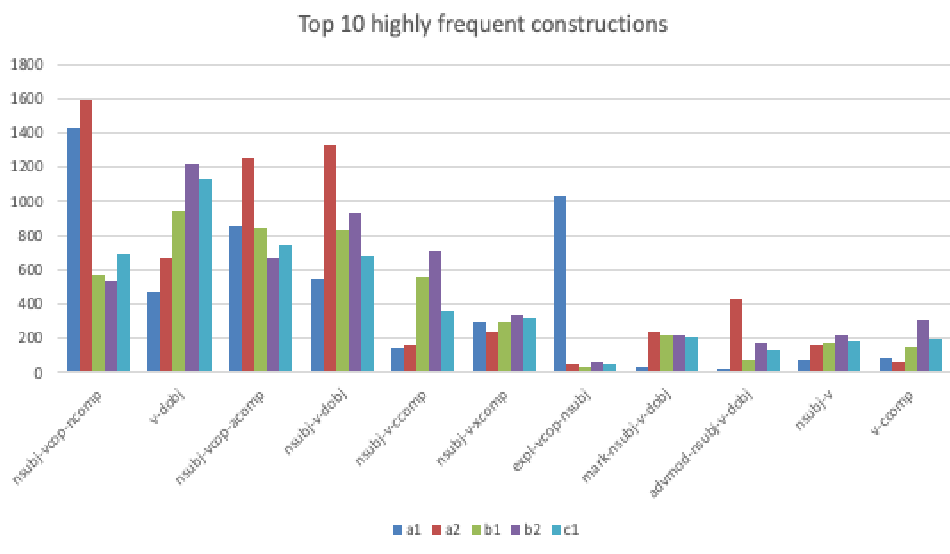
1) *All_av_lemma_freq* (r= -.693), *all_av_construction_freq* (r= -.332), *all_av_lemma_construction_freq* (r= -.706) and *all_av_approx_collexeme* (r= -.279) show inverse correlations. The strongest correlation seems to be between lemma, lemma-construction and CEFR levels. This suggests that on average, as proficiency increases, speakers move away from formulaic, fixed and highly repetitive lemmas and lemma-construction combinations. This supports the previous findings (Goldberg 2006, 45-65; Römer 2019). As for *approx_collexeme*, the results support the finding that with improving proficiency, students use less prototypical verbs in constructions.
2) *All_lemma_ttr* (r= .853); *all_construction_ttr* (r= .829) and *all_lemma_construction_ttr* (r= .850) demonstrate that with increasing proficiency, the type-token ratio per lemma, construction and lemma-construction combinations decrease. This indicates a more varied, or perhaps a lexicogrammatically richer, output. As such, it is possible to argue that construction learning occurs via experience with highly repetitive and prototypical examples at earlier stages (e.g., Goldberg 2006, 69-92), i.e., *I gave him a book* serves as a prototypical example for the ditransitive construction.
3) *All_lemma_attested* (r= .163), *all_construction_attested* (r= -.541) and *all_lemma_construction_attested* (r= -.725) results suggest that as CEFR levels increase, speakers' production of attested items decreases and attested lemmas data suggests that as proficiency increases, there are fewer attested lemmas. When descriptive statistics are taken into account, while the first two do not demonstrate a

considerably big leap between CEFR levels, *all_lemma_construction_attested* does. This could suggest that speakers are acquiring productivity, resulting in highly schematic constructions, similar to the findings of Römer (2019) in relation to productivity. Thus, productivity may be an important factor to investigate in future L2 studies, especially in relation to producing unconventional constructions.

## 4.2. Hypothesis 2: Highly Frequent Constructions in Early CEFR Levels

With approximately 1,193 constructions identified in total across levels, the frequency cut off for H2 and H3 was identified by using the *CF_per_million* data in the results files provided by TAASSC. As such, high frequency constructions were those which appeared >50000 per million and low frequency constructions were <1000 times per million (the vertical axis in figure 3 represents per million). To reiterate, per million is calculated based on within text data, not against a reference corpus. The ten most frequent constructions are shown in figure 3.

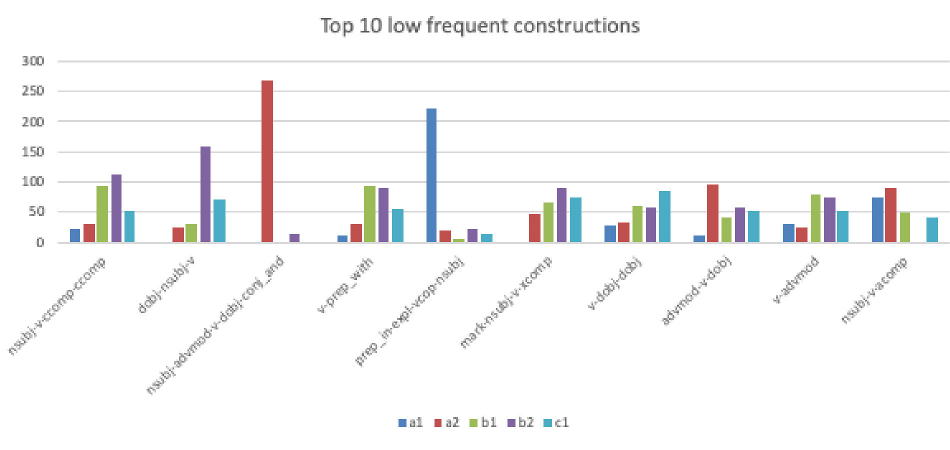
FIGURE 3. Top Ten Highly Frequent Constructions




As seen in figure 3, there does not seem to be a clear increasing trend in frequency of use per proficiency for any of the top ten constructions. As such, this hypothesis, that is, highly frequent constructions stem from earlier CEFR levels, is not confirmed. This may still be the case, although it may not be possible to capture using just corpus evidence. It may need to be triangulated with experimental studies investigating language aptitude in L1 and L2, working memory, print exposure and phonological abilities in L1. As for the idiosyncrasy of these constructions across CEFR levels, it is difficult to explain why that is the case. However, one possible explanation might be individual differences in grammatical knowledge (see Sparks 2022 for a detailed discussion). The assumption would be that L2 learners from a particular L1 would have constructions entrenched at different levels than speakers of other L1s at the same L2 proficiency level. These different levels of entrenchment stem not only from differences in exposure, but also individual differences in the cognitive machinery in speakers' L1 abilities. Sparks (2022) outlines fifty years' worth of research analyzing how having good phonological abilities in the L1 significantly predict L2 success or L2 aptitude. Thus, while corpora, which we take as a pseudo-measure of quantifying exposure, can account for a lot, this is not the only factor that needs to be taken into account (see Dabrowska 2016 on why individual differences should be taken into account in linguistic inquiry). Another possibility might be the communicative requirements of certain tasks, from which the subcorpora in the current study were sampled. These requirements may not have allowed for the need to use these constructions.

### 4.3. Hypothesis 3: Low Frequency Constructions in Later CEFR Levels

FIGURE 4. Ten Lowest Frequency Construction




Except for v-dobj-dobj, the other constructions do not follow a trend where constructions become more frequent as proficiency increases. Even then, the differences between levels are not statistically significant and there is a slight decrease between B1 and B2. Therefore, the hypothesis that low frequency constructions emerge in later CEFR levels is not confirmed. Once again, it is difficult to ascertain the idiosyncrasy of these constructions. It could be due to the communicative requirements of the tasks. Another explanation might be the fact that low frequency constructions possibly do emerge at earlier CEFR levels due to exposure in their lexically prefabricated instances. In other words, a low frequency construction may already be used by an A2 or B1 level student with a highly frequent item in the verbal slot of the construction. But the semanto-pragmatic features of the said construction may not be acquired fully or the construction is possibly not schematized until later CEFR levels.

See an example for v-dobj-dobj:

(1) v-dobj-dobj (*get*$_{\_v}$ *$ 50,000 salary*$_{\_dobj}$ *and 1 month holiday*$_{\_dobj}$)

### 4.4. Hypothesis 4: Attested Constructions and Lemma-construction Combinations

These findings, especially _attested indices combined with _ttr, confirm that students start learning their L2 with a set of limited lemmas and constructions, which then expands to incorporate greater variety. The change in *all_construction_attested* and *all_lemma_construction_attested* suggests that there is a probability that the high repetition of frequently-used construction and lemma-construction combinations helps with suppressing creativity or productivity for those who are at lower proficiency levels, which guides them to use more idiomatic and fixed phrases, i.e., lexically-prefabricated chunks. This is arguably because at lower levels learners may not have schematized constructions that are highly productive. With partial evidence from the indices in the present study and previous studies (Gedik and Uslu 2023; Goschler and Stefanowitsch 2023), H4 is confirmed. Increasing proficiency fosters productivity but this extension may be affected by certain factors, for example strongly entrenched items in the corresponding construction in the L1 (Goschler and Stefanowitsch 2023), and may potentially result in unattested instantiations of constructions, e.g., *I explain you the book*.

### 4.5. Aligning Constructions to CEFR Levels

In this section, constructions will be filtered to align with CEFR levels, using raw frequencies, to identify when they first emerge. In other words, how many times a construction was produced at a certain level. It is important to know that this analysis is not clear-cut and may differ based on many cognitive or frequency related factors in different L2 speakers of English. It is meant to serve as a rudimentary analysis of general tendencies in this study. For that reason, the 50 most frequent constructions will be aligned with CEFR levels. A construction was considered to have emerged in a specific level if it had not occurred more than 50% of the raw frequency in the previous level than that of the current level. Similarly, constructions that emerge in one level and continue occurring in other levels were eligible. Those that did not meet the criteria were not added to the list. The examples were taken from the respective levels.

TABLE 5. Constructional Alignment

| CEFR Levels | Constructions (raw/normalized frequency) |
|---|---|
| A1 | Expl-vcop-nsubj (A1: 1,182/1035; A2: 55/66.98): *there is a goat*<br>Expl-vcop-nsubj-nsubj (A1: 316/361.14; A2: 10/12.17): *There are three windows and a chair*<br>Prep_in-expl-vcop-nsubj (A1: 220/251.43; A2: 18/21.92): *In my office, there are many people* |
| A2 | Nsubj-v-dobj (A1: 549/627; A2: 1333/1623.51): *I have a daughter*<br>Mark-nsubj-vcop-acomp (A1: 16/18.28; A2: 215/261.85): *so I am busy*<br>V-prep_at (A1: 0/0; A2: 71/86.47): *study at*<br>Expl-vcop-ncomp (A1: 0/0; A2: 67/81.60): *There is a kitchen*<br>Nsubj-v-prep_to (A1: 0/0; A2: 80.38): *I go to Canada* |
| B1 | V-dobj-prep_on (A2: 0/0; B1: 119/143.04): *take me on a holiday*<br>Nsubj-v-ccomp-ccomp (A2: 29/35.32; B1: 215/76): *I hope I will understand and speak this language*<br>Nsubjpass-v (A2: 0/0; B1: 48/57.69): *the girl was hit*<br>Mark-nsubj-v-dobj-xcomp (A2: 8/9.7; B1: 16/19.23): *Because I try my best to learn* |
| B2 | Advmod-nsubj-v-prep_for (B1: 0/0; B2: 70/88.48): *Also I waited for the exam*<br>Nsubj-advmod-v (B1: 0/0; B2: 52/65.72): *I can also watch*<br>Dobj-nsubj-v-xcomp (B1: 15/18.03; B2: 37/46.76): *that he tried to learn*<br>Nsubj-v-iobj-dobj (B1: 13/15.62; B2: 36/45.50): *He gave her a box*<br>Mark-nsubj-vcop-xcomp (B1: 5/6.01; B2: 32/40.44): *Since she is to go* |
| C1 | V-dobj-prepc_by (B2: 0/0; C1: 39.17): *had a dream by creating*<br>V-prep_into (B2: 0/0; C1: 29/37.86): *got me into*<br>Mark-dep-v-dobj (B2: 13/16.43; C1: 28/36.56): *in order to know more friends*<br>Mark-nsubj-v-prep_for–ccomp (B2: 0/0; C1: 27/35.25): *If you vote for me, I will appeal*<br>V-prt-prep_on (B2: 0/0; C1: 27/35.25): *followed up on this*<br>Mark-nsubj-v-dobj-prep_as (B2: 0/0; C1: 26/33.95): *Whether they learned it as a second language*<br>Nsubj-v-prep_on-prep_for-mwe-prep_with (B2: 0/0; C1: 28.72): *I hope I can count on you for support because with your help we can make a difference* |

The constructions presented here are of varying specificity and abstractness. Table 5 demonstrates that with increasing proficiency, the use of more arguably complex constructions increases. That is, each level has a new construction in comparison to the previous one, except for A1 and A2, and this involves a subordinating word: for B1 this is *because*, for B2 *that* and *since* and for C1 *in order to* and *whether*. This classification does not necessarily mean that learners do not use such subordinating constructions at earlier levels, but in the current study the classification arises on the basis of the data at hand. Therefore, a more robust and valid attempt at classifying constructions per CEFR levels should triangulate data from different corpora and also include other indices apart from frequency only.

## 5. Conclusions

With growing proficiency, speakers expand their constructional knowledge and combine constructions with different items. Furthermore, as we have seen in frequency-based indices, (lower level) speakers mostly use fixed expressions and stay relatively loyal to them, though they increasingly depart from this as proficiency level increases.

Starting with the hypotheses, in H1, we reconfirmed some of the findings of previous studies, namely that there is a statistically significant correlation, especially considering the size of the subcorpora here, between selected syntactic indices and CEFR levels. More specifically, there is evidence that speakers with growing proficiency move away from fixed and highly repetitive expressions to include more lower-frequency lemma-construction combinations, which is backed up by the data from LFC per million and collexeme indices. Furthermore, when the writing samples are compared against a reference corpus, there is once again strong evidence that speakers expand their mental constructicon and the exemplar representations of lemma-construction combinations, which is supported by point 3 in section 4.1. Furthermore, there is evidence that as speakers gain more proficiency, the type-token ratios show a positive trend, meaning that output becomes lexicogrammatically more diverse. This provides partial evidence for how constructions are learned through highly repetitive and prototypical examples at earlier stages of proficiency. We rejected H2 and H3 because of lack of evidence, that is, high frequency constructions do not necessarily stem from earlier CEFR levels and low frequency constructions do not emerge in later CEFR levels. Finally, we confirmed H4 as a natural consequence and byproduct of H1, namely that as proficiency increases there will be a lower score for attested constructions in the texts against a reference corpus.

As such, then, the findings presented here confirm the central tenets of usage-based approaches. Namely, that language is an experience-based phenomenon and that constructions are first learned with high-frequency items being used in them, but are then expanded to incorporate other lower-frequency items. The indices confirm the findings of previous studies and present evidence that constructions are learned in a piecemeal fashion and language learners use the highly repetitive and fixed expressions as training wheels (or item-islands) to be able to incorporate and learn more constructions at higher proficiency levels and acquire more productivity, i.e., the constructions become more highly schematic. These findings can inform the teaching of foreign languages in the following ways:

a) repetition and recycling of constructions is important at earlier levels;
b) speakers will not always show developmental behavior with all constructions; and

c) presenting learners with highly repetitive and prototypical examples of constructions should ideally be of help in learning constructions, especially at earlier stages.

Future studies should be encouraged to take up this line of research and explore some of the shortcomings of this study, for instance the number of words in each subset and low number of constructions analyzed in constructional alignment and hypotheses.

The current study has presented a bird's eye view and insight into the implications of constructional emergence at different proficiency levels and aligned them with CEFR levels. As such, the findings suggest and confirm that there is a correlation between usage-based syntactic indices and CEFR levels, that language is learned in a piecemeal fashion and speakers use highly repetitive fixed expressions, which can range anywhere from small constructions such as collocations to argument structure rules such as the transitive construction [*nsubj-v-dobj*], but that with growing proficiency they move towards an expanded mental constructicon. Finally, some constructions were aligned with CEFR levels and compared against the expected skills of respective levels. This alignment suggests that speakers' knowledge of constructions partially follows a developmental sequence, with previous constructions combined with newer ones. The findings may prove useful for future SLA studies and applied construction grammar, where the teaching of constructions is concerned.

## Works Cited

Alexopoulou, Theodora et al. 2015. "Exploring Big Educational Learner Corpora for SLA Research: Perspectives on Relative Clauses." *International Journal of Learner Corpus Research* 1 (1): 96-129.

Ambridge, Ben and Elena Lieven. 2015. "A Constructivist Account of Child Language Acquisition." In MacWhinney and O'Grady 2015, 478-510.

Behrens, Heike. 2009. "Usage-Based and Emergentist Approaches to Language Acquisition." *Linguistics* 47: 383-411.

— and Stefan Pfänder, eds. 2016. *Experience Counts: Frequency Effects in Language.* Berlin: De Gruyter.

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.

Bybee, Joan. 2010. *Language, Usage and Cognition*. Cambridge: Cambridge UP.

Cadierno, Teresa and Søren Wind Eskildsen, eds. 2015. *Usage-Based Perspectives on Second Language Learning*. Berlin: De Gruyter.

Council of Europe. 2020. *Common European Framework of Reference for Languages: Learning, Teaching, Assessment: Companion Volume*. Strasbourg: Council of Europe Publishing.

—. n.d. "English Grammar Profile Online." *English Profile.* [Accessed January 19, 2022].

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 in Language Teaching* 49 (1): 1-26.

DĄBROWSKA, Ewa. 2016. "Cognitive Linguistics' Seven Deadly Sins." *Cognitive Linguistics* 27 (4): 479-91.

—. 2019. "Experience, Aptitude, and Individual Differences in Linguistic Attainment: A Comparison of Native and Nonnative Speakers: Experience, Aptitude, and Individual Differences." *Language Learning* 69 (1): 72-100.

DAVIES, Mark. 2010. "The Corpus of Contemporary American English as the First Reliable Monitor Corpus of English." *Literary and Linguistic Computing* 25 (4): 447-64.

DIESSEL, Holger. 2016. "Frequency and Lexical Specificity in Grammar: A Critical Review." In Behrens and Pfänder 2016, 209-38.

DIVJAK, Dagmar. 2019. *Frequency in Language: Memory, Attention and Learning*. Cambridge: Cambridge UP.

ELLIS, Nick. 2008. "The Dynamics of Second Language Emergence: Cycles of Language Use, Language Change, and Language Acquisition." *The Modern Language Journal* 92 (2): 232-49.

— and Fernando Ferreira-Junior. 2009. "Construction Learning as a Function of Frequency, Frequency Distribution, and Function." *The Modern Language Journal* 93 (3): 370-85.

— and Diane Larsen-Freeman. 2006. "Language Emergence: Implications for Applied Linguistics. Introduction to the Special Issue." *Applied Linguistics* 27 (4): 558-89.

ESKILDSEN, Søren W. 2009. "Constructing Another Language: Usage-Based Linguistics in Second Language Acquisition." *Applied Linguistics* 30 (3): 335-57.

—. 2012. "L2 Negation Constructions at Work." *Language Learning* 62 (2): 335-72.

— and Teresa Cadierno. 2007. "Are Recurring Multi-Word Expressions Really Syntactic Freezes? Second Language Acquisition from the Perspective of Usage-Based Linguistics." In Nenonen and Niemi 2007, 86-99.

—, Teresa Cadierno and Peiwen Li. 2015. "On the Development of Motion Constructions in Four Learners of L2 English." In Cadierno and Eskildsen 2015, 207-32.

GEDIK, Tan A. 2021. "An Analysis of Lexicogrammatical Development in English Textbooks in Turkey: A Usage-Based Construction Grammar Approach." *Explorations in English Language and Linguistics* 9 (1): 26-55.

— and Fatıma Uslu. 2023. "L1-L2 Transfer in Ditransitive Construction: A Usage-Based Replication Study with Turkish Speakers of English." In Georgiou, Giannakou and Savvidou 2023, 123-46.

GEERTZEN, Jeroen, Theodora Alexipoulou and Anna Korhonen. 2013. "Automatic Linguistic Annotation of Large Scale L2 Databases: The EF-Cambridge Open Language Database (EFCAMDAT)." In Prior et al. 2013, 240-54.

GEORGIOU, Georgios P., Aretousa Giannakou and Christine Savvidou, eds. 2023. *Advances in Second/Foreign Language Acquisition*. Berlin: Springer.

GOLDBERG, Adele E. 2006. *Constructions at Work: The Nature of Generalization in Language*. Oxford: Oxford UP.

—. 2014. "Fitting a Slim Dime between the Verb Template and Argument Structure Construction Approaches." *Theoretical Linguistics* 40 (1-2): 113-35.

—. 2019. *Explain Me This: Creativity, Competition, and the Partial Productivity of Constructions.* Princeton: Princeton UP.

—, Devin M. Casenhiser and Nitya Sethuraman. 2004. "Learning Argument Structure Generalizations." *Cognitive Linguistics* 15 (3): 289-316.

GOSCHLER, Julia and Anatol Stefanowitsch. 2023. "Generalization and Transfer in L2 Acquisition: The Role of Entrenchment in L1 and L2." Paper delivered at the 16th International Cognitive Linguistics Conference, Düsseldorf, August 2023.

GRANGER, Sylviane and Gaëtanelle Gilquin, eds. 2015. *The Cambridge Handbook of Learner Corpus Research*. Cambridge: Cambridge UP.

GRIES, Stefan T. and Anatol Stefanowitsch. 2003. "Collostructions: Investigating the Interaction of Words and Constructions." *International Journal of Corpus Linguistics* 8 (2): 209-43.

— and Stefanie Wulff. 2005. "Do Foreign Language Learners Also Have Constructions?" *Annual Review of Cognitive Linguistics* 3 (1): 182-200.

HERBST, Thomas. 2020. "Constructions, Generalizations, and the Unpredictability of Language: Moving towards Colloconstruction." *Constructions and Frames* 12 (1): 56-95.

HOFFMANN, Thomas and Graeme Trousdale, eds. 2013. *The Oxford Handbook of Construction Grammar*. Oxford: Oxford UP.

HOPPER, Paul. 1987. "Emergent Grammar." *Annual Meeting of the Berkeley Linguistics Society* 13: 139-57.

KYLE, Kristopher. 2016. "Measuring Syntactic Development in L2 Writing: Fine Grained Indices of Syntactic Complexity and Usage-Based Indices of Syntactic Sophistication." PhD diss., Georgia State University.

— and Scott Crossley. 2017. "Assessing Syntactic Sophistication in L2 Writing: A Usage-Based Approach." *Language Testing* 34 (4): 513-35.

— and Scott Crossley. 2018. "Measuring Syntactic Complexity in L2 Writing Using Fine-Grained Clausal and Phrasal Indices." *The Modern Language Journal* 102 (2): 333-49.

LARSEN-FREEMAN, Diane. 1997. "Chaos/Complexity Science and Second Language Acquisition." *Applied Linguistics* 18 (2): 141-65.

LEE, Jin-Hwa and Hye Min Kim. 2011. "The L2 Developmental Sequence of English Constructions and Underlying Factors." 영어학 11 (3): 577-600.

LI, Peiwen, Søren W Eskildsen and Teresa Cadierno. 2014. "Tracing an L2 Learner's Motion Constructions over Time: A Usage-Based Classroom Investigation." *The Modern Language Journal* 98 (2): 612-28.

LIEVEN, Elena V., Julian M. Pine and G. Baldwin. 1997. "Lexically-Based Learning and Early Grammatical Development." *Journal of Child Language* 24 (1): 187-219.

MacWhinney, Brian and William O'Grady, eds. 2015. *The Handbook of Language Emergence*. Malden: Wiley-Blackwell.

McCarthy, Michael. 2016. "Putting the CEFR to Good Use: Designing Grammars Based on Learner-Corpus Evidence." *Language Teaching* 49 (1): 99-115.

McCarthy, Philip M. and Scott Jarvis. 2010. "MTLD, Vocd-D, and HD-D: A Validation Study of Sophisticated Approaches to Lexical Diversity Assessment." *Behavior Research Methods* 42 (2): 381-92.

Meunier, Fanny. 2015. "Developmental Patterns in Learner Corpora." In Granger and Gilquin 2015, 379-400.

Nenonen, Marja and Sinikka Niemi, eds. 2007. *Collocations and Idioms. Papers from the First Nordic Conference on Syntactic Freezes*. PhilPapers.

Prior, Mathew T., ed. 2013. *Proceedings of the 31st Second Language Research Forum.* Somerville: Cascadilla Proceedings Project.

Roehr-Brackin, Karen. 2014. "Explicit Knowledge and Processes from a Usage-Based Perspective: The Developmental Trajectory of an Instructed L2 Learner." *Language Learning* 64 (4): 771-808.

Römer, Ute. 2019. "A Corpus Perspective on the Development of Verb Constructions in Second Language Learners." *International Journal of Corpus Linguistics* 24 (3): 268-90.

— et al. 2014. "Linking Learner Corpus and Experimental Data in Studying Second Language Learners' Knowledge of Verb-Argument Constructions." *ICAME Journal* 38 (1): 115-35.

— and Cynthia M. Berger. 2019. "Observing the Emergence of Constructional Knowledge: Verb Patterns in German and Spanish Learners of English at Different Proficiency Levels." *Studies in Second Language Acquisition* 41 (5): 1089-110.

—, Stephen C. Skalicky and Nick C. Ellis. 2018. "Verb-Argument Constructions in Advanced L2 English Learner Production: Insights from Corpora and Verbal Fluency Tasks." *Corpus Linguistics and Linguistic Theory* 16 (2): 303-31.

— and Selahattin Yilmaz. 2019. "Effects of L2 Usage and L1 Transfer on Turkish Learners' Production of English Verb-Argument Constructions." *Vigo International Journal of Applied Linguistics*16: 107-34.

Sparks, Richard L. 2022. *Exploring L1-L2 Relationships: The Impact of Individual Differences*. Bristol: Multilingual Matters.

Tabachnik, Barbara G. and Linda Fidell. 2014. *Using Multivariate Statistics.* Harlow: Pearson.

Tode, Tomoko and Hideki Sakai. 2016. "Exemplar-Based Instructed Second Language Development and Classroom Experience." *ITL-International Journal of Applied Linguistics* 167 (2): 210-34.

Tomasello, Michael. 2003. *Constructing a Language: A Usage-Based Theory of Language Acquisition*. Cambridge, MA: Harvard UP.

## Appendix

Table 6. Descriptive Statistics for Corpus-Based Indices

| Descriptive Statistics (corpus–based) | | | |
|---|---|---|---|
| | Mean | Std. Deviation | N |
| CEFR | 3.15 | 1.418 | 14120 |
| LF | 1517.63 | 1921.708 | 14120 |
| CF | 217.27 | 372.579 | 14120 |
| LFC_per_mil | 2447.806654771 | 11321.0685942004 | 14120 |
| collexeme_approx | 717.03977783535 | 1094.503550113793 | 14120 |

Table 7. Descriptive Statistics for Against-Reference-Corpus Indices

| Descriptive Statistics (against-reference-corpus) | | | |
|---|---|---|---|
| | Mean | Std. Deviation | N |
| all_av_lemma_freq | 2130726.451654 | 1089642.7990616 | 2347 |
| all_av_construction_freq | 578675.096749 | 117719.4343716 | 2347 |
| all_av_lemma_construction_freq | 236276.3095257 | 119155.86437477 | 2347 |
| all_av_approx_collexeme | 33032.2840899431 | 41865.47914957857 | 2347 |
| all_lemma_ttr | .23181847222 | .094126285048 | 2347 |
| all_construction_ttr | .45105434755 | .095545569046 | 2347 |
| all_lemma_construction_ttr | .62566841864 | .148603619021 | 2347 |
| all_lemma_attested | .99317114786 | .005132110376 | 2347 |
| all_construction_attested | .94516094431 | .022173839298 | 2347 |
| all_lemma_construction_attested | .86364407124 | .044220413735 | 2347 |

Received 17 March 2022 Revised version accepted 29 January 2023

Tan Arda Gedik is a PhD Candidate in linguistics specializing in the interplay between language, literacy acquisition and cognitive development. He works at FAU and is a visiting researcher at Bilkent University and conducts research at the intersection of linguistics, psychology and language learning. Gedik's contributions provide valuable insights into the cognitive benefits of literacy and the role of education in shaping language theories.