sciendo
ExELL (Explorations in English Language and Linguistics)
9.1 (2021): 26-55
DOI: 10.2478/exell-2022-0002
Original scientific article

# An analysis of lexicogrammatical development in English textbooks in Turkey: A usage-based construction grammar approach

**Tan Arda Gedik**
*Friedrich-Alexander-Universität Erlangen-Nürnberg, Germany*

## Abstract

This article measures the syntactic development indices in grade 5-12 English textbooks in Turkey. Through a usage-based construction grammar approach, it argues that the textbooks show an inconsistent development in verb-argument constructions (VAC) and other usage-based indices. The study employs an automatic software tool that detects variations in these indices and runs a statistical analysis on a corpus compiled by the author. Statistically significant results demonstrate that textbooks lack lexicogrammatical variation. As such, learners who use these textbooks are likely to experience a limited array of VACs that are limited in lemma-construction combinations. Findings also indicate that learners may not be exposed to the conventional usage patterns and frequencies of VACs when compared against a reference corpus. These findings may have an influence on the learners' generalization process, their low-proficiency level in English, and poor idiomatic uses of the language.

**Key words:** applied construction grammar; English textbooks; verb argument constructions; lexicogrammatical development; usage-based linguistics.

## 1. Introduction

In a usage-based approach to language learning, regardless of it being L1 or L2, the framework assumes that the more speakers are exposed to specific constructions, the more likely these constructions are to be entrenched in the mind (Wulff & Ellis, 2018). Another tenet of a usage-based approach is that it diverges from generative approaches of language acquisition and its production. As Wulff and Ellis (2018: 37) state, the processes involved in language

development are not specific to language learning, but rather manifest themselves in all kinds of knowledge acquisition. In other words, speakers do not necessarily acquire rules that generate sentences or acquire lexical items that can be inserted into slots according to these rules, but rather, they use general domain cognitive mechanisms such as pattern recognition, indirect negative evidence, and statistical preemption to learn a language, to name a few (Divjak, 2019; Goldberg, 2006, 2019). Combining this approach with construction grammar, as one may expect to see variation in ambient language, it becomes clear that certain constructions and verbs are more likely to be frequent in the input than others (Chater & Manning, 2006; Seidenberg & Plaut, 2014). If the input features a construction of low frequency, it becomes more difficult for the speaker to learn and reproduce that construction (Ellis, 2002, 2008; Ellis & Ferreira-Junior, 2009b; Wulff & Ellis, 2018). In other words, one could argue that the construction does not become as entrenched as other competing constructions in the mind of the learner due to low coverage. Therefore, it is safe to assume that there is a connection between familiarity, which can be thought of as overall proficiency, and coverage of low-frequency constructions in the input (Goldberg, 2019: 93–94).

In a recent study, Gedik and Kolsal (2022) examined English high school textbooks and English university entrance exams in Turkey to identify whether the two corpora aligned with each other syntactically and lexically. English language teaching in Turkey has been subject to a constantly changing curriculum and teaching materials. There have also been reports of a lack of proficiency in English across learners in Turkey. According to EPI (2021), Turkey ranked 70 out of 112 countries, indicating a low English proficiency. Gedik and Kolsal (2022) found a statistically significant mismatch between the two corpora ($p < 0.05$) for lexical sophistication, lexical diversity, and syntactic complexity levels. The interpretation of the results was that the exams consisted of much more complex language than what the textbooks taught. They argued that this mismatch further deepens the lack of proficiency in English due to what is called a negative backwash effect[1] in language learning. While the findings are certainly important, their study does not necessarily address L2 speakers' language development from a usage-based standpoint, especially regarding language teaching materials. As Ellis (2002a, 2002b) demonstrates, it is not the complexity of a particular linguistic item that leads to learning, but rather its frequency in the input. The previous studies, particularly the lexical sophistication and diversity

I would like to thank the editors, the two anonymous reviewers, and Prof. Stefanie Evert for their valuable feedback on an earlier version of this manuscript. All remaining errors are mine.

[1] Backwash effect is the influence a test has on test takers. A negative effect may take place when there is a mismatch between the goals of instruction and what the focus of the test. In this case, what is taught in Turkey does not reflect what is tested in the university entrance exams.

analyses, are partly based on frequency and occurrence of categories of words e.g., K1, K2, AWL, appearing in the corpora, and this seems to be in line with what some research demonstrates in language testing and evaluation (see for instance Cumming et al., 2005). Nevertheless, a usage-based approach includes and moves beyond word-level constructions i.e., the plural construction→ [N]s, to cover phrase-level constructions (Ellis, 2002a; Goldberg, 1995; Tomasello, 2003). Alongside phrases, research indicates that the frequency of verb argument constructions e.g., THE CAUSED-MOTION CONSTRUCTION, parallel syntactic development and the production of these constructions in learners (Ellis & Ferreira-Junior, 2009b; Lieven et al., 1997; Ninio, 1999). As such, the present paper improves upon Gedik and Kolsal (2022) by addressing the shortcomings of their study through a usage-based construction grammar approach to the lexicogrammatical development of verb argument constructions (VAC) in English textbooks used in Turkey for grades 5-12.

## 2. Merging construction grammar and usage-based linguistics

### 2.1. *Usage-based construction grammar*

Construction grammar, in its most basic sense, is a framework where lexis and grammar are united in form-meaning pairs e.g., words, phrases, idioms, sentences and so on (Goldberg, 1995). In its 1995 version (Goldberg, 1995: 4), the definition focused on unpredictability and idiomaticity of language, as seen below. That is, how the parts have an unpredictable meaning when combined. For instance, *it's raining cats and dogs* does not mean that there are cats and dogs free falling from the sky but rather it denotes the intensity of the rain. For this reason, idioms i.e., *a piece of cake*, and idiomatic expressions i.e., *by and large*, were a cornerstone within constructionist approaches because of their high frequency in ambient language and also because they were mostly sidelined by generative approaches.

> C is a construction iff*, C is a form-meaning pair <F,, S,> such that some aspect of F, or some aspect of S, is not strictly predictable from C's component parts or from other previously established constructions. (Goldberg, 1995: 4)

As seen in the definition, this unpredictability of semantics, in other words, such idiomatic expressions served as a ground for many succeeding constructionist studies and how constructions were defined. In her later work, Goldberg (2006: 5) included frequency into the picture and stated that constructions are constructions "even if they are fully predictable as long as they occur with sufficient frequency." With this, constructions no longer needed to be unpredictable by definition. The definition used in this work is the 2006 version.

Unlike projectionist approaches, i.e., generative approaches, construction grammar assumes that there is an interplay between the form and meaning creating pairs. In construction grammar, a verb argument construction has several participants and describes an event (e.g., the DITRANSITIVE CONSTRUCTION→transfer of X from person Y to Z→*I gave him a book*). While a generativist account of language might posit two entries for *sneeze,* a) as an intransitive verb, and b) as a transitive verb, construction grammar divides the labor of meaning between verbs and constructions. With the constructionist approach, grammar rules also become meaningful, contrary to what generativist approaches generally assume.

(1) *She sneezed the foam off the cappuccino*. (Goldberg, 2006: 42)

(1a) $NP_{she}$ $VERB_{sneezed}$ $OBJ_{the\ foam}$ $OBLIQUE_{off\ the\ cappuccino}$.

Example (1) illustrates the schematic use of the Caused-Motion construction combined with *sneeze.* A constructionist approach would argue that so long as the verb, its participant roles, and the construction at hand are semantically coherent (see the SEMANTIC COHERENCE PRINCIPLE, Goldberg, 1995: 50), the verb can occur in the construction. Therefore, it is safe to assume that the construction can also enrich the verb with different meanings, giving speakers a much more creative workspace in terms of combining schemas and never-heard-before combinations (Goldberg, 1995; Michaelis, 2004). The Caused-Motion construction in examples 1, and 1a display such a novel example. Goldberg (1995, 2006) names such constructions that span an entire sentence, which would be traditionally regarded as grammatical rules, ARGUMENT STRUCTURE CONSTRUCTIONS, which relate a highly schematic grammatical schema with an array of meanings. In other words, they "consist of a verb and all argument[s] [they] take" (Kyle, 2016: 29). However, in this study, argument structure constructions (see figure 1) are referred to as verb argument constructions (VACs) to keep the paper in line with Kyle and Crossley's (2017) usage of the term.

As such, the main difference between generative and constructionist approaches would be that, while in a generativist approach it is the verb and what subcategorization rules it allows for, i.e., transitive/intransitive, construction grammar eschews a verbocentric approach and calls for a division of labor between verbs and partially-filled or schematized form-meaning pairings, i.e., constructions, in meaning creation. Constructions vary with regard to schematicity and abstractness; some are completely lexically-filled and fixed such as idioms like *kick the bucket*, while some are fully schematized, and have slots to be filled by lexical items (for instance, the Ditransitive construction: subject verb object object), and some are gradient which have fixed items and slots such as the Xer the Yer construction, i.e., *the more the merrier*. Ideally, these constructions are on a spectrum of a lexicogram-

matical continuum (figure 1) and abstractness and item-specificity (figure 2). The gradience in figure 1 represents that, by nature, linguistic items are form-meaning pairings and arguably they cannot be split into grammar and lexis, as they are gradient. Figure 2 displays a partially filled expression, its specific items, and its taxonomic structure. Figure 3 shows VACs and examples.

Figure 1: The lexicogrammatical continuum.

| Prefixes | Words | Idioms | Partially filled expressions | Fully abstract/ schematic constructions |
|---|---|---|---|---|
| *-de, -re, -un…* | *Cat, dog, bird…* | *Kick the bucket* | The Xer the Yer<br>*the faster, the better*<br>The nice-of-you<br>*It was nice of you to come* | Argument Structure Constructions/VACs<br>The Ditransitive construction<br>*I gave him a book* |

Figure 2: Based on Herbst's analysis of some of the items that occur within the nice-of-you construction (Goldberg & Herbst, 2021: 20).

| **NP Verb AdjP Agent PP** (of) **INFINITIVE** (to Verb) |
|---|
| **NP** *BE NICE* **PP** (of) **INFINITIVE** (to Verb)<br>**NP** *BE GOOD* **PP** (of) **INFINITIVE** (to Verb) |
| *IT BE NICE* **PP** (of) *YOU TO BE HERE*<br>*IT BE GOOD* **PP** (of) *YOU TO COME* |

Figure 3: Several VACs from TAASSC, their conventional names, and examples.[2]

| VACs | Examples |
|---|---|
| Nsubj_verb_dobj (the transitive construction) | *I help my mom* |
| Nsubj_verb_iobj_dobj (the ditransitive construction) | *I gave my mom cookies* |
| Nsubj_verb_dobj_prepINTO (the caused-motion construction) | *I sliced chocolate into the batter* |

Perek (2015) also shows that the more frequently a verb is used in a specific construction, the more it absorbs the meaning which the construction embodies. That is, if a novel verb like *moop* is used in the Ditransitive construction very frequently, it will obtain the meaning of 'transfer' and its metaphorical extensions even if the verb does not have such a meaning prototypically, e.g., *I gabbledigooked her the news*, because the construction has such

[2] A full list of VACs extracted from the corpus with frequency data is available at https://drive.google.com/file/d/1phVPfTLXxESX4T5KYClLujivNtXt7LIs/view?usp=sharing

a meaning. In the same vein, some researchers argue for ITEM-SPECIFIC[3] (see Figure 2), especially verb-specific constructions (Herbst, 2018; Herbst, 2020). Herbst (2020) demonstrates that within the Ditransitive construction, *give* occurs almost 50% of the time and, as such, this Ditransitive construction might be stored as the *give*-ditransitive construction with usage information on which subjects and objects precede and succeed *give*. Item-specificity facilitates learning the meaning of a VAC because verbs that occur most frequently usually carry the most prototypical meanings of the VAC under scrutiny (see Bencini & Goldberg, 2000). Scholars have analyzed VACs both in L1 and L2 and revealed that VACs carry meaning for both groups of speakers (Bencini & Goldberg, 2000; Chang et al., 2003; Hare & Goldberg, 1999; Gries & Wulff, 2005). In other words, grammar is meaningful and contributes to the sentence meaning just as much as lexical items for L1 and L2 speakers. As Goldberg (2013) demonstrates with a nonsense verb in the Ditransitive construction, people associate the nonsense verb with literal or metaphorical transfer even if the construction lacks a verb that prototypically suggests the meaning of transfer. Studies also indicate that both L1 and L2 speakers learn VACs (among many other constructions) based on frequency (Goldberg et al., 2004; Lieven et al., 1997; Ninio, 1999; Ellis & Ferreira-Junior 2009a, 2009b). Alongside frequency, salience, attention (Wulff & Ellis, 2018), statistical preemption, and error-driven learning (Goldberg, 2019) are also other factors that influence construction learning, to name a few (see Divjak, 2019 for a lengthy discussion).

While basing proficiency on the frequency of any construction in a specific corpus may seem counter-intuitive, especially when frequency is not the only driving factor in construction learning, it still seems to be a predominant factor. One limitation in a usage-based construction approach is (i) the researchers can never fully account for all the variables in construction learning at once due to cost and time reasons, and (ii) only a limited section of a corpus can be analyzed, as it is very time-consuming. Nevertheless, as mentioned above, recent studies (Römer et al., 2014; Römer et al., 2015) demonstrate that the connection between reference corpora and the speakers' (L1 and L2) is stronger than previously thought. What these studies found was that the verbs used in VACs were generally in line and correlated to one another across L1, L2, and reference corpora. This, once again, indicates that frequency is a safe variable to test in learner materials, if one assumes a strong correlation between frequency and speaker proficiency. To overcome the above-mentioned shortcomings of a usage-based constructionist approach, researchers have developed automated ways of quantifying VACs (Kyle, 2016; Kyle & Crossley, 2017) using Contemporary Corpus of Ameri-

[3] Item-specificity is when "a property x is held by exactly one item $I_x$ and by no other item" (Herbst, 2020: 58).

can English (COCA, Davies, 2010) as a reference corpus. In their study, Kyle and Crossley (2017) employed automated VAC features in software called the Tool for the Automatic Analysis of Syntactic Sophistication and Complexity (TAASSC, Kyle, 2016) e.g., frequency data of verbs, constructions, collostructional analysis, and VAC frequencies to name a few. Furthermore, their study found a stronger correlation in L2 writing quality and higher essay scores in argumentative essays using usage-based indices compared to absolute complexity measures i.e., mean length of sentence, T-Unit, complex nominals per phrase, and so on. The results ultimately indicate that the higher a score received the essay, the more likely it included low-frequent VACs (Kyle & Crossley, 2017). This further strengthens the connection between learner proficiency and VAC learning based on frequency information. In light of previous research studies, as Ellis and Wulff (2015: 86) suggest, "a large and representative sample of language is required for the learner to abstract a rational model that is a good fit to the language data". While frequency plays an important role in the language learning process from a usage-based perspective, it is also important to mention the need to research the effects of other variables in the process of language learning (Divjak, 2019). Namely, attention spent on an item, auditory/visual perception, salience of an item, and memory. Divjak (2019: 130) comments on how corpus linguists have somewhat misinterpreted the connection between frequency and "what is linguistically experienced and encoded" in one's memory. Nevertheless, Divjak (2019), acknowledges the positive effects that frequency can have on memory and ultimately language development if one assumes that language learning is deeply connected to memory and experience. These effects are "the frequency effect, the spacing effect, and the serial position effect" (Divjak, 2019: 119). While it is important to attune to the outcomes of different variables on memory and language development, carrying out a study that accommodates all variables would be longitudinal and require more resources such as participants, questionnaires, and laboratories. Therefore, if frequency "is the most straightforward interpretation of the effect that repetition might have on the encoding and storage of a given event in memory" (Divjak, 2019: 131), then, it is safe to assume that the more a learner is exposed to a particular VAC or any other construction, the more likely these constructions will be entrenched and will be easier to produce.

Frequency, or coverage, helps learners distinguish constructions'[4] conventionalized forms from unconventionalized forms and produce them in line with the conventionalized usage patterns. As Herbst (2020: 84) exemplifies it by stating that "layers of usage events… become linked on the basis of

[4] Here and throughout the paper, I use the term 'construction' to refer to all shapes of linguistic pairings of form-meaning (e.g., words, prefixes, argument structure constructions to name a few).

recognized similarities between them". This means that usage events help learners identify conventionalized forms of a construction. For instance, when a learner experiences the Ditransitive construction in the following sentences (Herbst, 2020: 84), they recognize the similarities and links them to one schematic use of the construction.

(2) *Postman pat brought her some green Wellingtons.*

(3) *I'm going to my teddy a piece of apple.*

(4) *You could offer him a drink.*

(5) *You could offer him a drink.*

(6) *Bake me a cake.*

(7) *Shall I give you a kiss?*

(8) *We can give the teddy something to eat.*

In the same vein, it is important to mention the notion of entrenchment as it is an important term for the present study. In its most condensed form, entrenchment is the learning of more constraints on constructions in correlation with more exposure to them (Divjak, 2019: 48–49). Divjak (2019: 51) illustrates entrenchment as "repeated presentations of a verb in particular constructions, e.g., *The rabbit disappeared*, 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 is called simple entrenchment where frequency is "simply a proxy for familiarity." Another, perhaps more important type of entrenchment that needs attention is what Goldberg (2019: 77) calls conservatism via entrenchment. 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 have occurred together based on the frequency information of the item and the construction and based on this information arrive at a conclusion of generalizability of an item. This, however, does not mean one retains all the item-specific information for a construction, since memory is lossy (Goldberg, 2019), but whenever one experiences 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). Thus, ideally, entrenchment[5] can be sketched as the outcome of frequency and usage events (Divjak, 2019: 55–56), or in other words, linguistic experience. For Goldberg (2019: 94), entrenchment also explains how "more entrenched and better-covered constructions are easier to access,

[5] In this article, I refer to familiarity (frequency) and blocking effects (conservatism) of entrenchment as entrenchment and do not make a distinction.

which results in more conventional language being used more often, which further strengthens the association between conventional forms and particular messages-in-context." Therefore, the role of frequency, albeit equal to other factors, is salient as it can be detected using statistical measures.

Lastly, employing a usage-based construction approach overcomes the limitations of a traditional approach to syntactic complexity and development. For instance, Biber et al. (2011) mention that while phrasal complexity is a feature of academic writing, clausal complexity is not strongly correlated with it. Larsen-Freeman (2009) provides information on how large-grained indices e.g., mean length of sentence, clauses per T-Unit, and so on, may overlook the variation in the speakers' output and may lead to misleading results if one decides to use it for testing or research purposes. Finally, the traditional approach assumes a divide between lexis and grammar, which creates gaps in the theoretical understanding of how one should approach (syntactic) language development in light of new research studies.

## 2.2. *TAASSC and its indices*

TAASSC is a tool developed by Kyle (2016). The tool demonstrates a variety of different indices to measure syntactic sophistication levels.[6] Of particular interest for this study are its VAC indices. The tool calculates, by using a reference corpus, the "frequency of main verb lemmas and VACs and the strength of association between VACs and the verbs that fill them" (Kyle & Crossley, 2017: 517). Using COCA (Davies 2010) as the reference corpus, the tool introduces indices with frequency profiles based on all available registers on COCA, i.e., academic, fiction, magazine, and newspapers. By means of utilizing frequency profiles, the tool establishes frequency lists for main verb lemmas, VACs, verb–VAC combinations, and it uses indices to uncover the strength of association between VACs and main verb lemmas. For frequency indices, the tool computes an index score for the indices and "if a particular target structure, e.g., a VAC, that occurs in a text does not occur in the reference corpus, it is not counted toward the index score" (Kyle & Crossley, 2017: 522).

TAASSC calculates the association strength to identify the probability of two items occurring next to one another. These two items are a main verb lemma and a VAC, i.e., *write him a letter* → verb_indirectobject_directobject. As Ellis and Ferreira-Junior (2009a, 2009b) explain, strength of association and language development are interdependent, as the frequency of verbs and constructions are reflected in one's linguistic development. Tomasello's (2000) study also suggests that if a construction is already experienced with

[6] TAASSC 1.1 Index Description Spreadsheet (available on https://www.linguistic analysistools.org/tools.html).

a wide array of verbs, the speakers are more likely to extend the use of the construction. In line with these observations, TAASSC employs three indices related to association strength. Namely, these are faith (Gries et al. 2005), delta P (Ellis & Ferreira-Junior, 2009b), and collostructional strength (Stefanowitsch & Gries, 2003).

Gries et al. (2005) explain and calculate faith scores as the likelihood of a particular verb appearing in a particular VAC. As Kyle and Crossley (2017: 524) demonstrate, FAITH scores are calculated using $\left(\frac{a}{a+b}\right)$. To put it in perspective, they exemplify it using corpus data where they find the likelihood of the verb *have* appearing in the TRANSITIVE CONSTRUCTION to be 17.7% ($p$= *.177*) in COCA (Kyle & Crossley, 2017: 525).

In addition to FAITH, delta P predicts the likelihood of a VAC when triggered by a cue deduced by the likelihood of the VAC appearing without the triggering cue. Kyle and Crossley (2017: 525) calculate this with the following formula: $\left(\frac{a}{a+b}\right)-\left(\frac{c}{c+d}\right)$. They identify the likelihood of *have* appearing in the Transitive construction to be higher than the likelihood of the transitive construction appearing with another verb. Namely, .177-.053=.124 in COCA. This means, if the Transitive construction had a list of top ten frequent verbs, *have* would rank number one.

The final association strength index is collostructional analysis (Gries et al., 2005; Stefanowitsch & Gries, 2003). This index predicts the likelihood of two items from the corpus appearing right next to one another. While this index is originally calculated using the Fisher-Yates test (Fisher, 1934; Yates, 1934), Kyle and Crossley (2017: 525) employ the following formula, as it is easier to compute and 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 for "approximate collexeme strength" (Kyle & Crossley, 2017: 525).

Indices that analyze type-token ratio (TTR) are also available in TAASSC. Broadly explained, TTR calculates the total amount of different types divided by tokens in a corpus. For instance, Stefanowitsch and Flach (2016) employ TTR to identify the shortcomings of using just a frequency-based approach to linguistic development (or entrenchment). Their example (Stefanowitsch & Flach, 2016: 118–119) draws on two schemas (namely, *drive* [NP ADJ] and *color* [NP ADJ]). Frequency measures show that the former schema is far more entrenched (i.e., has a higher frequency count) than the latter schema (1028 > 46). However, when the TTR levels for the schemas are analyzed, it becomes clear that although specific occurrences of *drive* [NP ADJ] (e.g., *drive me crazy*) are more entrenched, the schema of *color* [NP ADJ] itself is more entrenched as "The instances of [drive NP ADJ] are all filled by adjectives meaning "insane" and/or "angry" (i.e., synonyms of *crazy*/*mad*), whereas the instances of [*color* NP ADJ] are filled by a semantically hetero-

geneous set of adjectives" (Stefanowitsch & Flach, 2016: 119). This suggests that some constructions, even though they have slots, e.g., *drive* [NP ADJ], have some instantiations that occur with specific items, which are more entrenched than their highly schematic forms, *drive me crazy* being hypothetically more entrenched than *drive* [NP ADJ]. In light of this, it is crucial to include indices that measure TTR levels for constructions as well.

Although there are thirty-five VAC indices available in TAASSC under syntactic sophistication, as explained later in the paper, only fourteen of the indices were eligible for the actual analysis. Due to space limitations, only those fourteen indices can be explained in detail[7]. Table 1 displays the indices used in this study. An important note regarding the examples given for association strength-based is that while TAASSC provides specific index data for each verb-VAC combination in the corpus, the indices utilized here are global, that is they provide an overall picture into constructional development. As such, the data for each verb-VAC combination is then added and divided by the total number of combinations to calculate the average scores for global indices. Nevertheless, the examples should help with a clearer presentation of what these global indices stand for.

Indices are split into frequency-based and association strength-based subgroups. Starting with association strength-based indices, ALL_AV_APPROX _COLLEXEME is the approximate joint probability of a verb-construction combination to appear in the corpus. In other words, it calculates the likelihood of two items occurring next to each other. For GET_V-PREP_INTO, (*They do this to* ***get into the skins of characters***, Teenwise 9th grade), there is a 2.5% chance that *get* and this construction will appear together in the entire corpus. Moving onto ALL_FAITH_VERB_CUE, it is the probability that a particular VAC will appear when a verb is given as cue. Put differently, it is a directional approach where it calculates how faithful a verb is to a construction e.g., what is the probability of *have* appearing in the *SVO* VAC?. For the verb-VAC combination LIVE_NSUBJ-V-PREP_WITH (*she married and went to* ***live with her husband***, CountMeIn 12th grade), the faithfulness of *live* to VAC is 0.001% in the corpus. Similarly, ALL_FAITH_CONST_CUE measures the probability that a verb will appear when a construction is given as cue and as such the probability of NSUBJ-V-PREP_WITH triggering the use of *live* in that VAC is 0.005%. Arguably, this construction is more faithful to the verb (*live*) than the verb is to the construction. ALL_DELTA_P_VERB_CUE calculates the probability of a VAC as an outcome when triggered by a cue (verb) deduced by the likelihood of the VAC appearing without the triggering cue (verb). For the verb-VAC combination DEAL_V-PREP_WITH (*I've had to* ***deal with a lot of repairs these days***, CountMeIn 12th grade), the probability of *deal* occurring in such a

[7] See TAASSC 1.1 Index Description Spreadsheet for further detail on the rest of the VAC indices (available on kristopherkyle.com).

VAC is 0.0029% when compared to the likelihood of the VAC appearing without *deal*. ALL_DELTA_P_CONST_CUE does the same task in the reverse order, namely, it calculates the same probability for a verb as an outcome when triggered by a cue (VAC) which is then deduced by the probability of the verb appearing without the triggering cue (VAC). Using the previous example, it is 0.00036%. This means *deal* is more likely to occur in NSUBJ-V-PREP_WITH than other verbs appearing in the VAC, whereas the VAC seems less likely to occur without the verb. As such, one could argue that *deal* might be an item-specific item to this construction.

Turning our attention to frequency-based indices, ALL_CONSTRUCTION _TTR measures construction type-token ratio, i.e., the number of different types of constructions divided by the total number of constructions. For instance, while the 6th grade MEB textbook has a TTR of 25.79%, the 7th grade counterpart has a total of 17.29%. The index ALL_LEMMA_CONSTRUCTION_TTR calculates the type-token ratio of lemmas in constructions where the unique instances of lemmas are divided by the total number of lemma-construction combinations and as such the 7th grade textbook has a score of 41.14% and the 8th grade textbook has a score of 46.52%. ALL_LEMMA_ATTESTED measures how many of the verbs that are in the corpus are also available in the reference corpus. To illustrate, the Teenwise 9th grade textbook covers 99.75% and the Gizem 10th grade textbook covers 99.95% of all lemmas in the reference corpus. ALL_CONSTRUCTION_ATTESTED calculates the approximate overlap of constructions in the corpus against the reference corpus (Teenwise: 97.19%; Gizem: 96.33%). In the same vein, ALL_LEMMA_CONSTRUCTION_ ATTESTED measures the approximate overlap of lemma-construction combinations in the corpus against the reference corpus (Teenwise: 91.05%; Gizem: 89.45%). ALL_AV_LEMMA_FREQ_LOG, ALL_AV_CONSTRUCTION_FREQ_LOG, and ALL_AV_LEMMA_CONSTRUCTION_FREQ_LOG calculate the lemma, construction, and lemma/construction frequency data respectively, but log transformed.

While the tool has been employed in several research studies, these studies mainly fall in L2 and syntactic development and L2 writing development areas (Kyle and Crossley 2018; Nakamura 2019; Diez-Bedmar and Perez-Paredes 2020). To this day, the tool has been used only once to identify linguistic complexity levels of various school textbooks (Green, 2019). Therefore, considering previous literature, the present study employs TAASSC for its validity and reliability (see Kyle (2016)) to answer the research questions in this article.

Table 1: Fourteen indices out of thirty-five used in the study.

| Frequency-based VAC indices | Association Strength-based VAC indices |
|---|---|
| all_construction_ttr | all_av_approx_collexeme |
| all_lemma_construction_ttr | all_av_faith_verb_cue |
| all_lemma_attested | all_av_faith_const_cue |
| all_construction_attested | all_av_delta_p_verb_cue |
| all_lemma_construction_attested | all_av_delta_p_const_cue |
| all_av_lemma_freq_log | |
| all_av_construction_freq_log | |
| all_av_lemma_construction_freq_log | |

## 2.3. *English language teaching materials from a usage-based perspective*

While various studies analyze English language teaching (ELT) materials in Turkey (Arıkan, 2005; Demir & Yavuz, 2017), these studies mainly scrutinize the sociocultural aspects of the materials. To this day and to the researcher's knowledge, only one study has examined ELT materials from a corpus linguistic point-of-view (Gedik & Kolsal, 2022).

In Turkey, students begin taking English classes starting at grade 2 until they graduate from high school. Milli Egitim Bakanligi (MEB), the 'Ministry of National Education', states that English textbooks used in grades 5 and 6 are at the A1 level, while 7 and 8 are A2 in accordance with the CEFR (MEB 2018: 10). As for the high school curriculum, MEB states that the textbooks begin at A1 and aims to achieve at least a minimum of B2 at the time of graduation (MEB, 2018: 7). MEB also states that grades below 5 have very limited reading and writing activities in the secondary school curriculum and consequently English textbooks (MEB, 2018: 10). Therefore, since the study draws on textual input, grades 5 and above were found eligible to be included in the corpus.

As stated before, EPI (2021) for Turkey is low, with the country ranking 70 out of 112 countries. There have been several studies investigating the possible reasons for consistent low English proficiency in Turkey (Coşkun, 2016; Yurtsever Bodur & Arıcak, 2017; Erarslan, 2019, to name a few). Some of the suggested reasons are parents' lack of knowledge of English, some economic and sociocultural causes (Yurtsever Bodur & Arıkan, 2019), e.g.,

not having access to TV channels that broadcast in English, teaching practices that heavily rely on grammar teaching and not communication, linguistic differences between Turkish and English (Coşkun, 2016), and a lack of motivation resulting from the implementation of, and a lack of proper equipment in, and the duration of English classes, and the number of students per classroom and the number of teachers (Erarslan, 2019).

While MEB states that the materials help students to gradually develop their language skills (MEB, 2018: 7), the mismatch previously uncovered (Gedik & Kolsal, 2022) raises concerns about this claim, as the high school textbooks severely lacked lexical and syntactic diversity. As many studies suggest, corpus-driven materials that mirror lexical, semantic, syntactic, and other linguistic information as they take place in real time, are needed (Carter & McCarthy, 2006; Dolgova & Tyler, 2019; Lee & Swales, 2006). Although the curriculum does not state that the materials are corpus driven or usage-based, its claim to help students gradually develop their language skills already poses the question: do the textbooks show a gradual progression in terms of lexical/syntactic items? Or in the context of the present study, do they gradually increase in usage-based syntactic sophistication levels e.g., VACs?

As of now, the implications of many studies (Monsell, 1991; Divjak & Caldwell-Harris, 2015; Baayen et al., 2016) help researchers argue that if a learner is exposed to a linguistic item frequently, disregarding other cognitive processes involved in the language learning journey such as attention, memory, and other biological factors (see Divjak (2019) for an extensive discussion of these), the learner is more likely to learn and retrieve that item. Thus, analyzing these materials from a usage-based perspective can provide textbook creators insight into their materials.

### 2.4. *Research Questions*

Considering previous literature, the present study expands on the previous findings and aims to answer the following research questions:

1. Is there a statistically significant difference from grades 5-12 in regard to VAC indices?
2. Is there a statistically significant difference in regard to the development of VAC indices in the corpus compiled for this study?

## 3. Methodology

In order to answer the research questions, the data were gathered from MEB's learning management system for state schools, Eğitim Bilişim Ağı

(EBA, Education and Information Network, eba.gov.tr). EBA offers online materials, quizzes, and other activities ranging from the 1st to 12th grade for all school subjects, one of which is English. The textbooks are also accessible by students and teachers for free in .pdf format. Using Gedik and Kolsal's (2022) corpus,[8] this study expands the textbook corpus to cover grades between 5 and 8 in addition to 9 and 12, which were already available in the previous corpus. According to the new Turkish education system, which was introduced in 2012, a 4+4+4 system where the first four years belong to elementary, the second part belongs to the secondary, and the final part belongs to the high schools was introduced. As it was not stated in the curriculum that the materials were corpus-driven, and as the nature of this study calls for a corpus-based i.e., text-based approach, the first four years were not included as they lacked reading and writing activities (MEB 2018: 5–10). The textbook publishers, which vary from grade to grade, were identified using the national curriculum (MEB 2018a, 2018b) and the appropriate textbooks were downloaded from the learning management system. However, because the secondary school grades (namely, 5th-8th grades) do not have workbooks that complement the student's books, the workbooks for 9th-12th grades were excluded from the finalized corpus. The textbooks for grades 5-12 have corresponding listening exercises on EBA which were also included in the corpus. Thus, the textbooks and the exercises were downloaded in .pdf format. The textbooks were published by the following publishers (publishing company: name of the textbook): MEB: English 5: 5th grade, Özgün: English 5: 5th grade; MEB: English 6: 6th grade; Monopol: English 6: 6th grade; MEB: English 7: 7th grade; Kök: English 7: 7th grade; MEB: English 8: 8th grade; Tutku: English 8th : 8th grade; MEB: Relearn, Teenwise, Progress: 9th grade; Gizem, Count Me In: 10th grade; Sunshine, Silverlining: 11th grade; MEB: Count Me In: 12th grade. As seen above, some grades have several textbooks available from multiple publishing companies. Grades 5-8 materials had a total token number of 220,997 while Grades 9-12 had a token number of 300,129. The entire corpus had 521,126 tokens. Although there are several different publishers, MEB ensures that all textbooks present the same content by evaluating the content in the textbooks (MEB, 2018).

While using TAASSC, the syntactic sophistication option was selected. Although it has various registers under that category, the ALL COCA WRITTEN register was chosen to reflect real-life usage of the language as the textbooks, or the curriculum do not state an academically oriented English proficiency outcome. Therefore, because an overall proficiency would ideally include all registers, ALL COCA WRITTEN was employed. All indices under ALL COCA WRITTEN were included in the analysis (see Kyle,

[8] The corpus is available at https://drive.google.com/drive/folders/1sG5hbekwavfoVrkhntow3dalwmCGT-lE?usp=sharing

2016: 68) for differences between frequency, type-token ratio, attested items, association strength, and variation). The following steps were carried to compile the corpus and analyze it: (a) check the national curriculum to identify the book publishers, (b) download the appropriate textbooks and listening transcripts in .pdf format from eba.gov.tr, (c) convert the .pdf files into .docx files using Adobe Pro, (d) check the files for typos that may have occurred during the conversion and clean if any, (e) export the .docx as .txt files using the UTF-8 setting, (f) run TAASSC on the files, (g) import the .csv output file into SPSS, (f) do a preliminary analysis of the indices using boxplots, (h) run MANOVA, (i) interpret the results.

## 4. Results

Before moving onto MANOVA, using boxplots, the trends in the corpus were identified. With this, it is possible to pinpoint which indices show a gradual development in the entirety of the corpus and as such the boxplots give the researcher insight into whether MANOVA is necessary or not. As seen in the boxplots,[9] except for all_lemma_attested, all other indices demonstrate an idiosyncratic developmental pattern. Thus, it is important to investigate the issue using MANOVA.

Several preliminary statistical analyses were conducted to ensure that the data were eligible to be run through MANOVA. At the beginning, the results of all thirty-five VAC indices were included in the preliminary analyses. Out of these indices, eighteen are related to frequency and the other seventeen are related to association strength-based measures. Out of thirty-five, ten VAC indices were removed from the dataset as they violated the assumption of normality. Eleven other VAC indices were removed since they had multicollinearity (Tabachnick & Fidell, 2014). In the end, there were fourteen VAC indices to report: nine frequency and five association strength-based indices as seen in Table 1. To account for the Zipfian nature of linguistic data (Zipf, 1935), logarithm transformations of the indices were preferred when available.

Multivariate tests indicate a statistically significant difference across the textbooks ($p = .034$) but these tests do not provide information as to which of the indices are actually statistically significant. Tests of between subjects demonstrate which exact indices are statistically significant in the corpus. The tests show that, out of the fourteen indices here, six indices (all_av_

[9] The boxplots were excluded from the article due to space related issues. They are available online at https://drive.google.com/file/d/1Tlp-p-TEXQpaLbH4tvXoWhNQY4pAKrIp/view?usp=sharing

approx_collexeme; all_av_faith_verb_cue; all_av_construction_freq_log; all_lemma_construction_ttr; all_construction_attested; and all_lemma_construction_attested) were statistically significant in the corpus (*p= .020; p= .046; p= .034; p= .002; p= .002; p= .003*). The table below demonstrates the descriptive statistics for the six indices.

Table 2: Descriptive Statistics.

| | corpus | mean | std. deviation | N |
|---|---|---|---|---|
| all_av_approx_ collexeme | 5thg | 22626.0942995000 | 6819.38233513515 | 4 |
| | 6thg | 32264.3018680250 | 25857.07570882010 | 4 |
| | 7thg | 14570.6977800425 | 16796.80057603693 | 4 |
| | 8thg | 14582.5365685250 | 13244.36775003682 | 4 |
| | 9thg | 46498.7883868400 | 25477.17193511307 | 5 |
| | 10thg | 15054.4046140025 | 7009.17840897714 | 4 |
| | 11thg | 22852.6977797250 | 7585.30857239037 | 4 |
| | 12thg | 17755.4525454500 | 6405.04416553623 | 2 |
| | Total | 24380.8960859477 | 18615.63478544160 | 31 |
| all_av_faith_verb_cue | 5thg | .054258697184975 | .003749124452027 | 4 |
| | 6thg | .045892419447875 | .007551042402815 | 4 |
| | 7thg | .050286428217250 | .003486311820491 | 4 |
| | 8thg | .048782578775850 | .005590998505363 | 4 |
| | 9thg | .045249052709040 | .003314724587964 | 5 |
| | 10thg | .045856776094950 | .006057381389849 | 4 |
| | 11thg | .042762171755625 | .002995281754255 | 4 |
| | 12thg | .043370474939750 | .001646968655421 | 2 |
| | Total | .047236854817123 | .005524769143966 | 31 |
| all_av_construction_ freq_log | 5thg | 5.0300504567625 | .05761488952018 | 4 |
| | 6thg | 4.9027903063275 | .09036908099108 | 4 |
| | 7thg | 5.0830581403525 | .07581551642962 | 4 |
| | 8thg | 4.9866251680850 | .13710246229790 | 4 |
| | 9thg | 4.9458508534080 | .08976926804642 | 5 |
| | 10thg | 4.9065102666500 | .05228934445291 | 4 |
| | 11thg | 4.8057141112525 | .06012805346630 | 4 |
| | 12thg | 4.8147938309750 | .04612656790007 | 2 |
| | Total | 4.9425108298939 | .11517848136891 | 31 |
| all_collexeme_ratio | 5thg | 4.8482872580000 | 1.58942190210636 | 4 |
| | 6thg | 4.0642918998875 | .88155263881843 | 4 |
| | 7thg | 4.7325241753050 | 1.43273718606458 | 4 |
| | 8thg | 3.9865642636125 | .19151521627241 | 4 |
| | 9thg | 3.7584198072040 | .26633479692968 | 5 |
| | 10thg | 4.1495971662100 | .74956212932176 | 4 |
| | 11thg | 4.4860379119875 | .53921685779674 | 4 |
| | 12thg | 4.3161017916200 | .09860601239462 | 2 |
| | Total | 4.2739843006216 | .88073227391212 | 31 |

| | | | | |
|---|---|---|---|---|
| all_lemma_ construction_ttr | 5thg | .44476035104900 | .047629095253088 | 4 |
| | 6thg | .52531460792725 | .088698526931072 | 4 |
| | 7thg | .45473064291925 | .038081171054581 | 4 |
| | 8thg | .55486746562600 | .073793950109292 | 4 |
| | 9thg | .56380012187700 | .052385692435703 | 5 |
| | 10thg | .53056152915500 | .049260273427495 | 4 |
| | 11thg | .62318603830825 | .057918852618848 | 4 |
| | 12thg | .64418327361950 | .032333838783351 | 2 |
| | Total | .53680805472784 | .081615847395950 | 31 |
| all_construction_ attested | 5thg | .97079525074000 | .012353285544158 | 4 |
| | 6thg | .96524057186050 | .005365059280805 | 4 |
| | 7thg | .97987191468975 | .008197261861174 | 4 |
| | 8thg | .96679747012925 | .006198104954633 | 4 |
| | 9thg | .96520382350480 | .015287849126296 | 5 |
| | 10thg | .96326338632275 | .008260660115429 | 4 |
| | 11thg | .94274276838475 | .008291395497167 | 4 |
| | 12thg | .95067822111950 | .008019984223576 | 2 |
| | Total | .96394261317003 | .013751016202038 | 31 |
| all_lemma_ construction_attested | 5thg | .91846697250300 | .020586231245591 | 4 |
| | 6thg | .90423063354050 | .010965996902301 | 4 |
| | 7thg | .93165029893775 | .016141684801632 | 4 |
| | 8thg | .87718049594200 | .023135746985442 | 4 |
| | 9thg | .89919489107000 | .031808973772774 | 5 |
| | 10thg | .88437750175650 | .031733262563515 | 4 |
| | 11thg | .85938238483900 | .018975256075964 | 4 |
| | 12thg | .86266094112550 | .010435116181630 | 2 |
| | Total | .89427256411858 | .031172986774765 | 31 |

The contrast results option (K Matrix, see Table 3) in MANOVA was utilized to uncover the fine-grained differences in indices across the grades. This setting compares grade 5 to 6, 6 to 7 and so on and helps to explore whether there is a statistically significant, in other words non-gradual, development of indices. As Table 3 demonstrates, on which statistically significant results were boldened, grades 5 through 12 show statistical significance for all_av_approx_collexeme; all_av_faith_verb_cue; all_av_construction_ freq_log; all_lemma_construction_ttr; all_construction_attested_ all_lemma_ construction_attested indices ($p < .05$).

When compared against Table 2, it is clear what changes occurred in the developmental history of the textbooks. Upwards and downwards pointing arrows in Table 3 indicate a statistically significant increase or decrease and G*n* shows in which grade the change happened. The following results ascends from the lowest to the highest grade.

- Grade 5 to 6, all_av_faith_verb_cue and all_av_construction_freq_log were found to be significant ($p < .05$), where grade 6 had a decrease in

both indices.

- Grade 6 to 7 had a statistically significant gap in all_av_construction _freq_log with grade 6 having a decrease in the index.
- Grade 7-8, all_lemma_construction_attested and all_lemma_construction_attested were found to statistically decrease in grade 8.
- Grade 9, on the other hand, demonstrated a statistically significant increase in all_av_approx_collexeme. Between grade 8-9, all_lemma_ construction_ttr statistically increased.
- The jump between grade 9 to 10 did not display any statistically significant results.
- Grade 10 to 11 had a statistically significant p value for all indices except for all_av_approx_collexeme. All indices, except all_lemma_construction_ttr, showed a decrease.
- These indices had a significant decrease from grade 10 to 11 except for all_lemma_construction_ttr.
- And finally, grade 11 to 12 showed a statistically significant increase in indices all_av_construction_freq_log and all_lemma_construction _ttr.

Table 3: K matrix results.

| **P value** | all_av_approx_collexeme | all_av_faith_verb_cue | all_av_constructi-on_freq_log | all_lemma_constructi-on_ttr | all_construction_attested | all_lemma_constructi-on_attested |
|---|---|---|---|---|---|---|
| Grade 5-6 | .425 | **.021 ↓ (G6)** | **.043 ↓ (G6)** | .067 | .437 | .393 |
| Grade 6-7 | .223 | .943 | **.033 ↓ (G6)** | .411 | .064 | .165 |
| Grade 7-8 | .385 | .627 | .703 | **.028 ↓ (G8)** | .377 | **.005 ↓ (G8)** |
| Grade 8-9 | **.007 ↑ (G9)** | .076 | .215 | **.033 ↑ (G9)** | .294 | .477 |
| Grade 9-10 | .240 | .257 | .082 | .505 | .256 | .098 |
| Grade 10-11 | .877 | **.040 ↓ (G11)** | **.001 ↓ (G11)** | **.002 ↑ (G11)** | **.000 ↓ (G11)** | **.002 ↓ (G11)** |
| Grade 11-12 | .612 | .241 | **.036 ↑ (G12)** | **.013 ↑ (G12)** | .064 | .058 |

Using Cohen's D (1988), it is possible to explore the effect size of these significant indices. The effect size reports the difference between the grades and gives an overall view of how much of a change there is between grade X and grade Y. Cohen (1988: 67) states that the significance of the effect sizes are as follows: small effect (0.2), middle effect (0.5), and large effect (0.8).

- The effect sizes of the statistically significant indices in the grade 5 to 6 cohort are found to be 1.40 and 1.67 for all_av_faith_verb_cue and all_av_construction_freq_log respectively which both have large effect sizes.
- For grade 6 to 7, the effect size was found to be large (2.16) for all_av_construction_freq_log.
- As for grades 7 to 8, all_lemma_construction_ttr showed a large effect size (1.70), and all_lemma_construction_attested displayed a large effect size (2.73).
- Grade 8 to 9 had an effect size of 1.57 (large effect) for all_av_approx_collexeme and of 0.13 (small effect) for all_lemma_construction_ttr.
- Grade 10-11 had the following effect sizes for the indices all_av_faith_verb_cue (0.64; middle effect); all_av_construction_freq_log (1.78; large effect); all_lemma_construction_ttr (1.72; large effect); all_construction_attested (2.47; large effect); and all_lemma_construction_attested (0.95; large effect).
- The effect sizes in the final cohort 11 to 12 were calculated to be 0.16 (small effect) for all_av_construction_freq_log and 0.44 (small effect) for all_lemma_construction_ttr.

## 5. Analysis

Out of thirty-five VAC indices, eleven of those were removed in accordance with MANOVA assumptions. Six out of the remaining fourteen indices were found statistically significant in comparison to between-grades in the corpus according to the tests between subjects and K Matrix results. Turning to interpreting these results, several findings become clear when the results are coupled with descriptive statistics (Table 2). Starting with all_av_approx_collexeme, it is the approximate joint probability of a verb-VAC combination to appear in the corpus. In other words, it calculates the probability of two items (i.e., what is the probability of give and indirect object occurring next to each other in the corpus?) occurring next to each other. Descriptive statistics (see Table 2) suggest that the cohort 8-9 textbooks have a higher approximate joint probability of verb-VAC combinations to appear. The effect size also confirms that this is a considerably large effect (1.57). Ideally, this is a

desirable outcome as from a purely frequency-based perspective students need the input with frequent repetition to strengthen the 'lossy memory' (Goldberg, 2019) paths for entrenchment to occur.

Moving onto all_faith_verb_cue, it is the probability that a particular VAC will appear when a verb is given as cue. Put differently, it is a directional approach where it calculates how faithful a verb is to a construction, e.g., what is the probability of *have* appearing in a Transitive construction? The decrease from grade 5 to 6 suggests that verb-construction combination probabilities are not as systematic between the two grades. A decrease in the verb-construction combination probabilities may also suggest that grade 6 students are not given as much of an opportunity to experience a wide array of verb-construction combinations. The effect size (1.40) is also found to be large for this cohort, thus it can be argued to possibly affect learners. The index also demonstrates another statistically significant decrease in the corpus' verb-construction combination probabilities at grade 10-11 and remain stagnant for the rest of the grades. The decrease has a middle size effect (0.64). Since the index is based on a reference corpus, this decrease might suggest that the learners may not always be exposed to the real-life usage patterns of a particular verb-construction combination. To take up Herbst's (2020) idea of a collo-profile (see Herbst, 2020: 81) as an example for a collo-profile), if a construction (e.g., the Ditransitive construction) in the corpus does not reflect the same usage pattern of a verb (e.g., *give*) in terms of raw frequency, then, as Herbst (2020: 56) argues, the non-representative usage pattern of the verb-construction combination "automatically distort[s] the collocational profiles of verbs in argument structure constructions."[10] Moreover, the more verbs that are used and to which students are exposed, the more likely it becomes for students to generalize the meaning of verbs onto the construction and vice versa as previously shown in literature. However, one can argue that according to the results, the learners who employ only these learning materials might have a harder time to generalize the patterns, for instance how the verb *earn* can be used both ditransitively and intransitively [*this earned him a prize* vs *they earned a prize*].

As discussed before, the logarithm transformation of all_av_construction_freq_log was chosen to account for the Zipfian nature of linguistic data (Zipf, 1935). This index calculates the approximate construction frequency e.g., verb object, verb indirectobject directobject and so on, of the corpus against the reference corpus, logarithm transformed. The results for this

[10] These indices sketched here assume that items are attracted to or repelled constructions. Herbst's (2020) proposal of items-in-constructions is inherently the other side of the coin as it presumes that items and constructions arise from usage-events at the same time. Nevertheless, both approaches aim for a probability of an item and construction appearing at the same time, albeit using different measuring methods (i.e., raw frequency and faith formulas).

index demonstrate that the overall average construction frequency is low. Grades 5 to 6 and 6 to 7 show a statistically significant decline and remain roughly stagnant until grade 12 where there is a minor increase. The effect sizes are found to be large, 1.67 and 2.16, respectively. While one might argue that the textbooks would show a decline in constructional variety in earlier grades since the textbooks for these grades are for children and they are compared against a corpus consisting of adult native speakers, this still raises concerns, because as grades advance the decline remains stagnant. From a purely usage-based understanding of language learning that takes frequency as its focus, students should be exposed to various constructions that are available in native speaker corpora consistently.

The index all_lemma_construction_ttr calculates the type-token ratio of lemmas in constructions where the unique instances of lemmas are divided by the total number of constructions. Grades 7 to 8, 8 to 9, and 11 to 12 show an increase in lemma-construction TTR. This is arguably a desired outcome because students experience more unique instances of lemmas (i.e., inflection) in constructions. It is important to note that the effect sizes are of small nature for the 8th to 9th and 11th to 12th grades (8 to 9= 0.13; 11 to 12= 0.44) except for the 7th to 8th grade cohort (1.70). Kyle (2016: 68) mentions that the lemma-construction TTR level is calculated only when the unique instance of an item exists in the reference corpus. Therefore, an increase in the means of this index suggests that students are more likely to experience a lexically more diverse set of lemma-construction combinations, especially in grades 7-8 since the effect size is found to be large. The other two grade cohorts demonstrate small effect size. Nevertheless, this result should be approached with precaution because although the lemma diversity might have statistically increased in above-mentioned grade cutoffs, the lemma diversity still appears in the same set of constructions when one considers the results of other indices. To put it in perspective, the textbook corpus at hand might use *earn* in two different constructions with its different inflections, e.g., *earned*, *earns, earning*, but if the collo-profile of the verb is not fully reflected in the textbook corpus (for instance the intransitive use of the verb), then this could still arguably lead to problems, albeit minor.

All_construction_attested calculates the approximate overlap of constructions in the corpus against the reference corpus. The large effect size (1.72) and the statistical significance in the 10th to 11th grade cohort suggest that the overlap of constructions in the corpus against the reference corpus decreased and students may have likely experienced a narrower array of constructions. The decrease in the approximate overlap of constructions in the grade 10-11 subcorpus remained stagnant in the final cohort as it did not display a statistical increase. This result can also be argued to reflect the

findings for all_av_construction_freq_log since both indices measure construction frequency, albeit from different perspectives.

In the same vein, all_lemma_construction_attested measures the approximate overlap of lemma-construction combinations in the corpus against the reference corpus. Starting with the 7th to 8th grade cohort, the effect size of the decrease was found to be large (2.73) and this stagnation seems to remain in the corpus until the 10th to 11th grade cohort where there was another decrease of large effect (1.72) nature in the lemma-construction combination against the reference corpus. When interpreted in the light of results of the other indices, this decrease should come as no surprise because even if lemma/TTR levels in the corpora increase grade by grade, quite possibly, the construction frequency decreases grade by grade. The lack of construction diversity in the corpus might be reflected onto the results of lemma-construction overlap with a reference corpus. Overall, this suggests that the students keep experiencing fewer lemma-construction combinations with a limited set of constructions over the years which may be detrimental for their language learning journey from a usage-based perspective.

In an ideal scenario, the changes in the developmental history of these textbooks from grade 5-12 would be expected to follow an upward linear trend, with each grade covering more of the natural language. While it would be unfair to criticize the materials based on corpus-driven methodologies especially when the textbooks were not corpus-driven, it is still logical to suggest that an increase in later grades would help students achieve a higher proficiency. For instance, descriptive statistics (Table 2) imply that all_lemma_construction_ttr already has a low percentage, if the percentage of the score is calculated, at grade 5 and it remains stagnant until grade 12. What this suggests is that students have a lower chance of being exposed to a number of lemma-construction combinations (see example sentences 1-6) and as such have a lower probability of recognizing similarities across and arriving at a schematized representation of the construction (see Herbst 2020: 84 for a discussion on recognizing similarities and linking them).

## 6. Discussion

This study compared the development of VACs of English textbooks in Turkey with a usage-based constructionist approach. Building on a freely available corpus by Gedik and Kolsal, this study expanded the corpus to include textbooks starting at grade 5 to grade 12. This paper demonstrated a novel way of analyzing learning materials, namely, the assumption that learning a language subsumes the unification of form and meaning and so should the learning materials. With this approach, researchers in the field can account

for discrepancies in learning materials in greater detail, especially with the help of natural language processing tools.

The results are similar to that of a previous study (Gedik & Kolsal, 2022). In that study, the authors uncovered a lack of variety in high school English learning materials with regard to lexical diversity and syntactic complexity levels. Whether the materials show a gradual increase or decrease regarding VAC development cannot be answered with a yes or no. However, in short, the VAC development of these materials can be certainly described as inconsistent in and across the corpus and do not necessarily show the development or increase they should.

Expanding on the previous study, the results of this present study also portray several implications for the Turkish English language teaching landscape. To begin with, indices that demonstrate statistically significant differences suggest that the learning materials lack a continuous development mainly in the probabilities of verb-construction combinations' occurrence in the textbooks or in other words, the conventionalized usage patterns of a verb-construction combination; they lack in the consistency of constructions' frequency in the corpus; and overall construction overlap of the textbook corpus when compared against a reference corpus. While there are trends in the positive direction in the corpora (for instance all_lemma_construction_ttr and all_approx_collexeme), these trends are mostly overshadowed by negative trends that have a large effect size. In defense of these positive changes, one could argue that students, even if the constructions that they are exposed to are limited especially in reference to a reference corpus (see all_av_construction_freq_log and all_construction _attested), learners still receive input on inflectional changes of verbs, e.g., *eat, ate, eaten, eating, eats*, in whatever constructions they appear. However, from a purely frequency-based point of view to account for the entrenchment of VACs in the learners' minds, the results demonstrate a statistically significant decrease, and it is safe to assume that the textbooks do not provide students with enough of a consistent input for entrenchment to take place. Turning back to the idea of collo-profiles, if the verb-construction combinations do not reflect the real-life usage patterns, then each time a student experiences the usage of a verb incompletely (for instance *earn* in Mono- and Ditransitive constructions but not in an intransitive construction), then the generalization process for this student might lead them to unidiomatic speech or arguably low(er) proficiency, since the schema might not have had enough input to form as other speakers' conventionalized schema. To put it in perspective, if the student experiences the verb *break* as in (9 and 10) but not (11), it might be likely that this student may experience a higher cognitive load when they are asked to describe (decode) or comprehend (encode) a situation that idiomatically or contextually requires the construction in (11).

(9) *The vase is broken*.

(10) *I broke the vase*.

(11) *The vase broke.*

Clearly, the discussion of intransitivity in constructions requires a great deal of research and is out of scope of this paper because this has not been researched from a usage-based approach yet to the researcher's knowledge. Ideally, in a real classroom, frequency cannot and should not be the only factor determining entrenchment (see Divjak (2019) on other determinants). One can argue salience of an item, attention, memory, and other factors to be interacting with the phenomenon of entrenchment. However, if mental representations of constructions (in this case, VACs) are affected and (re)configured in and through usage-events (Diessel, 2016), then one can argue that frequency should not be overlooked. Thus, this could be one of the reasons why Turkish learners of English have consistently ranked low in EPI, among many other reasons as cited previously. From another perspective, if the textbook corpus does not reflect real-life language use when compared against a reference corpus, it could be argued that students may not receive consistent input on the novel uses of the language as in (1). To paraphrase and apply Goldberg's view on entrenchment (2019: 94), if a linguistic item is covered enough in relation to other linguistic items, this item will become easier to access and this process will result in what we call IDIOMATIC LANGUAGE. This will ideally reinforce the form-meaning pairings and their respective forms and meanings. Ultimately, if we want to achieve near-native proficiency, this should include the ability of using language in novel and idiomatic ways since this is what we also do in our L1s.

One limitation to keep in mind when interpreting the results is that while it is possible to interpret statistically significant differences across subcorpora, i.e., across grades, and come to conclusions about how these changes might affect students' learning, the data is thus far only hypothetical, and it would be difficult to determine the real-life implications in classrooms without longitudinal studies (see however Madlener (2016) for short term effects of input optimization in teaching constructions). Another limitation is arguably the size of the corpus. Future studies should take these limitations into consideration.

Nevertheless, to improve this condition, textbook creators should be made aware of new advances in applied linguistics such as NLP tools that can detect variation of any construction, and also empirical findings that point at the fact that grammar and lexis are learned and stored together in chunks, i.e., constructions. In order to account for the ever-changing nature of language and linguistic patterns, textbook creators should employ reliable corpora, statistical measures, empirical findings in linguistics and software tools to create English textbooks. This is by no means an easy task and

should be handled with care. But as more work is done, this newly emerging field, applied construction grammar, can ultimately help English textbook creators create more cognitively plausible textbooks that account for the usage-based nature of language learning which might lead to more proficiency in learners.

## 7. Conclusion

This study analyzed the English textbooks in Turkey from grade 5 to grade 12 using a usage-based constructionist framework with a software tool. More specifically, the verb-argument constructions (VACs) and their development, i.e., their frequency, throughout grades were analyzed. The results replicate the findings of a similar study (Gedik & Kolsal, 2022) and indicate that students who use these textbooks to learn English at schools might be at a disadvantage. This is because the textbooks for the most part do not show a gradual increase in linguistic variation or frequency of VACs. Some of the key findings are: (i) VACs lack variety, in other words, the textbooks repeat the same VACs, (ii) for the present VACs it is likely that some of them do not occur frequently enough for entrenchment to take place, and (iii) this lack of frequency and variety might result in a failure of arriving at an overarching generalization for certain VACs, i.e., the ditransitive construction. As such, these findings might be one of the many reasons as to why Turkish speakers of English, i.e., students, consistently lack English proficiency and they might also explain their poor use of idiomatic expressions.

## References

Arıkan, Arda (2005). Age, gender and social class in ELT coursebooks: A critical study. *Hacettepe Üniversitesi Eğitim Fakültesi Dergisi* 28: 29–38.

Baayen, R. Harald, Petar Milin, Michael Ramscar (2016). Frequency in lexical processing. *Aphasiology* 30(11): 1174–1220.

Bencini, Giulia M.L, Adele E. Goldberg (2000). The contribution of argument structure constructions to sentence meaning. *Journal of Memory and Language* 43(4): 640–651.

Biber, Douglas, Bethany Gray, Kornwipa Poonpon (2011). Should we use characteristics of conversation to measure grammatical complexity in L2 writing development? *TESOL Quarterly* 45(1): 5–35.

Carter, Ronald, Michael McCarthy (2006). *Cambridge Grammar of English: a Comprehensive Guide; Spoken and Written English Grammar and Usage.* Stuttgart: Ernst Klett Sprachen.

Chang, Franklin, Kathryn Bock, Adele E. Goldberg (2003). Can thematic roles leave traces of their places? *Cognition* 90(1): 29–49.

Chater, Nick, Christopher D. Manning (2006). Probabilistic models of language processing and acquisition. *Trends in Cognitive Sciences* 10(7): 335–344.

Cohen, Jack (1988). *Statistical Power Analysis for the Behavioral*. Hillsdale, N.J.: L. Erlbaum Associates.

Coşkun, Abdullah (2016). Causes of the 'I can understand English but I can't speak' syndrome in Turkey. *I-Manager's Journal on English Language Teaching* 6(3): 1–12.

Davies, Mark (2010). The corpus of contemporary American English as the first reliable monitor corpus of English. *Literary and Linguistic Computing* 25(4): 447–464.

Demir Yusuf, Mustafa Yavuz (2017). Do ELT coursebooks still suffer from gender inequalities? A case study from Turkey. *Journal of Language and Linguistic Studies* 13(1): 103–122.

Diessel, Holger (2016). Frequency and lexical specificity in grammar: A critical review. In Behrens Heike, Stefan Pfänder, eds. *Experience Counts: Frequency Effects in Language*. Berlin: De Gruyter Mouton, 209–238.

Díez-Bedmar, María Belén, Pascual Pérez-Paredes (2020). Noun phrase complexity in young Spanish EFL learners' writing. *International Journal of Corpus Linguistics* 25(1): 4–35.

Divjak, Dagmar (2019). *Frequency in Language: Memory, Attention and Learning*. Cambridge: Cambridge University Press.

Divjak, Dagmar, Catherine Caldwell-Harris L. (2015). Frequency and entrenchment. In Dabrowska Ewa, Dagmar Divjak, eds. *Cognitive Linguistics. Foundation of Language*. Berlin: De Gruyter, 61–86.

Dolgova, Natalia, Andrea Tyler (2020). Applications of usage-based approaches to language teaching. In Gao Xuesong, ed. *Second Handbook of English Language Teaching*. Manhattan: Springer International Publishing, 1–23.

Education First EF English Proficiency Index (EPI). *A ranking of 100 countries and regions by English skills.* https://www.ef.com.tr/assetscdn/WIBIwq6RdJvcD9bc8RMd/legacy/__/~/media/centralefcom/epi/downloads/full-reports/v10/ef-epi-2020-english.pdf

Ellis, Nick C. (2002). Frequency effects in language processing: A review with implications for theories of implicit and explicit language acquisition. *SSLA*. 24: 143–188.

Ellis, Nick C. (2008). The dynamics of second language emergence: Cycles of language use, language change, and language acquisition. *The Modern Language Journal* 92.2: 232–249.

Ellis, Nick C., Fernando Ferreira-Junior (2009a). **Constructions and their acquisition:** Islands and the distinctiveness of their occupancy. *Annual Review of Cognitive Linguistics* 7(1): 188–221. https://doi.org/10.1075/arcl.7.08ell

Ellis, Nick C., Fernando Ferreira-Junior (2009b). Construction learning as a function of frequency, frequency distribution, and function. *The Modern Language Journal* 93(3): 370–385.

Erarslan, Ali (2019). Factors affecting the implementation of primary school English language teaching programs in Turkey. *Journal of Language Teaching and Learning* 9(2): 7–22.

Fisher, Ronald Aylmer (1934). Two new properties of mathematical likelihood. *Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical Character*. 144(852): 285–307.

Gedik, Tan Arda, Yağmur Su Kolsal (2022). A corpus-based approach to English university entrance exams and English high school textbooks in Turkey. *TAPSLA* 8(1). https://doi.org/10.31261/TAPSLA.9152

Goldberg, Adele E. (1995). *Constructions: A Construction Grammar Approach to Argument Structure Constructions.* Chicago: University of Chicago Press.

Goldberg, Adele E. (2006). *Constructions at Work: The Nature of Generalization in Language.* Oxford: Oxford University Press.

Goldberg, Adele E. (2013). Argument structure constructions versus lexical rules or derivational verb templates. *Mind & Language* 28(4): 435–465 https://doi.org/10.1111/mila.12026

Goldberg, Adele E. (2019). *Explain Me This: Creativity, Competition, and the Partial Productivity of Constructions.* New Jersey: Princeton University Press.

Goldberg, Adele E., Devin M. Casenhiser, Nitya Sethuraman (2004). Learning argument structure generalizations. *Cognitive Linguistics* 15(3): 1613-3641.

Goldberg, Adele E., Thomas Herbst (2021). The nice-of-you construction and its fragments. *Linguistics* 59(1): 285–318.

Green, Clarence (2019). A multilevel description of textbook linguistic complexity across disciplines: Leveraging NLP to support disciplinary literacy. *Linguistics and Education* 53: 1-11.

Gries, Stefan Th. , Beate Hampe, Doris Schönefeld (2005). Converging evidence: Bringing together experimental and corpus data on the association of verbs and constructions. *Cognitive Linguistics* 16(4): 635-676

Gries, Stefan Th., Stefanie Wulff (2005). Do foreign language learners also have constructions? *Annual Review of Cognitive Linguistics* 3(1): 182–200.

Hare, Mary L, Adele E Goldberg (1999). Structural priming: Purely syntactic? In *Proceedings of the twenty first annual conference of the Cognitive Science Society*. Psychology Press, 208–211

Herbst, Thomas (2018). Is language a collostructicon? A proposal for looking at collocations, valency, argument structure and other constructions. In Cantos-Gómez Pascual, Moisés Almela-Sánchez, eds. *Lexical Collocation Analysis.* Manhattan: Springer International, 1–22.

Herbst, Thomas (2020). Constructions, generalizations, and the unpredictability of language: Moving towards colloconstruction. *Constructions and Frames* 12(1): 56–95.

Kyle, Kristopher (2016). *Measuring syntactic development in L2 writing: Fine grained indices of syntactic complexity and usage-based indices of syntactic sophistication*. Georgia State University. http://scholarworks.gsu.edu/alesl_diss/35/

Kyle, Kristopher, Scott Crossley (2017). Assessing syntactic sophistication in L2 writing: A usage-based approach. *Language Testing* 34(4): 515-535.

Kyle, Kristopher, Scott Crossley (2018). Measuring syntactic complexity in L2 writing using fine-grained clausal and phrasal indices. *The Modern Language Journal* 102(2): 333–349.

Larsen-Freeman, Diane (2009). Adjusting expectations: The study of complexity, accuracy, and fluency in second language acquisition. *Applied Linguistics* 30(4): 579–589.

Lee, David, John Swales (2006). A corpus-based EAP course for NNS doctoral students: Moving from available specialized corpora to self-compiled corpora. *English for Specific Purposes* 25(1): 56–75.

Lieven, E. V., Julian M. Pine, Gillian Baldwin (1997). Lexically-based learning and early grammatical development. *Journal of Child Language* 24(1): 187–219.

Madlener, Karin (2016). Input optimization: Effects of type and token frequency manipulations in instructed second language learning. In Behrens Heike, Stefan Pfänder, eds. *Experience Counts: Frequency Effects in Language*. Berlin: De Gruyter, 133–174. https://doi.org/10.1515/9783110346916-007.

MEB (2018a). İngilizce öğretim programı ilkokul ve ortaokul. https://mufredat.meb.gov.tr/Dosyalar/201812411191321-%C4%B0NG%C4% B0L%C4%B0ZCE%20%C3%96%C4%9ERET%C4%B0M%20PROGRAMI%20Klas%C3%B6r%C3%BC.pdf

MEB (2018b). Ortaöğretim İngilizce dersi. http://mufredat.meb.gov.tr/ProgramDetay.aspx?PID=342

Michaelis, Laura A. (2004). Type shifting in construction grammar: An integrated approach to aspectual coercion. *Cognitive Linguistics* 15.1: 1–67.

Monsell, Stephen (1991). The nature and locus of word frequency effects in reading. In Besner Derek, Glyn W. Humphreys, eds. *Basic Processes in Reading*. London: Routledge, 156–205.

Nakamura, Daisuke (2019). Development of clausal and subclausal grammatical complexity and their relationship to overall length in second language writing over a year. *Bulletin of Japan Women's College of Physical Education* 49: 45–57.

Ninio, Anat (1999). Pathbreaking verbs in syntactic development and the question of prototypical transitivity. *Journal of Child Language* 26(3): 619–653.

Perek, Florent (2015). *Argument Structure in Usage-Based Construction Grammar*. Amsterdam: John Benjamins. https://doi.org/10.1075/cal.17

Römer, Ute, Audrey Roberson, Matthew B. O'Donnell, Nick C. Ellis (2014). Linking learner corpus and experimental data in studying second language learners' knowledge of verb-argument constructions. *ICAME Journal* 38(1): 115–135.

Römer, Ute, Matthew B O'Donnell, Nick C Ellis (2015). Using COBUILD grammar patterns for a large-scale analysis of verb-argument constructions. In Groom Nicholas, Maggie Charles, Suganthi John, eds. *Corpora, Grammar and Discourse: In Honour of Susan Hunston*. Amsterdam: John Benjamins, 43-72

Seidenberg, Mark S., David C. Plaut (2014). Quasiregularity and its discontents: the legacy of the past tense debate. *Cognitive Science* 38(6): 1190–1228.

Stefanowitsch, Anatol, Stefan Th Gries (2003). Collostructions: Investigating the interaction of words and constructions. *International Journal of Corpus Linguistics*. 8(2): 209–243.

Stefanowitsch, Anatol, Susanne Flach (2018). The corpus-based perspective on entrenchment. In Hans-Jorg Schmid, eds. *Entrenchment and the Psychology of Language Learning. How we Reorganize and Adapt Linguistic Knowledge*. Berlin: De Gruyter Mouton, 101–127.

Tabachnik, Barbara G., Linda Fidell S. (2014). *Using Multivariate Statistics*. Harlow, Essex: Pearson Education.

Tomasello, Michael (2000). First steps toward a usage-based theory of language acquisition. *Cognitive Linguistics* 11(1–2): 61–82.

Tomasello, Michael (2003). *Constructing a Language: A Usage-Based Theory of Language Acquisition*. Cambridge: Harvard University Press.

Wulff, Stefanie, Nick C. Ellis (2018). Usage-based approaches to second language acquisition. In Miller David, Fatih Bayram, Jason Rothman, Ludovica Serratri-

ce eds. *Bilingual Cognition and Language. The State of the Science Across its Subfields*. Philadelphia: John Benjamins, 37–56.

Yates, F. (1934). Contingency tables involving small numbers and the χ 2 test. *Supplement to the Journal of the Royal Statistical Society* 1(2): 217-235 https://doi.org/10.2307/2983604

Yurtsever Bodur, Gözde, Arda Arıkan (2019). Why can't we learn English?: Students' opinions at Akdeniz University. *Gaziantep University Journal of Educational Sciences* 1(1): 1–7.

Zipf, George Kingsley (1935). *The Psycho-Biology of Language: An Introduction to Dynamic Philology*. Boston: Houghton Mifflin.

## Author's address

Tan Arda Gedik
Friedrich-Alexander-Universität Erlangen-Nürnberg,
Schloßplatz 4, 91054 Erlangen, Germany
e-mail: tan.gedik@fau.de

Received: October 25, 2021

Accepted for publication: December 15, 2021