

A Usage-Based Constructionist Approach to Evidentiality in Turkish: The Unevidentiality Construction

Türkçe'de Kanıtsallığa Kullanıma Dayalı Yapı Gramer Yaklaşımı:
Kanıtsallık Yapısı

Tan Arda Gedik*

Friedrich-Alexander Universität Erlangen-Nürnberg

Abstract

Coupled with corpora, usage-based construction grammar aims to provide cognitive plausibility for linguistic phenomena. In this vein, this paper combines construction grammar and usage-based approaches to analyze evidentiality in Turkish. While Turkish has been analyzed from a usage-based perspective, evidentiality has not been taken up in a usage-based constructionist approach. By using corpora, association measures, and construction as a notion and a framework, this paper defines the Unevidentiality Construction in a taxonomic space. First, it outlines its semantic properties and then it uses association measures such as faith, delta (Δ) p, and ITECX to determine its usage pattern and statistical biases based on corpora. The paper demonstrates a superordinate and lower-level, item-specific instantiations of the construction. The results from association measures and the family of unevidentiality constructions can serve for future linguistic endeavors.

Keywords: construction grammar, usage-based approach, Turkish, evidentiality

Öz

Derlemler ile birleştğinde, kullanıma dayalı yapı gramer, dilsel fenomenler için bilişsel anlamda makul olmayı amaçlar. Bu bağlamda, bu makale Türkçede kanıtsallığı analiz etmek için yapı gramer ve kullanıma dayalı yaklaşımları bir araya getirmektedir. Türkçe kullanım temelli bir bakış açısıyla incelenmiş olsa da kullanım temelli yapı gramerci bir yaklaşımda kanıtsallık şu ana dek ele alınmamıştır. Derlemler, ilişkilendirme ölçüleri ve 'yapıyı' bir kavram ve çerçeve olarak kullanan bu makale, Kanıtsallık Yapısını taksonomik bir şekilde tanımlamaktadır. İlk olarak anlamsal özelliklerini ana hatlarıyla belirtir ve daha sonra faith, delta (Δ) p ve ITECX gibi ilişkilendirme ölçütlerini kullanarak kullanım modelini ve derlemlere dayalı istatistiksel tercihlerini açıklar. Makale bu yapının üst ve alt düzey, ögeye özel örneklerini göstermektedir. İlişkilendirme ölçümlerinden ve kanıtsallık yapıları ailesinden elde edilen sonuçlar gelecekteki dilbilim çalışmalarına hizmet edebilir.

Anahtar Kelimeler: yapı gramer, kullanıma dayalı dilbilim, Türkçe, kanıtsallık

Introduction

Usage-based approaches to language contrast to what generativist approaches offer to say in their assumptions about language. Briefly explained, usage-based approaches assume that language unfolds in and through usage events and is not dependent on an inherent grammar system, i.e., Universal Grammar. In other

words, language emerges from usage, frequency, and the interaction of items. Second, usage-based approaches mostly embody a lexicogrammatical view of language unlike generativists. This means that form is not independent of function and vice versa. One of the leading proponents of a lexicogrammatical view taken up in usage-based approaches is construction grammar (CxG) by Adele E. Goldberg. Finally, because usage-based approaches are based on usage, they analyze frequency data. Studies done in this framework mostly do away with employing an introspective analysis and utilize corpora.

Over the years, English and German have attracted quite a lot of attention in usage-based approaches. However, languages like Turkish have not received the same amount of attention to the same extent as those languages (however see Akkuş 1-13; Durant 1-38; Kiraz 1-15; Yılmaz 1-4; Römer and Yılmaz 108-109 to name a few for usage-based accounts of linguistic phenomena in Turkish).

Turning our attention to the focus of the study, (un)evidentiality, being one of the more popular topics in linguistic research, has attracted a lot of attention (e.g., Papafragou et al. 253-255). Evidentiality in Turkish as a definition presented at the end of this study is best captured as reporting how the information was obtained when making an utterance (e.g., Aijmer 63), however, it has been defined in various other ways (e.g., Banguoğlu 271; Gencan 423; Lewis 122-124; Cinque 47-60). While evidentiality has been analyzed from corpus-based perspectives in bilingual (Arslan and Bastiaanse 1), heritage speakers (Kaya-Soykan et al. 1), in Cypriot Turkish (Işık-Taş and Sağın-Şimşek 1), or acquisition of it in children (Aksu-Koç et al. 14; Uzundağ et al. 403), it has not been analyzed from a usage-based constructionist perspective to account for its taxonomic construction family. In one way, then, this paper serves constructicographic purposes subscribing to the tenets of such an approach, although there exist other corpus-based evidentiality studies, as briefly mentioned above. It also uncovers its statistical biases for verbs. Thus, in this paper, I aim to introduce a usage-based CxG account for the analysis of the unevidentiality construction in Turkish by using two corpora (OPUS2 Turkish Corpus and TrWaC), explain its form and meaning in a unified approach, i.e., lexicogrammar, and uncover its statistical preferences for the verbal slot to address the apparent research gap. However, what I do not intend to do in this paper is to convince readers that one approach is better than the other. This is merely a novel way of approaching a phenomenon in Turkish using a different view for opening other linguistic endeavors explained at the end.

Construction Grammar and Association Measures

Starting in the 1970s, especially with the work of John Sinclair and Ronald Langacker, researchers sought different ways of approaching linguistic phenomena, other than what the generativist approaches had to offer. What is meant by generativist approaches in this article is best captured by Guasti's (2-39) account of it. Briefly, the generativist approaches assumed that language consisted of separate and autonomous modules, i.e., syntax, lexis, pragmatics, morphology, phonology and other subsections. In this line of thought, researchers assumed that language was an innate capability of humans which

was acquired by means of the Language Acquisition Device. This hypothetical device analyzed the properties of ambient language then used Universal Grammar as a cloud-based storage, metaphorically speaking, to derive grammar rules of the respective language. This, however, arguably overlooks linguistic experience, exposure, and usage-events. For generativists, language is an organization that can be explained as minimally as possible (Chomsky, chapter 4). Consequently, this resulted in phrase structures (e.g., VP, NP and so on) and rules that attempted to generate the whole of language with as economically as possible. With the advent of powerful computers and corpora, researchers realized that language was in fact quite repetitive and not as never-heard-before as the generativist approaches put forward (see Dąbrowska 1-13 for a detailed discussion).

(1) *She*_{NP} *sneezed*_{VERB} *the foam*_{OBJ} *off the cappuccino*_{OBLIQUE}. (Goldberg, *Constructions at Work* 42)

As exemplified in (1), *sneeze* can be combined in a novel and meaningful way. From a generativist point of view, *sneeze* would be given two separate lexical entries in the mind, one with an intransitive and one with a transitive usage. These entries would then be put into a sentence by means of merge if its complementation requirements are met. This is arguably a verbocentric view.

Goldberg (*Constructions at Work* 42) names this (example 1) the caused-motion construction. This construction has such semantic properties (function) that when semantically coherent verbs¹ combine with it, it will result in the meaning “X moves Y along Z.” As such, CxG distributes the labor of creating meaning between constructions of different abstractions, e.g., verbs and schemas. This and many other similar findings from Langacker (*Foundations in Cognitive Grammar*, 27-42), and Bybee (*Morphology*, 81-109) led to a different understanding of language: a lexicogrammatical continuum (figure 1). On one side of this continuum, there are items that look like words, prefixes, and suffixes and on the other, there are items that are partially filled, idiomatic or fully schematic, i.e., the caused-motion construction. Another important aspect of CxG is, unlike what generative grammar postulates, constructions do not emerge from derivation. Each sequence experienced in ambient language is a construction of its right, or a less-abstract construction of a more highly abstracted construction (see Goldberg, *Constructions at Work* 45-68). To give an example, according to Herbst (“Constructions, generalizations, and the unpredictability of language” 69) because *give* is used almost 50% of the time in the ditransitive construction, e.g., *I gave her a book*, there could be a *give*-ditransitive construction, a less abstracted version of the highly abstracted ditransitive construction, since *give* is one of the most prototypical verbs that carry the meaning of the ditransitive construction. This is because the verbal slot in the *NP VERB OBJ OBJ* would be filled up by *give*, resulting in *NP GIVE OBJ OBJ*. As such, CxG argues that what linguists generally consider as grammar, i.e., the right-hand side of figure 1, carries meaning just as other items (see example 1).

¹ See Goldberg (*A Construction Grammar Approach to Argument Structure Constructions* 50) for the semantic coherence principle.

In other words, they were not passive in the creation of meaning, but had semantic/pragmatic functions. Constructions come in different sizes, as seen in figure 1. Following Goldberg (*Constructions at Work* 5), constructions are form-meaning pairings that “occur with sufficient frequency.”

Figure 1, the lexicogrammatical continuum

Prefixes, suffixes	Words	Fixed constructions	Partially-fixed constructions	Constructions with fixed items	Highly abstract constructions
<i>De-, re-, -ing, -ify..</i>	<i>Cat, dog, bird</i>	<i>Kick the bucket</i>	<i>X called, X wants Xs Y back</i> <i>1995 called, it wants its cord back</i>	<i>The Xer, the Yer</i> <i>The more, the merrier</i>	<i>NP VERB OBJ OBJ</i> <i>I gave her a book</i>
<i>-miş, na-, -Dir..</i>	<i>Kedi, köpek, kuş</i>	<i>Yangına körükle gitmek</i>	<i>NP vezir de eder rezil de eder</i> <i>Aile vezir de eder rezil de eder</i>	<i>Ne kadar X, o kadar Y</i> <i>Ne kadar ekmek, o kadar köfte</i>	<i>NP OBJ OBJ VERB</i> <i>Ben ona kitabı verdim</i>

In this approach, language is not a set of a priori rules but rather a dynamic web of interrelated signs, i.e., both grammar and lexis, that unfold over time and that are learned through domain-general cognitive abilities (Tomasello 144-193; Divjak 97-155) which are thought to be an innate capacity (e.g., attention, memory, automation, and abstraction). Thus, language learning is likened to learning any other skill, and does not require a special faculty. Thus, usage-based linguists formulate that language learning is based on exposure to a set of highly repetitive chunks via domain-general cognitive abilities, which help with the learning of other lower-frequency constructions (Goldberg, *Constructions at Work* 69-92) and generalizations occur due to frequency effects, as speakers try to predict what will come next in ambient language.

Usage-based studies² have been gradually taken up in Turkish linguistics over the last decade. The studies seem to be scattered across contact-linguistics (Akkuş 1-13; Backus and Demirçay 13-15), applied linguistics (Kiraz 1-15; Yılmaz 1-4; Römer and Yılmaz 108-109), acquisition studies (Altınkamaş-Altan 69-91, Ordem 190-195), formulaicity (Durant 1-38), constructicographically (Gedik, turkishconstruction.wordpress.com), and typology (Fried and Östman 11-86; Kawaguchi 247-268; Yılmaz 269-286) to name a few. While there are a few usage-based studies on Turkish suffixes (Durant 1-38; Karayayla 753-754), there do not seem to be many on evidentiality. One exception however is Aksu-Koç, Ögel-Balaban and Alp (13-28). In their seminal work, they focus on the

² Studies that subscribe to the main tenets of a usage-based approach and not just employ a corpus in the analysis.

learning of evidentiality in Turkish from a usage-based perspective in native speaker children. Another similar but earlier study is Aksu-Koç (15-28). These studies are important as they provide cognitive plausibility into child language learning by means of usage-based approaches. However, one important point that needs attention is that while it has been studied from usage-based approaches to the researcher's knowledge no other study to this date has examined the unevidentiality construction (UnCx) in Turkish from this perspective, i.e., usage-based construction grammar. Bridging this gap is important because it can give insight into the frequency profile of this construction and can serve as a reference work for future studies on other phenomena, for instance determining the productivity of this construction, translation purposes, or determining L1-L2 entrenchment levels and interference of the UnCx (see Goschler and Stefanowitsch 1 for a similar study in German and English). It can also serve for constructicographic purposes.

Turning our attention to frequency effects on language, terms such as entrenchment, and statistical preemption have been offered as mechanisms behind generalizations to capture frequency effects and in speakers. Starting with the former term, entrenchment suggests that the more an item is experienced, the more easily it is retrieved. The item to be entrenched can occur in varying shapes and sizes and it will gradually become easier to process. There is also ample evidence that points at a correlation between high entrenchment levels and easier accessibility, retrieval, and cognitive salience levels (Bybee, *Language* 33-57). Entrenchment has been associated with frequency levels (Bybee, *Morphology* 117; Langacker, *Foundations of Cognitive Linguistics* 59). In other words, high frequency levels of an item might suggest high degrees of conventionalization. Generalizations occur because as speakers see the usage of a construction in a particular social context with frequency, and they test it in similar social contexts instead of using a competing construction (e.g., the use of *-yor* versus *-Ir*), see Goldberg (*Explain me this* 51-94). Measuring entrenchment can be done using and analyzing data from acceptability tests and corpora.

Statistical preemption, on the other hand, is an error-avoidance system, as it will block the production of unattested constructions (see Boyd and Goldberg 55-83 as an example). To illustrate, one can suggest that learners record information³ on preemption whenever they see *she gave him the book* and *she gave the book to him*. Upon collecting the information, speakers then arrive at a general principle that blocks the use of TO-DATIVE in certain contexts and blocks the use of the DITRANSITIVE in others (see Perek 79-89 for a lengthy discussion on the usage conditions of these above-mentioned constructions). Preemption is therefore a powerful tool that extracts negative indirect evidence from the input. In other words, the non-existence of an item is also evidence.

From Goldberg's (*Explain me this* 122-123) perspective, entrenchment and statistical preemption are two entangled phenomena that are difficult to pick apart. Moreover, she presents evidence for how these two complement one

³ See Bybee, *Language* 14-33.

another. Thus, instead of differentiating between the two, she combines both terms under “entrenchment” and simply names them simple entrenchment for the effects of entrenchment, and conservatism via entrenchment for the effects of statistical preemption (Goldberg, *Explain me this* 122-123). In this study, I will subscribe to this line of thought.

Finally, item-specificity, as opposed to overarching generalizations are one of the central topics in CxG. Goldberg (*Constructions at Work* 12) acknowledges the need for a theory that can both accommodate item-specific knowledge and generalizations. Figure 2 is a visualization of this continuum of item-specificity/generalizations. For instance, as will be discussed later, while the UnCx would situate itself on the left-hand side of the continuum as it has a highly schematic schema, e.g., verb+(I)mlş, the hearsay-UnCx would be in the middle with its partial generalizability, e.g., *bana böyle demiş, - öyle yapmış - öyle mi yapmış?* This is because it either appears with reporting verbs such as *de-, söyle-* and so on, or it appears in combination with the preceding sentence, which makes it partially item-specific, partially open to higher generalizations. That is why researchers have advocated for lower-level constructions (Herbst “The Status of Generalizations” 347-368, “Is Language a Collostruction?” 1-22, “Constructions, generalizations, and the unpredictability of language” 56-80; Perek 105-111), e.g., the ditransitive construction → *give*-ditransitive construction. Alongside this, Herbst (“Constructions, generalizations, and the unpredictability of language” 58-90) argues for an items-in-constructions (ITECXs) approach to constructions, which sketches out the usage or frequency-profile of a construction. In his words, ITECXs indicate and capture “abstractions over many many usage events (all of which contain items)” (“Constructions, generalizations, and the unpredictability of language” 83). As such, it is possible to capture frequency effects, preemption and entrenchment, to identify how important the item is for the construction and how important the construction is for the item using raw frequencies. Herbst (“Constructions, generalizations, and the unpredictability of language” 67) calculates ITECX frequencies as “ITECX1: the proportion of a particular item as opposed to other items occurring in the same slot of the construction: ITEaCXA: ITEa-zCXA, ITECX2: the proportion of uses of a particular item in a construction as opposed to its use in other constructions: ITEaCXA: ITEaCXA-Z.” In other words, ITECX1 is calculated as follows: divide the raw frequency of the item by however many other items occur in the same slot. However, one problem in this approach for Turkish is the difficulty of determining all constructions that use a specific item, e.g., *demek ki (then)*. Since there is not a reliable constructicon⁴ for Turkish, ITECX2 will not be used in the analysis.

⁴ A constructicon is a reference work based on the assumptions of construction grammar. In other words, it is a lexicon, but it does not only include lexical items but the entirety of the lexicogrammatical continuum.

Figure 2, generalizations continuum (adapted from Herbst “Constructions, generalizations, and the unpredictability of language” 59)

overarching generalizations	partial generalizations	item-specific knowledge
-----------------------------	-------------------------	-------------------------

There are other measures researchers can use to identify how strongly the item and the construction are related to one another. In addition to ITECXs, there is delta (Δ) p, faith, and collostructional analysis (Gries and Stefanowitsch 209-243). However, this study uses the first three except collostructional analysis because ITECX already gives a collostruction-like insight into the construction.

Faith scores measure how faithful a verb (or any other item) is to a particular construction (Gries et al. 644-645). It is also possible to measure the faithfulness of a construction to a verb in the same vein. In other words, it measures the probability of a verb appearing in a particular construction (e.g., the ditransitive, the passive, the caused-motion construction to name a few). Faith scores are calculated using $\left(\frac{a}{a+b}\right)$. To put it in perspective, Kyle and Crossley (525) calculate the faithfulness of the verb ‘have’ in the transitive construction as 17.7%. This means that ‘have’ has a 17.7% probability that it will appear in the transitive construction. Measuring faith scores for the same verb in comparison to other competing constructions in the corpus can give a better insight into its usage pattern. By using faith scores, it is possible to determine which verbs have a higher chance of occurring in the UnCx.

ΔP , which is another bidirectional approach and a variant of faith, predicts the likelihood of a construction being used when triggered by a cue, i.e., a verb, and this score is deduced by the likelihood of the construction being used without the cue. Kyle and Crossley (525) calculate this with the following formula: $\left(\frac{a}{a+b}\right) - \left(\frac{c}{c+d}\right)$.⁵ In the same vein, they explain that the likelihood of ‘have’ appearing in the transitive construction is higher than the likelihood of the transitive construction appearing with another verb. The authors illustrate this as .177 (the likelihood of ‘have’ appearing in the transitive construction) - .053 (the likelihood of the transitive construction appearing without *have*) = .124 (.124 > .053). With this approach, it is possible to see which verbs have a higher chance of occurring with the UnCx than the others.

The Unevidentiality Construction (UnCx) and the Research Gap

The UnCx analyzed here has been discussed with approaches from structuralist linguistics, i.e., no lexis-grammar continuum, no form-function unity. For ease of referring to the description of the construction (cx) in other studies, the *-(I)mİş* notation will be used. Before moving further, it is important to note that, in this augmentation of CxG to Turkish, each suffix, i.e., *-yor*, *-Ir* and so on, is a construction that is nested within an argument structure alongside a verb. Figure 3 for sentence (2) reflects the idea (for space purposes, the focus is on

⁵ See the appendix to see what a, b, c, and d represent in this analysis.

the verb, for notating and illustrating the constructions, I follow Herbst and Hoffmann 197-218⁶):

(2) *Şu an siz-in ev-in ön-ün-de-yim de-miş.*

Now you-GEN house-GEN front-GEN-LOC-1PERSON say-3PERSONEVID
'She/He said she/he is in front of your house'

Figure 3, the quotative-hearsay-UnCx

"Sizin evin önündeyim"	<i>De</i>	<i>Miş</i>	The quotative-hearsay-UnCx ⁷
[The quotation-block cx]	[the Verb-root cx]	[the UnCx suffix cx]	

This construction has been researched by several authors from a usage-based perspective (Aksu-Koç et al. 14-28; Aksu-Koç 15-28; Işık & Sağın Şimşek 1). Aksu-Koç et al. (22) discuss that children learn the UnCx with 95% accuracy by the age of six. They postulate two versions of the UnCx, namely reportative and inferential. Some other researchers name this construction to be a 'hearsay' marking with past tense properties (e.g., Banguoğlu 271; Gencan 423). Other researchers (e.g., Lewis 122-124; Cinque 47-60) claim that this construction has inferential past tense properties. Alongside these authors, there are also studies that elaborate on the construction's aspectual properties, namely, its nature of completeness, i.e., perfectivity (Underhill 169-175; Lewis 122-124). Işık and Sağın Şimşek (1), using a corpus, sketch the differences of the usage of this construction between Turkey Turkish and Cypriot Turkish. They note that the Cypriot abstraction of the construction has changed in its pragmatic functions, reference to past, and its inferentiality due to language and dialect contact between Cypriot Turkish and Greek and Turkey Turkish.

While the construction itself may have been well-defined in terms of its form and function separately, i.e., tense and aspect, a fusion of its form and function with its frequency profiling is missing. To account for the construction, I will follow a holistic approach and suggest a network or family of the UnCx from a usage-based constructionist approach. The network has a taxonomic structure and has inheritance links. That is, the network does not imply a sense of hierarchy but rather a categorization of the construction's abstraction. Inheritance links ensure that the lower levels⁸ of the network also inherit the overarching properties of the higher-level constructions and that lower levels can add new features to themselves. For all purposes, I suggest that the construction is best analyzed as a suffix on a high level, and lower-level constructions (Perek 105-111) on smaller levels (see figure 5), i.e., the

⁶ Each building block, i.e., the root, suffix, prefix and so on, is taken as a construction which then form a bigger construction.

⁷ *De-* (to say) as an item semantically presupposes that somebody collected hearsay evidence from someone else and thus can be categorized as the hearsay UnCx.

⁸ Lower or higher level refers to the degree of abstraction of the construction in this context.

quotative-hearsay-UnCx. Semantic properties will be compiled from the OPUS2 Turkish corpus and frequency data for association measures will be compiled from TrWaC for reasons explained in the methodology section. In this study, I prefer to call this phenomenon unevidentiality as it suggests both un- and evidentiality. This, however, is merely a preference in notation. It may well be named the evidentiality construction.

The semantics of evidentiality that emerged from a corpus-based analysis in this study is compared against Plungian's (353) and Aksu-Koç et al's (14-16) research. Their proposals of evidentiality assure that there is a difference between personal and impersonal evidence, namely reportative and inferential.

Methodology

The study uses the OPUS2 Turkish Corpus,⁹ which is freely available on SketchEngine, to retrieve two randomized sets of 100 sentences to do a semantic analysis, i.e., a total of 200 sentences, using the following CQL query: [tag="V.*" & word=".*miş" | word=".*miş" | word=".*muş" | word=".*müş" | word=".*mişim" | word=".*mişsin" | word=".*mişlar" | word=".*mişiz" | word=".*mişsiniz"] [word=="." | word==","]. To run a query for everything else except *-(I)mış*, the following CQL query was run [tag="V.*" & word!=".*miş" | word!=".*miş" | word!=".*muş" | word!=".*müş" | word!=".*mişim" | word!=".*mişsin" | word!=".*mişlar" | word!=".*mişiz" | word!=".*mişsiniz"] [word=="." | word==","]. The corpus was preferred over other corpora such as the Turkish National Corpus, and TrTENTEN. The reason behind this was that they either lacked part-of-speech tagging or CQL query, which made the analysis almost impossible, or that the corpora compiled were not clean and had duplicates. The OPUS2 Turkish Corpus is an amalgamation of subtitles, newspapers, and documentation and is well-balanced. The OPUS2 corpora are parallel corpora, which means that the texts that a corpus has most likely appear in another language. It is also reliable because the translated documents were proof-read and edited by native speakers in the target language(s). However, to calculate association measures, TrWaC¹⁰ was preferred as the OPUS2 corpus did not have lemmatization available at the time of the study. Thus, data for association measures were gathered from TrWaC. Nevertheless, because the lemmatization of TrWaC was not reliable and brought up other non-verb results even with the part-of-speech tag embedded in the query, I only calculated association measures for those items that occurred at the end of a sentence or before a comma, as they were most likely verbs, which was ensured by manually checking via lemmatization and KWIC. As such, the total number of hits for the verb+UnCx combination was 52,172.¹¹

After retrieving the sentences from the OPUS2 Turkish Corpus for a semantic analysis, they were manually checked and those sentences that did not have a

⁹ It has a total number of 151,342,424 words. Accessible at sketchengine.eu with a free account. Access date August 14, 2021.

¹⁰ It has a total number of 32,791,491 words. Accessible at sketchengine.eu with a free account. Access date August 14, 2021.

¹¹ Frequency data from TrWaC is available on demand.

verb or the suffix were discarded. Running the query, the engine retrieved 497.398 hits. Then, to compile the actual verb+UnCx combination, verbs that occurred at the end of a sentence or before a comma were included in the analysis. This is because the OPUS2 Turkish Corpus also brought up results where the verb+Imİş combination was not a verb, but a different part-of-speech tag, i.e., adjective *yamulmuş tava*. In the end, there were a total of 175.565 verb+UnCx combinations. Using the random sampler, two sets of 100 sentences were gathered. Frequency of items within the verbal slot of the construction was not lemma based, i.e., *etmiş, edilmiş, edilmemiş* → *et*, because the corpus did not have the function available. The randomized sample only included hits from the subtitle and newspapers subcorpora, because the documentation subcorpus did not bring up any results. Out of 200 sentences, 30 were discarded because they were duplicates.

To form a network of the UnCx, the study follows compiling a pre-lexicographic database (Atkins and Rundell 100–101). Then, it employs Corpus Pattern Analysis (Hanks 404) alongside all the association measures mentioned above to identify the collo-profile (Herbst, “Constructions, generalizations, and the unpredictability of language” 81), or a frequency profile, of the construction. Finally, the network is compared against Plungian’s (353), Aksu-Koç et al’s (14-16), and Aksu-Koç’s (17-18) classification of evidentials to ensure reliability. The presentation of the analysis is twofold: a) a semantic analysis of the construction, and b) statistical preferences of the construction.

Analysis

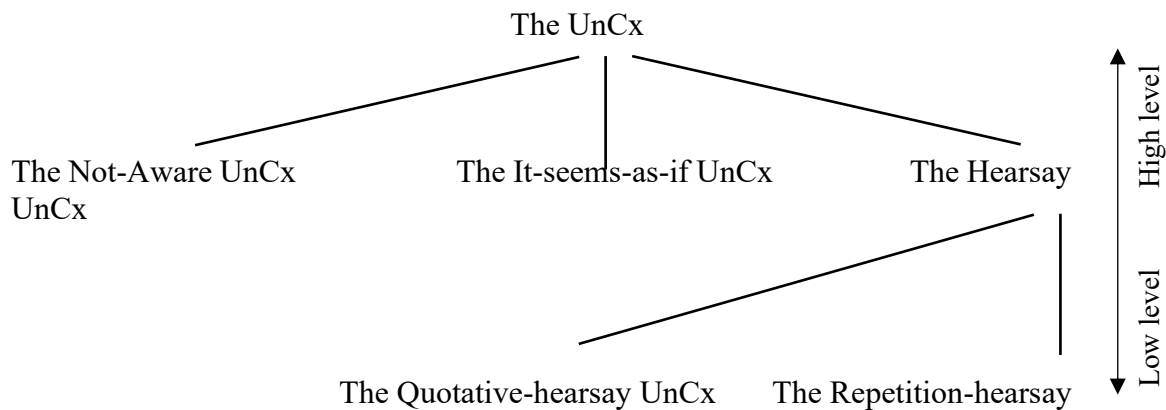
The network in figure 5 shows the relationship between different levels of abstraction of the UnCx in Turkish that emerged from the corpus analysis. Comparing this taxonomical structure to what previous evidentiality studies suggest with regard to semantic analysis (Aksu-Koç et al. 14-16; Aksu-Koç 17-18, Plungian 353), it appears to be comprehensive as it demonstrates a division between indirect and direct evidence, i.e., evidence collected from others in comparison to evidence collected by and within one’s body,¹² and some of the functions that arose here (Plungian 353). This also becomes clear throughout concordance lines with linguistic items that indicate how the evidence was collected. Across these three lower-level constructions, the not-aware-UnCx has the most frequent usage (80 hits), followed by the it-seems-as-if-UnCx (50 hits), and finally the hearsay-UnCx (40 hits). Turning back to Goldberg’s inheritance links,¹³ the lower-level constructions, e.g., the quotative-hearsay-UnCx, inherit the general properties of not-witnessing-an-event. In other words, these constructions are essentially different levels of abstraction of a highly abstract

¹² While it did not occur in the random sample, within-one’s-body refers to sentences such as *acıkmışım* or *uyumuşum*, where the speaker realizes the happening later during speech. In this paper’s account, such uses are categorized under the “not aware” usage.

¹³ Inheritance links (Goldberg, *A Construction Grammar Approach to Argument Structure Constructions* 73-81) are a way of connecting a more highly abstracted construction to a lower one, such that the lower one inherits some of the structural or semantic properties of the highly abstracted construction.

schema, and they all share the same property of ‘unevidentiality’. The higher levels of the structure indicate a higher degree of abstraction while lower levels indicate item-specificity.

Figure 5, the taxonomic family of the UnCx



Direct Evidence

The “not aware” Usage

Speakers in this usage realize a situation after the event took place. In other words, at the time of uttering the sentence, the event had already taken place. Out of 200 sentences, this usage had a total of 80 sentences. Some of the examples taken from the OPUS2 Turkish Corpus are as follows:

(3) *Kim-in koca-sı? -Jenny'-nin... -O evlen-miş.*

Who-GEN spouse-ACC? -Jenny-GEN -She marry-3PERSONEVID

‘Whose spouse is it? -Jenny’s... -She got married (I did not know that)’

(4) *Serçe parmağ-in-da iltihap farket-miş ama önemse-me-miş.*

Pinky finger-GEN-LOC infection realize-3PERSONEVID but care-NEG-3PERSONEVID

‘She/he realized that her/his pinky finger got infected but she/he did not care’

(5) *Bak! Ne çok kar yağ-mış.*

Look-IMP! What much snow snow-3PERSONEVID

‘Look! It has snowed a lot’

(6) *... kilise-de-yken büyük darbe al-mış,*

Pikul. Kablo-lar-ın-dan biri-si sökül-müş.

...church-LOC-ADV big impact receive-3PERSONEVID, Pikul. Cable-PL-GEN-ABL one-ACC rip-3PERSONEVID.

‘... it got damaged in the church, Pikul. One of the cables were ripped open (and I was not aware until now)’

This usage is accompanied by items that signify a later-realization in the concordance lines, i.e., *fark etmiş, bak, Jenny’nin*. Restricting the analysis of this

usage from a grammatical point-of-view only would not help with the semantics or the function of it. By using this construction in combination with other constructions, speakers point at and redirect the focus onto what is now newly introduced to the discussion. This usage can sometimes semantically merge with the "it-seems-as-if" usage if the concordance lines lack enough information. However, this usage has retrospectivity as a feature. In other words, it has more emphasis on a retrospective analysis of an event than the "it-seems-as-if" usage.

The "it-seems-as-if" Usage

Speakers collect evidence on events that took place before the moment of utterance. Using various linguistic or external cues, speakers then arrive at a conclusion by reasoning. This concluding or reasoning is the vague semantic line that separates it from the previous usage, i.e., the "not-aware" usage. This usage had a total of 50 sentences.

(7) *...çünkü basit-çe yanlış tuş-a bas-ıyor-muş... dinleme tuş-u-na bas-mak yerine.*

...because simple-ADV wrong button-DAT press-PROG-3PERSONEVID...
listening button-ACC-DAT press-INF instead.

'... because s/he simply kept pressing the wrong button... instead of pressing the tune-in button (reasoning)'

(8) *Baş-ta-ki yara-ya bak-ılır-sa kısa mesafe-den vur-ul-muş.*

Head-LOC-ACC wound-DAT look-PASS-COND short distance-ABL shoot-PASS-3PERSONEVID

'Inspecting the wound on the scalp, s/he must have been shot from a short distance'

(9) *Bir hata ol-malı. Daha emir-ler ulaş-ma-mış ol-abilir.*

One error be-AUX. Yet order-PL arrive-NEG-3PERSONEVID be-AUX

'There must be a mistake. The orders may not have made it yet'

(10) *Maymuncuk kulan-ıl-ma-mış. Kurban kapı-yı açık bırak-mış ol-malı.*

Picklock use-PASS-NEG-3PERSONEVID. Victim door-ACC open leave-3PERSONEVID be-AUX

'The picklock was not used. The victim must have left the door open'

These sentences have linguistic cues that serve as a reason that led speakers to conclude certain ideas. For instance, in (7), the fact that the other person did not press the correct button serves as linguistic evidence for the speaker to conclude that the other interlocutor was pressing another button. In (8), the wound is the evidence. Similarly, in (9), the intuition that there is a mistake is linguistic evidence. And finally, in (10), the fact that the picklock was never used helps speakers arrive at a conclusion. This lower-level abstraction of the UnCx has linguistic items that serve as or direct the speaker to a piece of evidence, e.g., the use of the UnCx in the preceding sentence.

*Indirect Evidence**The "hearsay" Usage*

The semantics of this usage is straightforward. Speakers collect evidence from some other source, usually an agent, and report on it. This usage was the least frequent usage with only 40 sentences.

(11) *-Bu hiç adil değil. -Bu hiç adil değil-miş.*

-This any fair not. -This any fair not-3PERSONEVID.

'-This is not fair. -(S/he says) this is not fair'

(12) *-Hayır ama Philip ısır-ıl-dı. -Isır-ıl-mış - mı?*

-No but Philip bite-PASS-PAST. -Bite-PASS-3PERSONEVID-Q

'-No but Philip was bitten. -(I just heard it from you) he was bitten?'

(13) *"...Calut'-la ben döv-üş-ür-üm" de-miş.*

"...Calut-INS I fight-RECP-PRES-1PERSON" say-3PERSONEVID

'S/he says "I will fight Calut"'

(14) *Bu bilgi için çavuş-a 500 dolar ver-di-m... Hana tam bir işkolik. Mastır yap-mış, donanma-ya gir-miş.*

This information for officer-DAT 500 dollars give-PAST-1PERSON... Hana total one workaholic. Master do-3PERSONEVID, navy-DAT enter-3PERSONEVID.

'I bribed the officer 500 dollars for this information... Hana is a workaholic. (I heard it from the officer that) she did her masters and joined the navy'

These sentences demonstrate either immediate hearsay evidence, as in (11) or (12) or hearsay evidence which was collected some unknown time ago. (13) is a common example for how the verb *de-* (to say) co-occurs with *-(I)mış*. (14), on the other hand, is a rare example of how the hearsay-UnCx can also be realized without *de-*, other reporting verbs, i.e., *anlat-* (to tell), *söyle-* (to say) and so on, or repetition. Out of those 40 sentences, 28 of them had reporting verbs and a quotation. The rest employed repetition. In (14), the speaker provides the fact that they bribed someone for intel on someone else, which is hearsay information in the end. Furthermore, a striking characteristic of the hearsay-UnCx is that in all the corpus examples, it is either in combination with *de-* or is the repetition of the entire preceding sentence with *-(I)mış*. As such, it is possible to propose two lower-level constructions under the hearsay-UnCx, namely the quotative-hearsay, and the repetition UnCxs. This is because there is no derivation in CxG and each experienced form is a construction on its own in a network.

Association Measures

The figure below demonstrates the top ten lemmatized verbs¹⁴ that co-occur with the UnCx in TrWaC. Using faith, ΔP, and IT∈CX1, it is possible to get a

¹⁴ Due to space issues, only the top ten verbs were included for presentation.

glimpse of the overall statistical preference of the UnCx for verbs. For faith and ΔP , Gries (Coll.analysis 3.5) was used for the automatic statistical analysis. IT \in CX1 was manually calculated following the guidelines previously mentioned. For ease of reference, the scores were converted into percentages. The data for association measures of the top ten verbs in figure 6 were gathered from TrWaC. The total frequency count for the verbs listed below was 49,678, which occurred with the construction at the end of a sentence or before a comma. The total frequency for other constructions except the UnCx at the end of a sentence or before a comma was 4,261,653. There was also a strong correlation between the measures¹⁵ ($r = .80$).

Figure 6, association scores

Verbs (lemmatized)	IT \in CX1 Value (rank)	Faith Score (rank)	ΔP [verb to construction] (rank)	ΔP [construction to verb] (rank)
<i>Et- (cause)</i>	91.48% (1)	2.07% (8)	0.92% (8)	2.40% (1)
<i>Ol- (be)</i>	91.25% (2)	1.59% (9)	0.46% (9)	1.46% (5)
<i>Yap- (do)</i>	61.56% (3)	3.24% (3)	2.01% (3)	2.38% (2)
<i>Al- (take)</i>	44.23% (4)	2.86% (5)	1.64% (5)	1.58% (4)
<i>Ver- (give)</i>	42.68% (5)	2.50% (6)	1.30% (6)	1.37% (7)
<i>Kal- (stay)</i>	35.88% (6)	4.62% (1)	3.27% (1)	1.62% (3)
<i>Çık- (leave)</i>	35.78% (7)	3.26% (2)	2.02% (2)	1.39% (6)
<i>De- (say)</i>	33.89% (8)	1.55% (10)	0.38% (10)	0.50% (10)
<i>Başla- (start)</i>	33.09% (9)	3.08% (4)	1.85% (4)	1.24% (8)
<i>Gel- (come)</i>	32.45% (10)	2.47% (7)	1.27% (7)	1.04% (9)

Starting with IT \in CX1, it is based on raw frequencies and indicates the importance or simple entrenchment of an item in the construction. Interpreting these results, the top ten verbs have a descending importance for the UnCx. The striking result is that in TrWaC, *et-* and *ol-* appear to be the two most significant and competing items-in-construction. This, however, is not surprising according to the top 50 frequent verbs list in Turkish¹⁶ gathered by the TNC team (TNC). These are followed by the rest of the verbs in the figure. Furthermore, faith and ΔP verb to construction show that these verbs, *et-* & *ol-*, are not that faithful to this construction, as they are used more frequently in other constructions, i.e.,

¹⁵ Identified using SPSSv26.

¹⁶ http://www.tudfrekans.org.tr/dosyalar/first_50_verbs_w.pdf.

competition. However, from the UnCx's perspective, it is possible to claim that *et-* and *ol-* have become entrenched across speakers' mental construction because they occur very frequently. This is also partially further supported by ΔP construction to word scores, which shows a similar descending rank for the attraction of the construction to the verbs as in IT \in CX1. The frequency profile can be visualized as in the collo-profile (figure 7). Typeface indicates the frequency and consequently the importance of the item. It is possible to claim that *et-*, *ol-*, and *yap-* point at the superordinate construction's prototypical meaning, namely that speakers report on directly or indirectly collected evidence (see Herbst "Constructions, generalizations, and the unpredictability of language" 70).

Figure 7, collo-profile of the UnCx

The verb-root cx	The UnCx-suffix cx
et- ol- yap- al- ver- kal- çık- de- başla- gel-	-(I)mış

Turning our attention to faith scores, faith calculates the probability of a verb occurring with a construction. For instance, *kal-* has a much higher probability of occurring in the UnCx than *ver-*. In other words, *kal-* as a verb may not be as frequent as *ver-*, but when it occurs, it has a bigger tendency to appear with *-(I)mış* than *ver-*. In the same vein, the least faithful verb to the construction is *de-*. This quantitative result can also be traced back to the quantitative analysis, i.e., the semantic analysis, of this paper where the least frequent lower-level construction was the hearsay-UnCx. Another possible assumption is that *de-* may be used more frequently with the past-tense construction, i.e., *-DI*, than the UnCx. It, however, requires further research as it is not possible to uncover it with the present data set.

ΔP verb to cx results tell a similar story. Being a variant of faith, the result here is a calculation of how likely it is to see *-(I)mış* when triggered by a verb minus the probability of *-(I)mış* occurring without the triggering verb. If one of the scores resulted in a negative ΔP value, it would have indicated that the probability of the verb triggering *-(I)mış* would be low. In this vein, *kal-* ranks first with a 3.27% chance of occurring with *-(I)mış*, e.g., *kalmışlar*, when compared to *kal-* appearing without *-(I)mış*, e.g., *kaldılar*. Another way of putting it is whenever speakers encounter *kal-*, it has a higher chance of occurring with the UnCx than with other constructions. Faith and ΔP scores show that their rankings of the verbs are the most faithful to the construction. Even then, the highest percentage of a conditional probability for the *verb+(I)mış* schema is 4.62%, which is quite low. The verbs with a higher rank in faith and ΔP columns are more likely to be entrenched with the UnCx in speakers' minds while lower-ranking verbs might be in competition with other constructions, i.e., *de-*. These scores can prove useful in the formation of grammaticality judgment tests or translation tasks in usage-based studies as they provide a vague insight into speakers' mental construction representation of this construction. ΔP cx to

word scores, on the other hand, show that $IT \in CX1$ values have validity, and that the construction is attracted to such verbs, possibly pointing at collocations.

In the end, it is possible to claim that while $IT \in CX1$ values indicate which verbs have more entrenchment or familiarity in the construction which can show collocations, i.e., items-in-constructions, higher values in faith and ΔP suggest higher degrees of collocation-ness which is useful for lexical priming studies and network activation studies (see Cangir 58 for a similar discussion).

Discussion and Implications

Starting with the semantic analysis, unlike many studies that focus only on grammar or semantics of the construction separately, with the current approach, it is possible to merge the two and arrive at a lexicogrammatical continuum and identify items-in-constructions, i.e., collocations. While in its core, the analysis here agrees with the semantic distinctions made in previous cognitively oriented studies (e.g., Aksu-Koç et al. 14-28), it presents statistical biases and item-specific lower-level constructions of the UnCx. As shown in figure 5, the UnCx has both overarching and partially generalizable properties, i.e., compare the UnCx in figure 8 and the quotative-hearsay-UnCx in figure 9. This shows that some items or discourse tendencies tend to occur more frequently in some constructions. For instance, the partial generalization of the quotative-hearsay-UnCx is the reporting verbs and some information in quotation marks or reported linguistic evidence. This partial generalization can be illustrated as in figure 9, where the verb-root construction is a slot that can be filled by various reporting verbs. Moreover, linguistic evidence that is reported is also a part of this construction. As such, it is argued that the quotative-hearsay-UnCx is a lower-level item-specific construction.

Figure 8, the Unevidentiality construction template

The UnCx	
FORM	
<i>Verb</i>	<i>-(I)mIş</i>
The verb-root cx	The UnCx-suffix cx
MEANING: speakers report on directly or indirectly collected evidence. This is a highly abstracted construction.	

Figure 9, the quotative-hearsay-UnCx template

The quotative-hearsay-UnCx		
FORM		
Linguistic elements to be	<i>de-, söyle-, anlat...</i>	<i>-(I)mIş</i>

reported		
The quotation-block cx	The verb-root cx	The UnCx-suffix cx
MEANING: speakers quote a piece of information that they collected indirectly and report on it		

Outlining the collo-profile of the quotative-hearsay-UnCx would illustrate which reporting verbs are the most important for the verb-root construction slot, however, due to space limitations this cannot be done. Following the idea of Tomasello's verb islands (5-18), hypothetically if *de-* is the most frequent verb that occurs in that slot within the quotative-hearsay-UnCx, then speakers will gradually generalize over other reporting verbs that they encounter in ambient language for this specific lower-level construction. In other words, *de-* will act as a training wheel for other reporting verbs to cluster on. However, it is possible to claim it for the superordinate UnCx and suggest that *et-*, *ol-*, *yap-*, and *kal-* probably act as prototypical items for constructional learning. Similarly, Goldberg (*Constructions at Work* 103-128) found out that such islands or clusters help speakers learn the construction faster as they represent the most prototypical meaning for that specific construction. It is safe to argue that *de-*, *söyle-* or *anlat-* represent a prototypical meaning of quoting and act as training wheels for the acquisition of the construction. Over time, processes such as conservatism via entrenchment will avoid semantically incoherent or discourse-ill verbs, i.e., verbs that do not meet the communicative purposes of discourse. As such, as speakers are exposed to this construction in usage-events, they generalize over items and what discourse specific functions this construction has. This is also in line with Aksu-Koç et al's findings (22) because as seen in the corpus data, the hearsay-UnCx does not occur as frequently as the other semantic classifications. Based on frequency effects, it is expected that speakers generalize over the highly frequent uses of the construction first, i.e., *et-*, *ol-*, which will be expanded onto lower frequency uses of it by means of simple entrenchment. In this vein, the statistical biases for the verbal slot would be useful for studies that test acceptability ratings of the construction in native speakers of Turkish (or creating Turkish as a foreign language material, see for instance Cangır 45-66).

Linguists know that a theory of language should be able to account for idiosyncrasies and other phenomena in all languages. In that sense, construction grammar as a theory when combined with usage-based linguistics can well account for phenomena in Turkish in a unified way. What is important in analyzing sentences in agglutinating languages such as Turkish or Finnish is that we take the central idea of construction grammar and apply it to each building block to arrive at bigger blocks. The central idea is that each form-meaning pair is a construction at different levels of abstraction (Goldberg, *Constructions at Work* 69-103). By looking at Turkish from a usage-based construction grammar view, linguists can get a better glimpse of how Turkish constructions are taxonomically organized. This can help with typology studies and whether there

exist cross-linguistic constructions. Finally, such studies can also help with explaining idiosyncratic errors of Turkish speakers of English when they speak English. There is also evidence for this from Turkish in a recent study done by Gedik and Uslu (1), in which they found that statistically biased verbs in the ditransitive construction interfere with the output of advanced Turkish speakers of English, resulting in unconventional sentences, e.g., *I transfer you the money, I explain you the situation*. As such, the descriptive adequacy of a usage-based constructionist approach to Turkish is quite high, especially if one were to consider its special attention to cognitive plausibility with relation to frequency effects and how speakers' emergent grammar might be organized.

It is important to note that this study has several limitations. First, it only gives a statistical overview of the highly abstracted UnCx. A future research study can analyze the statistical biases of lower-level constructions in the UnCx family for applied or cognitive linguistics purposes. Second, corpora in Turkish are somewhat problematic and usually employing two corpora in one study is not desired. However, for reasons explained before, it was preferred. Finally, a bigger sample for the semantic analysis might add more lower-level constructions to the UnCx family.

Conclusion

This study analyzed the unevidentiality in Turkish from a usage-based constructionist approach. While the theoretical grounds of the study are not new and have been used by several other studies to analyze various aspects in Turkish, this study couples the usage-based and constructionist approaches to account for a constructicographic perspective of the unevidentiality construction (UnCx). Such a coupling enabled for not only a unified account of how certain lexemes are more attracted to specific instantiations of the UnCx, but also allowed for a more cognitively plausible analysis of the construction with statistical biases for the verbal slot, giving way to collostructions, which can serve as the basis for the preparation of acceptability tests, or an analysis of linguistic productivity, for instance. The findings indicate that the UnCx, *-(I)mış*, has lower-level constructions with various semantic functions which are connected to the highly schematized UnCx via inheritance links. Furthermore, an association-measure analysis reveals verbal preferences of the UnCxs, for example *kal-* has a higher probability of occurring with the UnCx than *çık-*. *Et-*, *ol-*, *yap-*, and *kal-* likely act as prototypical items for constructional acquisition, possibly in children or L2 learners of Turkish, though this requires further longitudinal research. The association measure results confirm naturalistic data obtained from child language acquisition studies on the UnCx, and points in the direction of simple entrenchment, namely that speakers will be more likely to produce highly frequent variants of a construction, only to expand the construction to other lower frequent variants of it. The findings in this study can serve as a reference work for future usage-based constructionist approaches to unevidentiality to detect for instance L1-L2 entrenchment interference levels. We hope that this study inspires more studies analyzing Turkish from a usage-based constructionist perspective.

Acknowledgement

I would like to thank the anonymous reviewers for their detailed feedback on an earlier version of this article. All remaining errors are mine.

Works Cited

- Akkuş, Mehmet. *A Usage-Based Investigation of Converbial Constructions in Heritage Speakers' Turkish Living in the Netherlands*. METU, 2019, <https://open.metu.edu.tr/bitstream/handle/11511/45140/index.pdf>.
- Aksu-Koç, Ayhan, et al. "Evidentials and Source Knowledge in Turkish." *New Directions for Child and Adolescent Development*, vol. 2009, no. 125, June 2009, pp. 13–28. <https://doi.org/10.1002/cd.247>.
- Aksu-Koç, Ayhan, et al. "Some Aspects of the Acquisition of Evidentials in Turkish." *Empirical Approaches to Language Typology*, 2000, pp. 15–28.
- Altınkamaş, Feyza, and Aslı Altan. "Türkçede İlgi Tümceciklerinin Edinimine Kullanım Temelli Bir Yaklaşım." *Dilbilim Araştırmaları Dergisi*, vol. 0, no. 1, July 2016, pp. 69–91. <https://doi.org/10.18492/dad.47469>.
- Arslan, Seckin, and Roelien Bastiaanse. "Tense/Aspect and Evidentiality in Narrative Speech of Turkish-Dutch Bilingual Speakers." *Proceedings of the 17th International Conference on Turkish Linguistics*, University of Rouen France, 2014.
- Atkins, B. T. Sue, and Michael Rundell. *The Oxford Guide to Practical Lexicography*. Oxford University Press, 2008.
- Backus, Ad, and Derya Demirçay. "Intense Turkish-Dutch Bilingualism Leads to Intense Turkish-Dutch Mixing: A Usage-Based Account of Increasing Integration of Two Typologically Different Languages." *Belgian Journal of Linguistics*, vol. 35, no. 1, 2021, pp. 13–33.
- Banguoğlu, Tahsin. *Türkçenin grameri*. 10. bsk, Türk Dil Kurumu, 2015.
- Boyd, Jeremy, and Adele E. Goldberg. "Learning What Not to Say: The Role of Statistical Preemption and Categorization in a-Adjective Production." *Language*, vol. 87, no. 1, 2011, pp. 55–83.
- Bybee, Joan. *Language, Usage and Cognition*. Cambridge University Press, 2010, <https://doi.org/10.1017/CBO9780511750526>.
- Bybee, Joan L. *Morphology A Study of the Relation between Meaning and Form*. John Benjamins Publishing Company, 1985, <https://doi.org/10.1075/tsl.9>.
- Cangır, Hakan. "Lexical Associations in the L1 Turkish Mental Lexicon: Can L1 Lexical Intuition and a Representative Corpus Guide Teaching of Turkish as a Foreign Language Materials?" *Dilbilim Dergisi / The Journal of Linguistics*, vol. 0, no. 37, Dec. 2021, pp. 45–66. <https://doi.org/10.26650/jol.2021.972166>.
- Chomsky, Noam. *The Minimalist Program*. MIT Press, 1995.
- Cinque, Guglielmo. "A Note on Mood, Modality, Tense and Aspect Affixes in Turkish." *The Verb in Turkish*, edited by Eser Erguvanlı Taylan, J. Benjamins Pub. Co, 2001, pp. 47–60.

- Dąbrowska, Ewa. "What Exactly Is Universal Grammar, and Has Anyone Seen It?" *Frontiers in Psychology*, vol. 6, June 2015. <https://doi.org/10.3389/fpsyg.2015.00852>.
- Durrant, Philip. "Formulaicity in an Agglutinating Language: The Case of Turkish." *Corpus Linguistics and Linguistic Theory*, vol. 9, no. 1, 2013, pp. 1–38.
- Fried, Mirjam, and Jan-Ola Östman. "Construction Grammar: A Thumbnail Sketch." *Constructional Approaches to Language*, edited by Mirjam Fried and Jan-Ola Östman, vol. 2, John Benjamins Publishing Company, 2004, pp. 11–86. <https://doi.org/10.1075/cal.2.02fri>.
- Gedik, Tan Arda, and Fatıma Uslu. "L1-L2 Transfer in the Ditransitive Construction: A Usage-Based Replication Study with Turkish Speakers of English." *Advances in Second/Foreign Language Acquisition*, 2022, https://asefola.weebly.com/uploads/1/4/0/2/140216775/book_of_abstracts_-_asefola_conference.pdf.
- Gencan, Tahir Nejat. *Dilbilgisi*. Ağaç, 2007.
- Goldberg, Adele E. *Constructions: A Construction Grammar Approach to Argument Structure*. University of California, 1995.
- Goldberg, Adele E. *Constructions at Work: The Nature of Generalization in Language*. Oxford University Press, 2006.
- Goldberg, Adele E. *Explain Me This: Creativity, Competition, and the Partial Productivity of Constructions / Adele E. Goldberg*. Princeton University Press, 2019.
- Goschler, Julia, and Anatol Stefanowitsch. *Generalization and Transfer in L2 Acquisition: The Role of Entrenchment in L1 and L2*. 2019, https://iclc2019.site/wp-content/uploads/abstracts/applied/ICLC-15_paper_202.pdf.
- Gries, Stefan Th. *Coll.Analysis 3.5. A Script for R to Compute Perform Collostructional Analyses*. 2014.
- Gries, Stefan Th., et al. "Converging Evidence: Bringing Together Experimental and Corpus Data on the Association of Verbs and Constructions." *Cognitive Linguistics*, vol. 16, no. 4, 2005, pp. 635–76, <https://doi.org/10.1515/cogl.2005.16.4.635>.
- Gries, Stefan Th., and Anatol Stefanowitsch. "Collostructions: Investigating the Interaction of Words and Constructions." *International Journal of Corpus Linguistics*, vol. 8, no. 2, 2003, pp. 209–43, <https://doi.org/10.1075/ijcl.8.2.03ste>.
- Guasti, Maria Teresa. *Language Acquisition. The Growth of Grammar*. The MIT Press, 2004.
- Hanks, Patrick. *Lexical Analysis: Norms and Exploitations*. The MIT Press, 2013.
- Herbst, Thomas. "Constructions, Generalizations, and the Unpredictability of Language: Moving towards Collostruction." *Constructions and Frames*, vol. 12, no. 1, 2020, pp. 56–95, <https://doi.org/10.1075/cf.00035.her>.
- Herbst, Thomas. "Is Language a Collostruction? A Proposal for Looking at Collocations, Valency, Argument Structure and Other Constructions." *Lexical Collocation Analysis*, edited by Pascual Cantos-Gómez and Moisés Almela-

- Sánchez, Springer International Publishing, 2018, pp. 1–22. https://doi.org/10.1007/978-3-319-92582-0_1.
- Herbst, Thomas. “The Status of Generalizations: Valency and Argument Structure Constructions.” *Zeitschrift Für Anglistik Und Amerikanistik*, vol. 59, no. 4, 2011, pp. 347–68, <https://doi.org/10.1515/zaa-2011-0406>.
- Herbst, Thomas, and Thomas Hoffmann. “Construction Grammar for Students: A Constructionist Approach to Syntactic Analysis (CASA).” *Yearbook of the German Cognitive Linguistics Association*, vol. 6, no. 1, Nov. 2018, pp. 197–218. <https://doi.org/10.1515/gcla-2018-0010>.
- Işık-Taş, Elvan Eda, and Çiğdem Sağın-Şimşek. “A Corpus-Based Study of Evidentials in the Turkish Cypriot Dialect.” *Lingua*, vol. 231, Nov. 2019, p. 102743. <https://doi.org/10.1016/j.lingua.2019.102743>.
- Karayayla, Tuğba. “A Usage-Based Approach to Productive Use of Inflectional Patterns and Level of Lemma Sophistication in Adult Heritage Speakers’ Performance: Convergence on the Immigrant Variety.” *Linguistic Approaches to Bilingualism*, vol. 11, no. 6, Nov. 2021, pp. 753–82. <https://doi.org/10.1075/lab.18019.kar>.
- Kawaguchi, Yuji. “Usage-Based Approach to Linguistic Variation –Evidence from French and Turkish–.” *Spoken Language Corpus and Linguistic Informatics*, edited by Yuji Kawaguchi et al., John Benjamins Publishing, 2006, pp. 247–68.
- Kaya-Soykan, Didem, et al. “The Production and Perception of Turkish Evidentiality Markers by Turkish-German Returnees.” *Applied Linguistics Review*, vol. 0, no. 0, Nov. 2020. <https://doi.org/10.1515/applirev-2020-0042>.
- Kiraz, Meryem. *A Construction Grammar Approach to How Turkish Learners of English Use Auxiliary Verbs in Terms of Tense, Aspect and Voice*. Kent State University, 2010.
- Kyle, Kristopher, and Scott Crossley. “Assessing Syntactic Sophistication in L2 Writing: A Usage-Based Approach.” *Language Testing*, vol. 34, no. 4, Oct. 2017, pp. 513–35. <https://doi.org/10.1177/0265532217712554>.
- Langacker, Ronald W. *Foundations of Cognitive Grammar: Theoretical Prerequisites*. Stanford university press, 1987.
- Lewis, Geoffrey. *Turkish Grammar*. 2nd ed, Oxford University Press, 2000.
- Muharrem, Ergin. *Türk Dil Bilgisi*. Boğaziçi Yayınları, 1985.
- Ordem, Eser. “Acquisition of Zero Relative Clauses in English by Adult Turkish Learners of English.” *Journal of Education and Training Studies*, vol. 5, no. 1, Jan. 2017, pp. 190–95. <https://doi.org/10.11114/jets.v5i1.2056>.
- Papafragou, Anna, et al. “Evidentiality in Language and Cognition.” *Cognition*, vol. 103, no. 2, 2007, pp. 253–99.
- Perek, Florent. *Argument Structure in Usage-Based Construction Grammar*. John Benjamins Publishing Company, 2015, <https://doi.org/10.1075/cal.17>.
- Plungian, Vladimir A. “The Place of Evidentiality within the Universal Grammatical Space.” *Journal of Pragmatics*, vol. 33, no. 3, Mar. 2001, pp. 349–57. [https://doi.org/10.1016/S0378-2166\(00\)00006-0](https://doi.org/10.1016/S0378-2166(00)00006-0).

- Römer, Ute, and Selahattin Yılmaz. “Effects of L2 Usage and L1 Transfer on Turkish Learners’ Production of English Verb-Argument Constructions.” *Vigo International Journal of Applied Linguistics*, no. 16, 2019, pp. 107–34.
- Sketch Engine. *OPUS2 Turkish Corpus*. sketchengine.eu. Accessed 12 Aug. 2021.
- Sketch Engine. *TrWaC*. sketchengine.eu. Accessed 12 Aug. 2021.
- TNC. *Most Frequent 50 VERBS in TNC Written Texts*. http://www.tudfrekans.org.tr/dosyalar/first_50_verbs_w.pdf. Accessed 19 Feb. 2022.
- Tomasello, Michael. *Constructing a Language: A Usage-Based Theory of Language Acquisition*. Harvard University Press, 2003.
- Underhill, Robert. *Turkish Grammar*. MIT Press, 1976.
- Uzundağ, Berna A., et al. “Functions of Evidentials in Turkish Child and Child-Directed Speech in Early Child-Caregiver Interactions.” *Proceedings of the 40th Annual Boston University Conference on Language Development*, 2016, pp. 403–16.
- Yılmaz, Selahattin. *A Constructional Analysis of Written Academic English as a Lingua Franca: The Case of Unedited and Edited Research Writing*. Georgia State University, 2020.
- Yılmaz, Selim. “Viewpoint and Posttheme in Spoken Turkish.” *Spoken Language Corpus and Linguistic Informatics*, edited by Yuji Kawaguchi et al., John Benjamins Publishing, 2006, pp. 269–86.

Appendix

Example contingency table for faith and ΔP

	Construction (cx) -(I)mlş	Not Construction Not -(I)mlş
Verb (et-)	A (verb+cx) 1798	B (verb+other cxs) 168.496
Not Verb (not et-)	C (other verbs+cx) 50.374	D (other verbs+other cxs) 32,622,995