## Article

Jane Klavan*, Alice Blumenthal-Dramé and Ann Veismann

# Predicting native speaker choice: the role of corpus-based frequency metrics in morpho-syntactic alternations

https://doi.org/10.1515/cllt-2023-0110
Received November 18, 2023; accepted December 9, 2024; published online March 14, 2025

**Abstract:** This paper investigates the role of corpus-based frequency metrics in predicting native speaker choice in Estonian, a Finno-Ugric language with rich inflectional morphology. We test whether frequency affects the choice between two constructions which represent the alternants of a morphosyntactic alternation in Estonian – the adessive case construction and the postpositional *peal* construction. Both constructions are used, *inter alia*, to express the spatial relationship of one object located on top of another object, e.g. *Raamat on laual/laua peal* 'The book is on the table'. We further compare which out of several well-established frequency metrics yields the most accurate predictions and whether frequency on its own outperforms the predictiveness of significant semantic and morpho-syntactic variables identified in previous research (e.g. Klavan, Jane. 2012. *Evidence in linguistics: Corpus-linguistic and experimental methods for studying grammatical synonymy* (Dissertationes Linguisticae Universitatis Tartuensis). Tartu: University of Tartu Press, Klavan, Jane. 2020. Pitting corpus-based classification models against each other: A case study for predicting constructional choice in written Estonian. *Corpus Linguistics and Linguistic Theory* 16(2). 363–391). The comparison of several mixed-effects logistic regression models showed that out of seven frequency metrics considered overall, only those related to the postposition *peal* are predictive. Adding previously established semantic and morphosyntatic variables to the winning frequency metrics significantly improved model fit and outperformed any simpler model. We conclude that frequency affects the choice between two constructions in different ways and that it does not explain away other functional predictors.

***Corresponding author: Jane Klavan**, University of Tartu, Tartu, Estonia, E-mail: jane.klavan@ut.ee. https://orcid.org/0000-0002-0114-3674
**Alice Blumenthal-Dramé,** Department of English, Albert-Ludwigs-Universitat Freiburg, Freiburg im Breisgau, Germany
**Ann Veismann,** University of Tartu, Tartu, Estonia

**Keywords:** alternations; frequency; Estonian; forced choice task

# 1 Introduction

Frequency is among the most robust predictors of human performance (Hasher and Zacks 1984). In usage-based linguistics, it has long been acknowledged that items and structures that are more frequent in language exposure and use are mentally processed with greater ease than those that are less frequent. Frequency has typically been gauged in terms of corpus frequencies, and a growing body of experimental research has revealed principled correlations between statistical generalizations over corpus data and subjects' behaviour in different experimental settings and at various levels of language (for an overview, see Blumenthal-Dramé 2016b).

Although numerous studies have provided support for the view that frequency in language use correlates with ease in language processing, a number of questions are still unresolved. First, a range of competing frequency metrics is used in the literature, and it is not clear which one (if any) should be considered as the most cognitively realistic. Second, it remains to be seen whether metrics that have been successfully applied to model lexical and syntactic processing can be extended to other levels of language description, notably morphology. Third, most relevant studies so far have been conducted on English, and the extent to which their findings apply to other, typologically different, languages is not clear. Fourth, although few linguists today seriously deny the existence of frequency effects in language, there is still a debate as to how far these effects go, in particular concerning the question of whether frequency should be considered as a driver of language processing by itself or as a proxy for (or effect of) a bunch of underlying functional predictors (Blumenthal-Dramé 2012: 190).

Our study aims to contribute to filling these gaps. It explores the extent to which frequency metrics that have been successfully applied to the processing of English can be extended to account for processing in Estonian. Moreover, it investigates whether purported frequency effects can be explained away by other predictors that are independently known to modulate morphosyntactic processing in Estonian.

More specifically, this will be done by comparing the predictive power of several well-established frequency metrics to model subjects' forced offline choices between the alternants of a morphosyntactic alternation in Estonian. Although modelling native speakers' preferences for constructional alternants has received a substantial amount of attention in recent usage-based linguistic research (Bresnan and Ford 2010; Bresnan et al. 2007; Cappelle 2006; Divjak et al. 2016; Gries and Stefanowitsch 2004; Klavan and Divjak 2016; Perek 2012; Szmrecsanyi et al. 2016; Van de Velde 2014;

to name just a few studies here), only a limited number of studies have linked speakers' preferences in experimental settings with corpus-based frequency measures. The few studies that have done so have mostly focused on argument structure constructions in English, notably the dative alternation (e.g. Allen et al. 2012; Ambridge et al. 2013; Ambridge et al. 2014; Ambridge et al. 2015; Bridgwater et al. 2019; Goldberg 2016; Jaeger and Snider 2013; Kuperman and Bresnan 2012; Robenalt and Goldberg 2015; for a notable exception, see Divjak 2017).

In our own study of Estonian morpho-syntactic alternations, we focus on the following research questions:

– Is frequency associated with the choice between two (broadly) synonymous constructions which represent the alternants of a morphosyntactic alternation in Estonian?
– If so, which out of several well-established frequency metrics yields the most accurate predictions?
– Does the inclusion of frequency significantly improve model fit?

In the next section (Section 2), we introduce four issues that we perceive as particularly important in relation to the topic of frequency from a usage-based perspective:

(1) Which corpus-derived frequency metrics are cognitively most realistic?
(2) Are different levels of language (e.g. morphology vs. syntax) sensitive to different frequency metrics?
(3) Are different languages sensitive to different frequency metrics? and, finally
(4) What is the epistemological status of frequency?

This will set the stage for our exploration of the role of frequency metrics in accounting for morphosyntactic processing in Estonian. In Section 3, we introduce our study on the alternation between an exterior locative case construction (the adessive construction) and the corresponding postpositional construction (the *peal* construction), which are both used to express the spatial relationship of one object located on top of another object (e.g. *Raamat on laual/laua peal* 'The book is on the table'). In Section 4, we present our data – both the corpus-based frequency metrics employed in the study and the forced choice task that was used as a testing ground for the frequency metrics. The data analysis of the data from the forced choice task (Section 5) is divided into three sub-sections. We first explore the performance of a model that only includes semantic and morpho-syntactic predictors identified in past research (Section 5.1), we then proceed to look at different models with only one frequency metric included in the model formula (Section 5.2). In our third and final data analysis section, Section 5.3, we explore the performance of models with the winning frequency metrics in combination with the previously established predictors. In Section 6 we discuss the findings of our study in the context of previous studies on corpus-based frequency metrics in usage-based linguistics. We also provide an outlook of how the present study can be extended by incorporating a different experimental design and a range of other potentially pertinent frequency metrics.

# 2 Frequency metrics and morpho-syntactic alternations in Estonian

In the following introductory discussion, we approach the status of frequency from four different angles: different frequency metrics, different levels of language, different languages, and the epistemological status of frequency.

## 2.1 Which metrics are the cognitively most realistic?

The issue of which frequency metric is the cognitively most realistic one has received increasing attention in corpus-based cognitive linguistics (Blumenthal-Dramé 2016b; McConnell and Blumenthal-Dramé 2022). In these studies, the benchmark for cognitive realism has been the extent to which a metric predicts behavioural data in language processing experiments. In line with these studies, we see cognitive realism a relative notion, and as an ideal that can only be approximated: A metric cannot be cognitively real in an absolute sense, but it can outperform competing metrics in terms of cognitive realism, and this can be tested by comparing how well competing metrics align with experimental data. Moreover, any winning metric can only be realistic in the rather loose sense of imposing restrictions on the neural and psychological levels at which the human mind can be modelled (for a more in-depth discussion, see Blumenthal-Dramé 2016b).

Indeed, a vast panoply of competing metrics is available in the literature, with different primary uses in mind. For example, metrics inspired by information theory (like entropy reduction or surprisal) have served as the gold standard in psycho- and neurolinguistics (Frank et al. 2015; Hale 2016; Lowder et al. 2018; Shetreet et al. 2016; Smith and Levy 2013), different types of association scores between words (e.g. LogLikelihood, Dice, T-Score) have been widely used in corpus-linguistics (McConnell and Blumenthal-Dramé 2022), and still other researchers have strongly relied on raw frequencies (Arnon and Christiansen 2017; Arnon and Cohen Priva 2013). While all of those metrics have been insightful in different ways, most of them have not been designed to model cognition; their cognitive realism should therefore not be taken for granted. For example, corpus-derived association scores have first and foremost been developed for the purposes of computational linguistics, lexicography, and applied linguistics (e.g. for corpus-based dictionaries, translation resources, or automated language processing systems), and Evert (2009) rightly cautions that they merely capture "a statistical attraction between certain events and must not be confused with psychological association". Brysbaert et al. (2018) point out that word knowledge and word frequency are not directly related; they introduce word prevalence measure instead.

On the other hand, corpus-derived metrics inherently capture collective behavioural patterns; it therefore stands to reason that they must, however indirectly, be related to cognition. It does not come as a surprise that many empirical studies have identified significant correlations between corpus-derived association measures and different aspects of language processing (for review, see Blumenthal-Dramé 2016a). However, as cautioned by several researchers, the fact that collective patterns are describable in terms of some statistical generalisation does not entail that the individuals giving rise to these patterns unconsciously apply or follow this specific generalisation (cf. Frank and Bod 2011; Kronenfeld 2006; Perruchet and Peereman 2004: 116). This clearly calls for studies adjudicating between competing corpus-based metrics with the aim of identifying the cognitively most insightful one.

## 2.2 Different levels, different metrics?

The second open question at the interface of corpus and cognitive research concerns the range of linguistic levels at which established metrics can successfully predict aspects of language processing. Most cognitive linguistic research so far has looked into attraction strengths between two or more content items, or between content items and abstract grammatical constructions (Blumenthal-Dramé 2016a; Gries 2012; Gries et al. 2005). By contrast, much less research has focused on attraction strengths between content and grammatical items (like inflectional suffixes or adpositions). This is conspicuous given the fact that cognitive linguists generally subscribe to the tenet that language is "constructions all the way down" (Goldberg 2006: 18). This widely cited slogan captures the view that constructions at different levels of language description from pragmatics down to morphology are not qualitatively different but vary only in terms of complexity (in terms of syntagmatic length) and abstractness (i.e. the number of different realisations that a constructional schema can subsume). A strict interpretation of this claim suggests that it should also be "the same metrics all the way down". In other words, metrics that have proven to be predictive in collocational or collostructional research should also work at the morphological level or between content and function words. Conversely, if it was found that the same metrics are not equally insightful across different levels of language, this would arguably weaken the cognitive plausibility of the constructionist uniformity claim. Also note that the it's-the-same-metrics-all-the-way-down hypothesis that we are putting forward is less trivial, or more contentious, than might seem at first glance. Recent research suggests that humans do not possess one unified statistical learning and processing ability; their statistical sensitivity rather varies as a function of modality (e.g. auditory vs. visual signals) and information domain (e.g. musical vs. linguistic acoustic sequences) (Conway and Christiansen 2009; Frost et al. 2015; Siegelman 2020; Siegelman et al. 2017; Thiessen 2017). Against the light of these findings, it makes sense to ask whether the same statistics apply at least across different levels of (one) language.

## 2.3 Different languages, different metrics?

Related to the question of whether different levels of language are affected by different frequency metrics is the third question of whether the same corpus metrics are equally cognitively realistic across different languages. Over the last years, an increasing number of studies have suggested that language users' sensitivity to statistics might be modulated by typological features of their language and its spelling system (Potter et al. 2017; Saksida et al. 2017). This could mean that statistical metrics are computed over different units, that different statistical metrics are used, or that different thresholds are set depending on the language under consideration. This highlights that it would be fruitful to explore whether metrics that are well-established for quantifying attraction strength in English can and should be extended to typologically different languages without further qualification.

## 2.4 The epistemological status of frequency

The fourth and final frequency-related question that this research seeks to address is the epistemological status of frequency-based cognitive explanations. More specifically, to what extent does frequency in itself have a driving influence on language processing, and to what extent should it rather be considered as a spurious by-product of other underlying variables? This question is notoriously difficult to answer, but given that the usage-based paradigm understands itself as an offshoot of American functionalism (Bybee 2010: 195) that is officially committed to explaining "aspects of grammar through reference to meaning and discourse function" (Bybee 2010: 10), the idea that non-frequency related predictors might be the actual driving forces of language processing should not be discarded too quickly (Baayen 2010;

Bley-Vroman 2002; Caldwell-Harris 2021; Jolsvai et al. 2020; Newmeyer 2003). Bybee (2007) has stressed that the complex question of whether frequency is a cause or an effect can be answered in both ways. As an observable pattern in text, it is an effect, at the same time "repetition of experiences has an impact on cognitive representations and in this way becomes a cause" (Bybee 2007: 17–18).

One way of achieving a more nuanced understanding of the relation between frequency and other, well-established functional predictors is to gauge their respective roles when combined into the same statistical model. Do they account for the same share of variance (in which either frequency or the other predictors could rightly be deemed epiphenomenal), do they show independent effects (in which case frequency could be assigned the role of one functional predictor among others), or does frequency supersede the other predictors (in which case the share of variance explained by the other predictors would be a proper subset of the share of variance accounted for by frequency, which would point to frequency being the actual driving force)? We will investigate these questions by taking a morphosyntactic alternation in Estonian as our case study.

## 3 Estonian as a morphologically rich language

Our case study comes from a morphologically rich Finno-Ugric language that exhibits a high degree of morphological homonymy and polysemy. By using data from a non-Indo-European language and by focusing on morphological rather than syntactic alternations (cf. the dative alternation and the particle placement alternations in English), we can offer more nuanced data for the discussion of how frequency affects language. Our focus in this paper is the alternation between a case construction and a postpositional construction to express the spatial relationship of one object located on top of another object. For example, if we have a relationship between a book being located on top of a table (e.g. *the book is on the table*), speakers of Estonian can choose between two alternants of a morpho-syntactic alternation: a morphological construction involving the adessive case (case construction; example 1), or a syntactic construction involving the postposition '*peal*' (postpositional construction; example 2).

(1) Raamat on laual.
book.SG.NOM be-PRS.3SG table.SG.ADE
'The book is on the table.'

(2) Raamat on laua peal.
book.SG.NOM be-PRS.3SG table.SG.GEN on
'The book is on the table.'

Estonian is one of the few languages to have a system to mark the same function with a postposition and case affix in parallel manner. It has been claimed in Estonian reference grammars that the meaning of adpositions is more concrete and specific than that of the cases, while the usage range of the cases is much broader (Erelt et al. 1995: 33–34; Erelt et al. 2007: 191). This is in line with the general claims made in the literature concerning the differences between adpositions and case affixes – the case construction is more abstract and expresses more frequent relations than the postpositional construction (Comrie 1986; Hagège 2010; Lestrade 2010). In other Finno-Ugric languages, Bartens (1978) and Ojutkangas (2008) have found that the analytic adpositional construction, compared to the synthetic case construction, places more stress on location and is used together with smaller, manipulable things as Landmarks.

There are not very many empirical studies conducted on the alternation between case affixes and postpositions in Estonian. One prominent exception is the alternation between the adessive case construction and the postpositional construction with *peal* described above. Previous corpus-based and experimental studies (Klavan 2012, 2020, 2021; Klavan and Divjak 2016; Klavan and Schützler 2023; Klavan and Veismann 2017; Klavan et al. 2015) have looked at different semantic and morphosyntactic factors to predict the choice between the two competing constructions. Across the different studies, confirmatory evidence has been found for the importance of the following variables: word lemma, length, complexity, mobility, and the syntactic function of the Landmark phrase, and the word class of the Trajector. These variables are discussed in more detail in Section 4.2 of the paper. Shorter phrases and smaller entities prefer the postpositional construction, while longer and more complex Landmark phrases prefer the case construction.

Klavan and Veismann (2017) compared mixed effects logistic regression with two experimental tasks (forced choice and rating). They point out that the four predictors singled out by the mixed effects model (length, complexity, mobility and word class) correlate with both forced choice and rating tasks. Based on the experiment they conclude that the default choice for Estonian native speakers is the adessive construction. The choice for the adessive construction was proportionally higher than for the *peal* construction in the forced choice task and the mean residualised rating for the adessive construction was much higher than the mean residualised rating for the peal construction (Klavan and Veismann 2017). A finding from Klavan (2020), where the author compared logistic regression model and naïve discriminative learning (NDL) on the same dataset, shows that both models predict the postpositional construction better than the adessive construction. We will turn back to the issues of the adessive as the default choice for speakers and the finding that the models behave “better” for the postpositional construction in our discussion where we place the previous results and the results from the present study in the context of the inverse frequency effect.

In light of the previous studies on the alternation between the adessive and the *peal* construction, there are a number of issues that the studies do not address in sufficient detail. Both Klavan (2020) and Klavan and Veismann (2017) acknowledge the importance of frequency, but they do not go beyond the simple observation that the adessive is the preferred construction. We propose to zoom in on the issue of frequency by looking at different frequency measures and how they contribute to the choice between the two competing constructions both on their own and in combination with the semantic and morpho-syntactic variables that have been found to play a role in this alternation in the previous studies. We will now introduce the dataset we are working on and explain both the dependent and independent variables of our study.

# 4 Data and method

Our focus in this study is on the question whether frequency affects the choice between two competing constructions and, if so, which frequency metric yields the most accurate predictions. Furthermore, we want to explore whether frequency outperforms the predictiveness of other relevant variables identified in previous research, namely the complexity, length, mobility and the syntactic function of the Landmark phrase and the word class of the Trajector. In order to answer these questions, we will report the results of a forced choice task carried out with 103 native speakers of Estonian where the participants had to choose between the two competing constructions. We will then model the results using the binary choice between the case construction and the postpositional construction as the dependent variable and the various frequency metrics and the linguistic factors as independent variables. First, the materials and procedure of the forced choice task is discussed, followed by the description of the frequency metrics and the linguistic variables.

## 4.1 Forced choice task

### 4.1.1 Materials

The experiment consisted of 60 corpus sentences with a blank for the original construction followed by the two constructional alternatives. There were 30 experimental items and 30 filler items in the experiment. The list of filler items were alternations of a different type, e.g. pro-drop alternation *oletan* ~ *ma oletan* "I assume". The experimental items were randomly sampled from five probability bins defined by the logistic regression model fitted by Klavan (2012: 176–181). The sentences therefore represent the full probability scale – the stimuli ranged from sentences where one construction was very probable (near-categorical preferences) to sentences where both constructions were equally probable (approximately equal probability estimates for both choices). Six sentences were randomly sampled from each probability bin, adding up to a total of 30 (5 × 6) experimental items.

We decided to sample items across the corpus model predicted probabilities rather than the actual frequency measures used in our study because we did not know which frequency metric would turn out to be the winning metric. We did not want to prejudge the analysis by sampling our data according to only one metric which might have reduced variability according to the other metrics. This study is part of a larger scope research which was not only focused on frequency metrics. Using Spearman's correlation coefficients, we looked at the relationship between the different frequency metrics and the corpus probabilities among the experimental stimuli (see Appendix A for an overview of the results). These findings indicate that frequency and probability are associated.

Figure 1 presents the item probabilities estimated by the original logistic regression corpus model. For each sentence an alternative paraphrase was

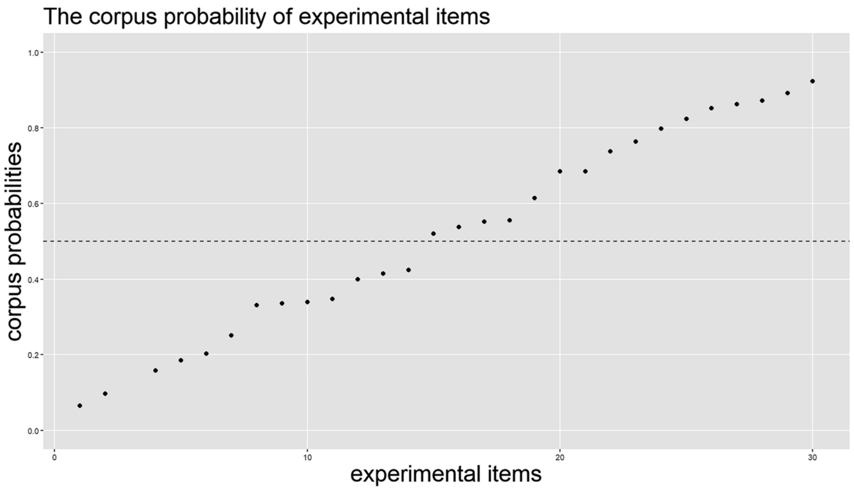



**Figure 1:** Estimated probabilities for experimental items based on the binary regression model fitted in Klavan (2012).

constructed for the original construction and both alternatives were presented together with the original sentence context. Items were pseudo-randomized so that no two sentences from the same probability bin followed each other. There were four versions of the questionnaire to diminish potential order effects. The order of the construction choices was alternated between the versions. For example, in versions 1 and 3 the *peal* construction was given first, in versions 2 and 4 the adessive construction was presented first for the same item.

#### 4.1.2 Participants

103 native speakers of Estonian were recruited via the online language technology platform qlaara (developed by TripleDev together with Arvi Tavast). They were randomly assigned to one of the four versions of the experiment (v1 = 23, v2 = 27, v3 = 29, v4 = 24). The participants (85 female, 18 male) ranged in age from 18 to 47 (mean 26, SD = 7.2).

#### 4.1.3 Procedure

Participants were asked to choose which of the two constructions suits into the blank better. They were asked not to spend too much time on making the decision and to try and guess which of the two alternatives is the more natural alternative for Estonian native speakers in written language. The two constructional alternatives were presented next to each other horizontally after the corpus sentence with a blank. Participants saw only one sentence at a time and were not able to go back and change their answers. Each subject completed the task with the same 60 sentences. On average, it took about 10 min for the participants to complete the experiment.

### 4.2 Annotation: semantic and morphosyntactic variables

The 30 experimental items were annotated for 5 variables based on the results of the previous studies on the adessive and *peal* alternation in Estonian (Klavan 2012; Klavan 2020; Klavan and Veismann 2017; Klavan et al. 2015): complexity, length, mobility and the syntactic function of the Landmark phrase and the word class of the Trajector. Following is a very brief overview of these variables; for more details on the annotation scheme, see Klavan (2012). Appendix B lists all the experimental items together with the annotated categories.

The variable COMPLEXITY takes the reference value “simple” if the Landmark lemma is not a compound (e.g. *laud* ‘table’) and the value “compound” if it is (e.g. *kirjutuslaud* ‘writing desk’). LENGTH is the number of syllables in the Landmark phrase (log-transformed). The variable MOBILITY has two levels: mobile and static. Following de Vega et al. (2002), mobile Landmarks are those that do not have a fixed position in the environment, either because they move by themselves (e.g. humans, animals) or can be moved by an external agent (e.g. a table, a car). Static Landmarks have a fixed position in the environment (e.g. street, market).

The variable SYNTACTIC FUNCTION denotes the grammatical function of the Landmark phrase: adverbial means that the Landmark is a sentence constituent in its own right, while modifier means that the Landmark modifies the Trajector within the same constituent. For example, both the adessive and the postpositional construction with *peal* can fulfil two syntactic functions in a clause – that of an adverbial, as *Vaas on laua peal/laual* 'The vase is on the table', or an adverbial modifier as *vaas laual/laua peal* 'the vase on the table' in *Vaas laual/laua peal on ilus* 'The vase on the table is pretty". The variable has two levels: 'adverbial' and 'modifier'.

The word class of Trajector (TRWC) has two levels – it can either be a noun or it can belong to other word classes (verb phrases and pronouns). For example, in the sentence *Vaas on laua peal* 'The vase is on the table', the Trajector is *vaas* 'vase' and it is labelled a 'noun'. However, in the sentence *Istusin voodi peal* 'I was sitting on the bed', the Trajector is *istusin* 'I was sitting', a verb phrase and it is labelled as 'other'.

Among the 30 experimental items, there are 4 compound and 26 simple Landmarks, 13 mobile and 17 static Landmarks with a mean length of 2, corresponding to 4 syllables. 27 of the Landmark phrases were used in the adverbial syntactic function and 3 in the modifier function. 15 of the Trajector phrases were nouns and 15 from other word classes. For the subsequent modelling of the results, it should be noted that in the previous studies the importance of these variables was assessed in a corpus-based dataset where the distribution of the variable levels is very much the same, but the entire number of lexemes included in the studies extends drastically beyond the 30 items included in the present study (cf. Klavan 2021 whose corpus sample covered 438 different lemmas). It is thus expected that these semantic and morpho-syntactic predictors may not yield models with the same predictive power as models built on corpus data provide.

## 4.3 Corpus-based frequency measures

This section gives an overview of the corpus-based frequency metrics used in the present study. All of the frequency metrics of interest were extracted from the Estonian National Corpus (ENC 2017; size 1.1 billion words in total; Kallas and Koppel 2017). ENC is one of the most comprehensive corpora of present-day written Estonian. It consists of the Estonian Reference Corpus (2013), the Estonian Web corpus (2013 and 2017) and Estonian Wikipedia (2017). The corpus was accessed using the Sketch Engine interface which provides access to the morphologically annotated corpus.[1] The automatic morphological annotation for ENC 2017 has been done using the tagging tool Filosoft,[2] which assigns the correct morphological tag to a word in 97 % of cases (Kaalep 1998: 23). We used the special Corpus Query Language (CQL) of Sketch Engine for extracting the necessary frequency metrics. The corpus data was extracted in March 2019.

Table 1 gives a quick overview of the different frequency metrics used in the study; the details of how each metric was calculated are discussed below. Appendix C lists all of the relevant values of the frequency metrics for the experimental items. Appendix D lists all of the relevant measures from collostructional analysis.

In total, we looked at seven frequency measures for the present study: (1) word lemma frequency, (2) the frequency of the particular word in the adessive case, (3) the frequency of the genitive word form in combination with the postposition *peal*, (4) the ratio of the particular word in combination with the adessive construction divided by the word in combination with the *peal* construction, (5) collostructional strength of the word in the adessive construction, (6) collostructional strength of the word in the *peal* construction, and (7) the ratio of the two collostructional measures. Following is a description how we operationalized the frequency measures.

LEMMA. The frequency of the word lemma was calculated by logarithmically transforming the raw frequency count of the lemma extracted from the ENC 2017

**Table 1:** Description of frequency metrics.

| Frequency metric | Description |
|---|---|
| LEMMA | Log transformed frequency of the word lemma |
| ADE_FORM | Log transformed frequency of adessive word form used in experiment (sg. or pl.) |
| PEAL_FORM | Log transformed frequency of genitive word form and peal combination (sg. or pl.) |
| ADEPEAL_RATIO | Log transformed ratio of ADE_FORM/PEAL_FORM |
| ADE_COLL | Collostructional strength of word in ade-cx (scaled and centered) |
| PEAL_COLL | Collostructional strength of word in *peal*-cx (scaled and centered) |
| ADECOLL_PEALCOLL_RATIO | Ratio of ADE_COLL/PEAL_COLL |

**1** https://www.sketchengine.eu/estonian-reference-corpus/.

**2** https://github.com/Filosoft/vabamorf.

corpus. For example, from Appendix C we can see that the raw frequency count of the lemma for the word *aken* "window" is 139,805; for our modelling we used the logarithmically transformed value of 5.15.

ADE_FORM. The frequency of the particular word in the adessive case was extracted from the ENC 2017 corpus by searching for the exact word form used in the experiment. If the word was used in the singular adessive form in the experiment, then we looked up only the frequency of this form in the corpus, e.g. *aknal* (window-ADE.SG) and ignored the plural forms, e.g. *akendel, aknatel* (window-ADE.PL). We used the same procedure when the word was used in the plural adessive form in the experiment, e.g. searching only the plural adessive form *piltidel* (photo-ADE.PL) and ignoring the singular form *pildil* (photo-ADE.SG).

PEAL_FORM. The frequency of the particular word used together with the postposition *peal* was extracted from the ENC 2017 corpus by searching for the exact combination used in the experiment. If the word was used in the experiment in singular genitive form together with *peal*, then we looked up only the frequency of this combination, e.g. *akna peal* (window-GEN.SG on) and ignored the plural forms, e.g. *akende peal, aknate peal* (window-GEN.PL on). We used the same procedure when the word was used in the plural genitive form, e.g. searching only the plural genitive form together for *piltide peal* (photo-GEN.PL on) and ignoring the singular form *pildi peal* (photo-GEN.SG on).

ADE_PEAL_RATIO. It is the log transformed ratio of two frequency counts: ADE_FORM and PEAL_FORM, i.e. a relative frequency of the word in the adessive construction and the word in the *peal* construction. Because the two forms compete, it may be that the absolute frequency of the two constructions is not as important as the frequency of one construction relative to the frequency of the other construction with which it is competing.

ADE_COLL. It is the collostructional strength of a word in the adessive construction.

PEAL_COLL. It is the collostructional strength of a word in the *peal* construction.

ADECOLL_PEALCOLL_RATIO. It is the ratio of the two collostructional values: ADE_COLL and PEAL_COLL, i.e. a relative frequency of the collostructional strength of a word in the adessive construction and the collostructional strength of a word in the *peal* construction. Similarly to the ADE_PEAL_RATIO measure, we decided to calculate the ratio of the two collostructional measures.

The computations for both of the collostructional metrics were done with the R-script "coll.analysis 3.5" developed by Stefan Gries (Gries 2014). The script uses an exact binomial test to quantify the association strength between the words and their use with the adessive and *peal* construction. The script allows us to calculate the association strength on the basis of a crosstabulation of the individual frequencies of the word and the construction in question as well as their joint

**Table 2:** Frequencies of the word *aken* and the adessive construction in ENC-17.

| | ***aken*** **'window'** | **Other words** | **Total** |
|---|---|---|---|
| Adessive construction | 615 | 6,302,504 | 6,303,119 |
| Other constructions | 33,212 | 1,101,248,138 | 1,101,281,350 |
| Total | 33,827 | 1,107,550,642 | 1,107,584,469 |

frequency, using the Fisher exact test. For example, to calculate the association strength (i.e. collostruction strength) between the word *aken* and the adessive, the data in Table 2 are needed.

The Fischer exact $p$-value is calculated for all 30 words occurring with both the adessive and the *peal*-construction. From the script output we used the measure "collostructional strength" which is based on $p$-values of the Fischer exact test. The collostructional strength for the 30 words for both constructions is given in Appendix D. The measure can be interpreted as follows: collostructional strength > 3 = $p < 0.001$, collostructional strength 1.30103 = $p < 0.005$. For statistical modelling we centered and scaled the values.

## 5 Results: statistical modelling of the data

There are three sets of data analyses that we conducted for our study. The aim of the first analysis was to replicate the previous findings and we only included in the model the five predictors that have been found significant in previous studies on the Estonian adessive ~ *peal* alternation: length, mobility, complexity, and the syntactic function of the Landmark phrase and the word class of the Trajector. The second analyses looked at to what extent the different frequency measures calculated for the experimental items can be used to explain the proportion of *peal* choices in the forced choice task. In the third analysis, we combined the winning frequency metrics with the five previous predictors. We will first say a few words about the statistical modelling technique applied in the paper and we then present the results of each of the three analyses separately.

Mixed-effects logistic regression (Harrell 2001; Hosmer et al. 2013; Pinheiro and Bates 2002) was used for the analysis of data in order to find out whether the choice between the case construction and the postpositional construction in the experimental data can be accounted for by the proposed frequency metrics. The data were analysed using the statistical computing software *R* (version 3.6.1, R development core team 2019). For the data analysis in this paper, the *lme4* package was used (Bates 2014; Bates et al. 2015).

In model evaluation, two aspects are considered: reduction in AIC and model accuracy. Model accuracy is evaluated by two measures – percentage of overall accuracy and the $C$ measure (Hosmer et al. 2013: 173–182). Overall accuracy is estimated by cross-tabulating the two possible outcomes by high and low probabilities based on a cutoff point set at 0.5. The model makes a correct prediction if the estimated probability for the *peal* construction is greater than or equal to 0.5 and the *peal* construction was actually chosen by the participants in the forced choice task. The $C$ measure ranges from 0.5 to 1.0 and reflects the ability of the model to discriminate between the two outcomes. The following general guidelines are given as a rule of thumb: $C = 0.5$ – no discrimination; $0.5 < C < 0.7$ – poor discrimination; $0.7 \leq C < 0.8$ – acceptable discrimination; $0.8 \leq C < 0.9$ – excellent discrimination; $C \geq 0.9$ – outstanding discrimination (Hosmer et al. 2013: 177).

We excluded one of the stimuli items from the data analysis – *ala* 'field'. Only 1 participant out of 103 had chosen the *peal* construction for this item. When inspecting the random effect structure, it became clear that it created potential problems for modelling, and it was decided to leave data regarding this item out. The data analysis is based on 29 items for which we have choices by 103 participants (2,987 data points).

## 5.1 Model with the past predictors

To test whether the predictors that have been found to be significant in the previous studies conducted on the Estonian morphosyntactic alternation between the adessive case and the postposition *peal* construction, we built a mixed-effects model using five linguistic features as fixed effects (length, mobility, complexity, syntactic function, Trajector word class) and subject as a random effect. The goodness of fit measures for this model are given in Table 3. According to Hosmer et al. (2013), a $C$ value of 0.75 is acceptable discrimination. We can see that both the accuracy score (70 %) and the $C$ value have increased compared to the intercept only model (accuracy = 67 %, $C$ value = 0.69). Still, the model is not doing a particularly good job in predicting the choice. We may conclude that the model may be missing some

**Table 3:** Model accuracies (overall accuracy and *C* measure) for the model with the past predictors.

| Model | Accuracy | *C* value |
|---|---|---|
| Model 0: intercept only | 67 % | 0.69 |
| Model: length + mobility + complexity + synfun + trwc + (1\|subject) | 70 % | 0.75 |

important predictors and we now proceed to reporting models that take into account one frequency metric at the time. In the last section of our results, we combine the winning frequency metric(s) with the linguistic metrics to see if we are able to build a model with higher predictive power. We need to keep in mind that we are modelling the results of an experiment that has only 29 lexical items for which 103 native speakers of Estonian have provided a forced choice. Overall, the dataset is sufficiently large (2,987 datapoints), but it does not represent the linguistic diversity of lexemes we would find in a corpus-based study.

## 5.2 Models with one frequency metric per model

The aim of the second analysis of the data was to see to what extent the different frequency measures calculated for the experimental items can be used to explain the proportion of *peal* choices in the forced choice task. We decided not to fit all of the frequency metrics to the data in one single model and instead we ran 8 different models, one of them the so-called zero model with intercept only. Figure 2 is a correlation matrix of the seven frequency measures studied.

We used the specific frequency metric as the fixed effect, and subject as random effect. The overall model formula for all the seven models is the following:

CHOICE_CX ~ FREQUENCY METRIC + (1|SUBJECT)

The overall accuracy and the *C* measure for the eight models are given in Table 4. We also include among our models the so-called null model or the intercept only model. In terms of the model fit, the models are not performing at a very good level of accuracy. The accuracy scores range from 67 % (the intercept only model) to 71 % (models that include frequency measures for the word in combination with the *peal* construction). The *C* values tell us that the fit of the models is on the borderline between poor discrimination and acceptable discrimination (range: 0.69–0.76).

As for assessing the importance of specific frequency measures, we take decrease in the Akaike information criterion (AIC; Hosmer et al. 2013: 120) as an indicator of the importance of a particular predictor (Baayen et al. 2013: 264). AIC is used to compare the fit of models with different numbers of parameters – a smaller value is taken as an indication of a better model fit. Individual parameter estimates were tested by the likelihood ratio test, a test based on the difference in deviances. All models except Model 5 (ade_coll) were significantly better than the null model, based on the likelihood ratio tests (DF = 1; see Appendix E for details). The frequency

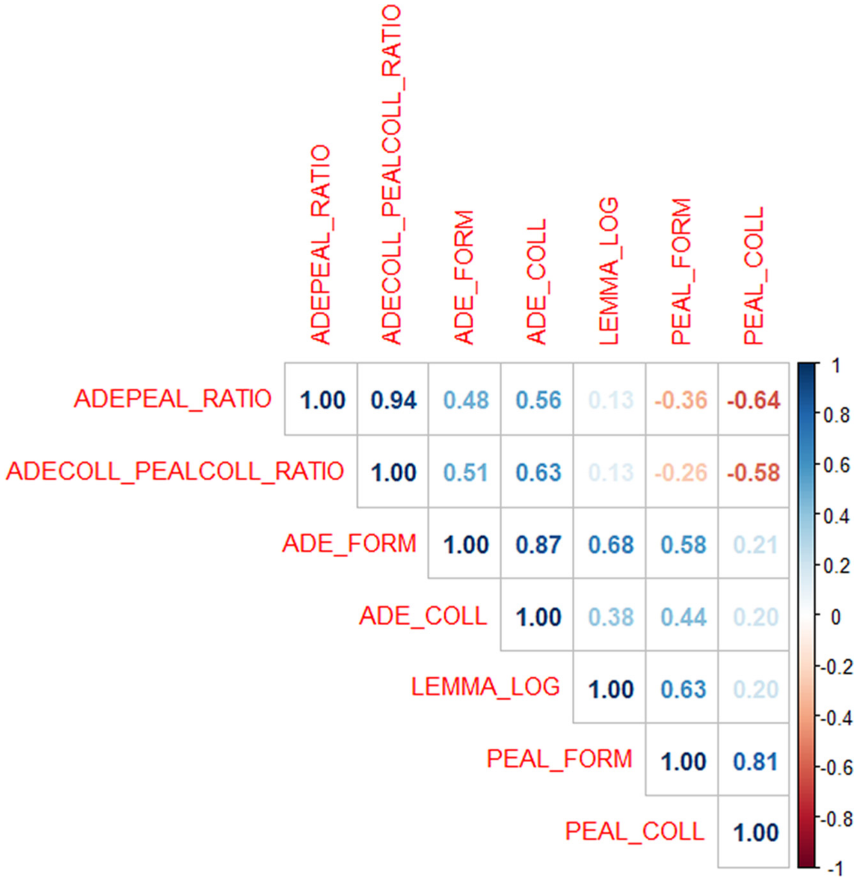



**Figure 2:** Correlation of frequency measures.

**Table 4:** Model accuracies (overall accuracy and *C* measure) for the seven models.

| Model | Accuracy | *C* value | Reduction in AIC |
|---|---|---|---|
| Model 0: intercept only | 67 % | 0.69 | |
| Model 1: ~ ade_form | 68 % | 0.70 | 27.1 |
| Model 2: ~ peal_form | 71 % | 0.74 | 194.4 |
| Model 3: ~ adepeal_ratio | 70 % | 0.73 | 147.3 |
| Model 4: ~ lemma | 68 % | 0.71 | 60.3 |
| Model 5: ~ ade_coll | 68 % | 0.70 | –1.9 |
| Model 6: ~ peal_coll | 71 % | 0.75 | 188.9 |
| Model 7: ~ adecoll_pealcoll_ratio | 70 % | 0.73 | 145 |

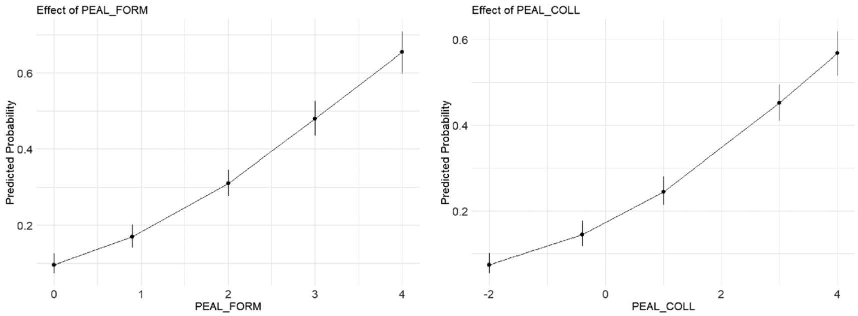



**Figure 3:** Effect plots for Model 2 (peal_form) and Model 6 (peal_coll).

measures were added to the intercept only model (the null model). The last column in Table 4 lists the reduction in AIC – the larger the reduction in AIC compared to the null model, the more important we consider the frequency measure to be. Based on the different model accuracy measures we have selected the following two models as our winning models: Model 2 (peal_form) and Model 6 (peal_coll).

As to the specific predictions made by the models, we may inspect the estimated coefficients of the main effects (Figure 3). We discuss here only the coefficients of our two winning models significant (Model 2 and Model 6; see Appendix E for the coefficients of the other models). The coefficients in Figure 3 indicate that for Model 2 (peal_form) the higher the frequency of the *peal* form with the lemma in the corpus data, the higher probability of *peal* construction being chosen in the forced choice task. This result is mirrored in Model 6 (peal_coll), where a higher collostructional strength of the *peal* construction for a lemma indicates the higher probability of the *peal* construction being chosen in the forced choice task. It seems that the frequency of the lemma in the adessive form (Model 1), the collostructional strength of the lemma in the adessive construction (Model 5), and the overall frequency of the lemma (Model 4) make a less prominent contribution towards explaining the choices made by the participants in the forced choice task. We return to the discussion of why frequency measures related to the *peal* construction play a more prominent role in explaining the variation in the forced choice data in the next section.

## 5.3 Models with frequency in combination with previous predictors

The aim of the third set of statistical analysis is to combine the winning frequency metrics with the five previous predictors from the mixed-effects regression model

**Table 5:** Model accuracies (overall accuracy and *C* measure) for the two winning models in combination with other predictors.

| Model | Accuracy | *C* value | Reduction in AIC |
|---|---|---|---|
| Model 0: length + mobility + complexity + synfun + trwc + (1\|subject) | 70 % | 0.75 | |
| Model 1: peal_form + length + mobility + complexity + synfun + trwc + (1\|subject) | 71 % | 0.76 | 39 |
| Model 2: ~ peal_coll + length + mobility + complexity + synfun + trwc + (1\|subject) | 72 % | 0.77 | 101.2 |

presented in Section 4.1. For the purposes of comparison, Table 5 also includes the model evaluation measures for the model with only the previous five predictors. We can see from the table that both frequency measures add some explanatory power – there is an increase in both accuracy and the *C* value and a drop in AIC when we compare the two models with a frequency measure in combination with the previous predictors to the model that only has the previous predictors (Model 0).

In order to evaluate the contribution of each predictor in each of the models likelihood ratio tests were conducted. It should be kept in mind that our focus was to test whether the frequency measures add explanatory power to the models or not. Hence, all of the predictors were kept in the final models irrespective of whether they significantly improved the model fit or not. The results of the likelihood ratio tests for each model are reported in Appendix F. From these tables it can be seen that the results of our mixed-effects logistic regression models highlight the importance of the frequency metrics in explaining the part of the variability in the choice between the adessive and *peal* construction in our data.

# 6 Discussion

We set out to explore which frequency metrics that have been widely used in research on English can be extended to account for morphosyntactic processing in Estonian. We compared the predictive power of several well-established frequency metrics to model subjects' forced offline choices of the Estonian morphosyntactic alternation between the adessive case and the postposition *peal*. In addition, the purported frequency metrics were pitted against other predictors that are known to modulate the use of this morphosyntactic alternation in Estonian. We will now discuss the research questions raised at the beginning:

## 6.1 Does frequency correlate with alternant choice?

A positive answer can be provided for whether frequency affects the choice between the adessive and *peal* constructions in Estonian. The following two frequency metrics were particularly significant: frequency of the genitive form in combination with the postposition *peal* (peal_form, Model 2) and the collostructional strength of the word in the *peal* construction (peal_coll, Model 6).

By contrast, the frequency of the adessive form (Model 1 and Model 5) does not seem to play a role in explaining the variation in the forced choice data. Neither was the overall frequency of the lemma used with the constructions particularly significant (Model 4). In one of the few previous studies that have reported the effects of lemma and whole-word frequencies in Estonian, Lõo et al. (2018: 90) show that the variable importance of lemma frequency was inferior to that of whole-word frequency as predictors for both production latencies and articulation durations. They conclude that, at least for Estonian, lemma frequency does not afford the predictive precision that comes with whole-word frequency (Lõo et al. 2018: 90). Our study is in line with this finding concerning word lemma frequency, however, we did find word frequency to be predictive with the *peal* construction.

Lõo et al. (2018: 91) claim that the support for a whole-word frequency effect for Estonian case-inflected nouns fits well with the findings that in English, sequences of words also show frequency effects; the same applies to our study. Follow-up research will be needed to explore whether our winning metrics are equally predictive of subjects' offline choices at other levels of language processing (e.g. at the level of fully abstract syntactic constructions).

## 6.2 Which frequency metrics yield the most accurate predictions?

In looking for an answer to our second research question – which out of several frequency metrics yields the most accurate predictions – we retain that frequency effects reflect a complex interaction between the various frequency measures of different grain-sizes. We found two frequency measures to play a particularly significant role in explaining the choices made by native speakers in the forced choice task. Such an assumption about different frequency metrics being pivotal aligns with the exemplar models of language (Bod 1998, 2006; Pierrehumbert 2001, 2006; Snider 2008). What matters in the context of the present study is the frequency of the words used with the *peal* construction and the relative frequency of the word in the adessive construction and the word in the *peal* construction. Our analysis shows that the more frequent the word form is used with the *peal* construction, the more probable the choice of *peal* construction in the forced choice task. The same is echoed by the collostructional measure – the stronger the collostructional strength for a particular word lemma with the *peal* construction, the more probable the choice of the *peal* construction in the forced choice task. For the relative frequency metric we found that if the adessive and *peal* ratio is small, i.e. the smaller the difference in the overall frequency of the word with the adessive and *peal*, the more probable the choice of the *peal* construction in the forced choice task.

Our study shows that out of the different frequency metrics that we tested, the collostructional model performed the best when looking at the frequency metrics individually. A similar finding was observed by Colleman and Bernolet (2012) who offer a detailed comparison of the results of the Dutch dative alternation based on a corpus analysis and a series of picture description experiments. As with the Estonian data, where one of the constructions, the adessive construction, is much more frequent overall compared to the other, postpositional, construction, the Dutch data demonstrates that the double object construction is by far the most frequently realized option in natural language. Yet, the experimental data display a distinct bias toward the prepositional dative. One of the explanatory factors offered by the authors is the alternation biases of the individual dative alternating verbs as measured not simply in terms of the raw observed frequencies, but in collostructional terms. Furthermore, Colleman and Bernolet (2012) also take into account the fact that different senses of a polysemous verb may well exhibit different argument structure preferences.

We set out to test, *inter alia*, whether collostructional analysis is extendable to Estonian, a morphologically rich language. We conclude that, at least for the *peal* construction, collostructional analysis provided meaningful insights into the alternation between the two Estonian morphosyntactic alternations. As to whether the collostructional analysis is extendable to morphology, we are left in the dark. We only found an effect of the collostructional strength for the *peal* construction, which in terms of collexeme analysis plays out nicely since the construction with *peal* has more to do with measuring the attraction of two separate words rather than attraction between a word and its inflection. Further research needs to be done with Estonian or other morphologically rich languages to test the applicability of the collostructional analysis to their morphology. Our models show that the collostructional strength between the word and the adessive case affix did not predict the choices in our forced choice task.

## 6.3 Does the inclusion of frequency significantly improve model fit?

For our third research question, we pitted the models with frequency metrics against a model with five other relevant variables identified in the previous research (length, mobility, complexity, syntactic function, trajector word class). Out of this "competition", the model with the collostructional measure for the *peal* construction was the "winner". The model with the collostructional measure in the formula on its own does a better job at predicting the choices in the task than a model with all the five variables combined. Yet, adding the collostructional frequency metric to the model in addition to the other variables yields an even better model. We, therefore, infer that frequency on its own is a particularly strong predictor, but does not explain away the other predictors since adding them into the picture is still significant.

In light of the result that the ratio variable is also significant, we may entertain the idea of whether the token frequency of the *peal* construction is actually an artifact of the relative frequency measure, i.e. the ratio variable. Hay (2001) discusses the role of relative versus absolute frequency effects in the decomposability of morphologically complex words and whether the two measures are correlated. Hay (2001: 1053) concludes that absolute frequency and relative frequency are not independent of one another. Our paper provides experimental evidence that a relative-frequency effect exists – participants consistently chose the *peal* construction more when the ratio of the adessive construction and the *peal* construction with the same word was small.

An important insight provided by our study is the fact that frequency affects the adessive and *peal* construction differently. The *peal* construction may not necessarily have to be very frequent for it to win out in the forced choice task; similarly, the adessive construction in combination with a particular word may be very frequent but still not affect the choices. What we observe in our data is the phenomenon referred to as the *inverse frequency effect*. This effect explains the observation that less frequent structures tend to prime more strongly (Bock 1986; Colleman and Bernolet 2012; Ferreira 2003; Jaeger and Snider 2013). For example, Bock (1986) showed syntactic persistence for passives, which are overall rather infrequent in her production experiments; the much more frequent active structure were not found to prime. In our study, we have also observed stronger frequency effects for the less frequent construction (the postpositional variant, in our case). Klavan and Schützler (2023) report the frequencies of the Estonian exterior locative cases (including the adessive) and the corresponding postpositions (including the postposition *peal*) from the Estonian National Corpus (ENC 2017; 1.1 billion words). These figures show that the adessive construction is 127 times more frequent in the language compared to the *peal* construction (30,661,120 adessive case occurrences vs. 241,263 *peal* construction occurrences).

Behind these drastically different raw frequency measures of adessive and *peal* lies the fact that the adessive case has other essential functions to fulfil in the Estonian language, for example expressing the possessor in the possessive construction (***Maril** on kaks last.* 'Mari has two kids.') and agents with finite verb forms (*See asi ununes **mul** kiiresti.* 'I quickly forgot about that thing.') (see Appendix G for more details about the different functions of the adessive case in Estonian). Support for the inverse frequency effect is found in Klavan (2020) who has shown that both the mixed-effects logistic regression model and the naïve discriminative learning model do a slightly better job in predicting the *peal* construction than the adessive construction. Since we only found an effect of the *peal* construction frequency to affect the choices, we can conclude that for adult native speakers of Estonian, a word in combination with *peal* is a unit of representation while a word in combination of the adessive is not. This conclusion needs further study, especially psycholinguistic experiments; particularly interesting would be to look at the role of these two constructions in language learning.

There are, of course, other frequency metrics which we can further investigate that are currently missing from our study. One such metric relates to the semantics of the adessive construction. In Estonian, the adessive case expresses other relations besides the locative (cf. Appendix G). Importantly, the adessive case is used to temporal relations and agents with finite verb forms, including the possessor in the possessive construction. For these functions of the adessive, postpositional construction is not an attested alternative construction. The polysemy of the adessive case creates a possible ambiguity for language users since it is not clear from the adessive form itself whether location or other more abstract functions are intended. In a pilot study, we calculated the adessive function ratio by counting the usages of the word in the adessive expressing the locative functions versus other functions. We found that if the ratio is negative, i.e. the adessive form with a particular word has a low proportion of uses where a locative function is expressed, the more probable the choice of the *peal* construction in the forced choice task. Further studies will hopefully shed light on whether the semasiological frequency of the adessive form influences the onomasiological frequency (expressing the locative function of the same word with adessive or postpositional construction). This can be tied in with research carried out by Mehl (2021: 228) who has shown that semasiological approaches are useful for research questions related to exposure rates while onomasiological approaches are useful for research questions related to selection preferences.

Similarly to the potential effect of the adessive case polysemy, verb sense has been suggested as an important predictor that needs to be considered in argument structure studies. Hare et al. (2003) suggest that the empirical discrepancies found in studies that have looked at how comprehension is affected by subcategorization biases of the verb arise because verb-sense has not been taken into account. For example, Gahl (2000), Garnsey et al. (1997), MacDonald (1994), and Trueswell et al. (1994) found an effect of subcategorization biases of the verb, while Ferreira and Henderson (1990), Kennison (2001), and Pickering et al. (2000) did not. The conclusion to be drawn from these studies is that different senses of a polysemous verb may exhibit different argument structure preferences. The results of our preliminary follow-up study about the polysemy of the adessive case clearly demonstrate that it is important to take into account the relative frequency of the locative adessive function versus other functions – an additional frequency measure of a different grain-size.

# 7 Conclusions

This paper investigated the role of corpus-based frequency metrics in predicting native speaker choice. We tested whether frequency affects the choice between two (broadly) synonymous constructions which represent the alternants of a morpho-syntactic alternation in Estonian – one between the adessive case construction and the corresponding postposition *peal* ‘onto’. We further tested which out of several well-established frequency metrics yields the most accurate predictions and whether frequency outperforms the predictiveness of other relevant semantic and morpho-syntactic variables identified in the previous research. In total, we looked at seven frequency measures: (1) word lemma frequency, (2) the frequency of the particular word in the adessive case, (3) the frequency of the genitive word form in combination with the postposition peal, (4) the ratio of the particular word in combination with the adessive construction divided by the word in combination with the peal construction, (5) collostructional strength of the word in the adessive construction, (6) collostructional strength of the word in the peal construction, and (7) the ratio of the two collostructional measures.

We report the findings of a forced choice task carried out with 103 native speakers of Estonian. Using mixed-effects logistic regression models with subjects as the random effect and the previous semantic and morpho-syntactic variables as well as the frequency metrics as fixed effects, we predict the choice of *peal* over adessive-marked realisations. Our experimental results show that frequency affects the choice between alternating morpho-syntactic constructions in Estonian. The cognitively most realistic metrics turned out to be the frequency of the particular word used together with the postposition *peal*; relative frequency of the word in the adessive construction and the word in the *peal* construction; the collostructional strength of the word in the *peal* construction. These results echo the results of previous studies that observe the inverse frequency effect – less frequent structures tend to prime more strongly. The frequency metrics that turned out to be insignificant were related to the adessive construction – the frequency of the particular word in the adessive case, the frequency of the word lemma as calculated logarithmically transforming the raw frequency counts of the lemma, and the collostructional strength of the word in the adessive construction.

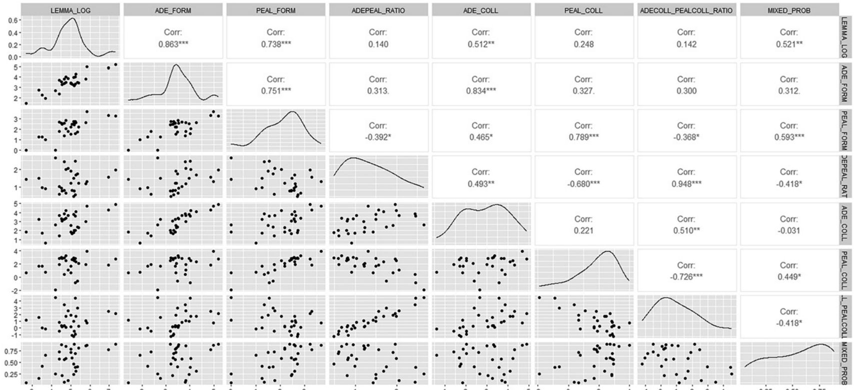
In terms of the direction of the effects, we found that the smaller the ratio between the two frequency measures (frequency of the word in the adessive construction/frequency of the word in the *peal* construction), the more probable the choice of the *peal* construction in the forced choice task. We conclude that frequency affects the two constructions in our study differently – the *peal* construction may not necessarily be very frequent for it to be win out in the forced choice task; likewise, the adessive construction that is very frequent may still not be chosen in the task. Based on these results we can say that if the frequency of the word is the same in both constructions, the preferred construction is the *peal* construction. Furthermore, we conclude that frequency shows independent effects alongside other semantic and morphosyntactic variables – frequency can be assigned the role of one functional predictor among others. Frequency on its own captures less variance than frequency combined with other predictors – frequency on its own does not explain away everything, or, in other words, frequency cannot be reduced to a correlate or epiphenomenon of other levels of languages.

Follow-up studies with other metrics (e.g. information-theoretic ones such as inflectional paradigm size, paradigm entropy, inflectional entropy, and relative inflectional entropy) and online experiments would be helpful in gaining a more complete understanding of the alternation between the synthetic case construction and the postpositional construction in Estonian. The choice of an experimental task may also influence whether frequency effects are found to play a role in explaining language variation or not since some experimental tasks may be more susceptible to frequency effects than others. It may further be predicted that forward looking metrics are more relevant in online processing tasks than in off-line tasks. This calls for further studies using the same set of stimuli, but different experimental methods. In the follow-up studies, we may want to use more fillers in order to make the competition of two constructions less salient.

It may very well be that the metrics taking into account the whole probability distribution between words and construction (not only the competing ones) become significant (e.g. entropy) for this type of morpho-syntactic alternation in Estonian and in other languages. Each experimental item in our study does not only co-occur with the two constructions that are of interest here, but also with other constructions (i.e. with other cases besides the adessive and with other postpositions besides *peal*). To achieve a more nuanced view of language users' probabilistic choices, it would be useful to take into account the whole range of morphosyntactic combinatorial probabilities. All in all, to gain a more complete understanding of the extent to which the constructionist uniformity claim applies to metrics "all the way down", follow-up research will be needed to explore whether our winning metrics are equally predictive of subjects' offline choices at other levels of language processing (e.g. at the level of fully abstract syntactic constructions).

**Data availability:** The dataset and the R-scripts used in the analysis of the data can be retrieved from https://osf.io/pkz32/. Note that the format of variables in the original dataset is not the same as the one used in this study: This article depends on the original (published) dataset in combination with the analytic scripts, which involve some re-formatting of variables. For the entire repository, we selected a CC BY 4.0 licence (https://creativecommons.org/licenses/by/4.0/).

# Appendix A: Correlation matrices of the stimuli (frequency metrics vs. probabilities)



# Appendix B: Experimental items together with the annotated categories

| Experimental item | LENGTH[3] | MOBIL. | COMPL. | SYNFUN | TRWC |
|---|---|---|---|---|---|
| aken ‘window’ | 2 | Static | Simple | Adverbial | Noun |
| diivan ‘couch’ | 3 | Mobile | Simple | Adverbial | Pronoun |
| ehitus ‘building site’ | 4 | Static | Simple | Adverbial | Pronoun |
| hing ‘soul’ | 2 | Mobile | Simple | Adverbial | Pronoun |
| jalg ‘leg’ | 3 | Mobile | Simple | Adverbial | VP |
| jää ‘ice’ | 1 | Static | Simple | Adverbial | Noun |
| kaal ‘scale’ | 2 | Mobile | Simple | Modifier | Noun |
| kallas ‘shore’ | 2 | Static | Simple | Adverbial | Pronoun |
| kirjutuslaud ‘writing desk’ | 5 | Mobile | Compound | Adverbial | VP |
| käis ‘cuff’ | 2 | Mobile | Simple | Adverbial | Noun |
| lehekülg ‘page’ | 7 | Mobile | Compound | Adverbial | Noun |
| lumi ‘snow’ | 2 | Static | Simple | Adverbial | Noun |
| maastik ‘landscape’ | 15 | Static | Simple | Adverbial | Noun |
| meri ‘sea’ | 2 | Static | Simple | Modifier | Noun |
| mis ‘what’ | 2 | Mobile | Simple | Adverbial | Noun |
| müür ‘stonewall’ | 6 | Static | Simple | Adverbial | VP |
| niit ‘meadow’ | 7 | Static | Simple | Adverbial | Noun |
| pilt ‘picture’ | 3 | Mobile | Simple | Adverbial | Pronoun |
| pink ‘bench’ | 2 | Mobile | Simple | Adverbial | Pronoun |
| puu ‘tree’ | 2 | Static | Simple | Adverbial | Noun |
| rada ‘path’ | 2 | Static | Simple | Modifier | Noun |
| rannakarjamaa ‘pastureland’ | 6 | Static | Compound | Adverbial | Noun |
| saar ‘island’ | 2 | Static | Simple | Adverbial | VP |
| see ‘this’ | 2 | Mobile | Simple | Adverbial | Pronoun |
| selg ‘back’ | 2 | Mobile | Simple | Adverbial | Noun |
| tee ‘road’ | 1 | Static | Simple | Adverbial | Pronoun |
| tool ‘chair’ | 7 | Mobile | Simple | Adverbial | Pronoun |
| trepp ‘stairs’ | 13 | Static | Simple | Adverbial | VP |
| tänavanurk ‘street corner’ | 8 | Static | Compound | Adverbial | Pronoun |

3 Length is measured in syllables and it refers not only to the lemma length, but the whole phrase, e.g. *tänavanurk* ‘street corner’ forms a part of the phrase *millisel tänavanurgal* ‘which street corner’ (8 syllables).

# Appendix C: Frequency metrics[4] of the experimental items (log transformed)

| Experimental item | ADEFORM | PEALFORM | ADEPEAL_RATIO | LEMMA |
|---|---|---|---|---|
| aken ‘window’ | 3.19 | 2.71 | 0.49 | 5.15 |
| diivan ‘couch’ | 3.68 | 2.64 | 1.04 | 4.40 |
| ehitus ‘building site’ | 3.42 | 1.51 | 1.91 | 5.02 |
| hing ‘soul’ | 3.19 | 2.39 | 0.79 | 5.19 |
| jalg ‘leg’ | 3.43 | 2.20 | 1.23 | 5.41 |
| jää ‘ice’ | 3.50 | 2.84 | 0.67 | 4.75 |
| kaal ‘scale’ | 3.40 | 2.18 | 1.22 | 5.01 |
| kallas ‘shore’ | 4.05 | 2.05 | 1.99 | 4.76 |
| kirjutuslaud ‘writing desk’ | 2.24 | 1.26 | 0.99 | 3.52 |
| käis ‘cuff’ | 1.91 | 1.00 | 0.91 | 3.69 |
| lehekülg ‘page’ | 3.84 | 1.78 | 2.06 | 4.87 |
| lumi ‘snow’ | 3.30 | 2.47 | 0.83 | 5.07 |
| maastik ‘landscape’ | 3.81 | 1.36 | 2.45 | 4.67 |
| meri ‘sea’ | 3.97 | 2.48 | 1.49 | 5.16 |
| mis ‘what’ | 4.87 | 3.30 | 1.57 | 6.97 |
| müür ‘stonewall’ | 2.36 | 1.76 | 0.60 | 4.41 |
| niit ‘meadow’ | 2.66 | 0.00 | 2.66 | 4.22 |
| pilt ‘picture’ | 3.82 | 2.61 | 1.20 | 5.77 |
| pink ‘bench’ | 3.53 | 2.58 | 0.95 | 4.53 |
| puu ‘tree’ | 3.47 | 1.72 | 1.75 | 5.36 |
| rada ‘path’ | 4.28 | 2.59 | 1.69 | 5.26 |
| rannakarjamaa ‘pastureland’ | 1.43 | 0.00 | 1.43 | 2.68 |
| saar ‘island’ | 4.01 | 1.54 | 2.47 | 5.25 |
| see ‘this’ | 5.19 | 3.22 | 1.97 | 7.33 |
| selg ‘back’ | 3.31 | 2.73 | 0.59 | 5.25 |
| tee ‘road’ | 5.00 | 3.68 | 1.32 | 5.84 |
| tool ‘chair’ | 3.39 | 2.57 | 0.82 | 4.67 |
| trepp ‘stairs’ | 3.26 | 2.45 | 0.81 | 4.55 |
| tänavanurk ‘street corner’ | 2.75 | 1.26 | 1.50 | 3.36 |

**4** All of the frequency metrics are reported by taking a natural log() from the raw frequency counts in ENC 2017.

# Appendix D: Collostructional strength of the experimental items

| Experimental item | ADE_COLL | PEAL_COLL | ADECOLL_PEALCOLL_RATIO |
|---|---:|---:|---:|
| aken ‘window’ | 56.559 | 891.0848 | 0.063472 |
| diivan ‘couch’ | 4,908.719 | 746.5966 | 6.574795 |
| ehitus ‘building site’ | 385.7847 | 4.622685 | 83.45468 |
| hing ‘soul’ | 103.2095 | 333.45 | 0.30952 |
| jalg ‘leg’ | 217.0486 | 111.9933 | 1.938051 |
| jää ‘ice’ | 2,292.129 | 1,988.093 | 1.152928 |
| kaal ‘scale’ | 72.38913 | 49.30108 | 1.468307 |
| kallas ‘shore’ | 16,978.1 | 303.6411 | 55.91503 |
| kirjutuslaud ‘writing desk’ | 50.20747 | 47.33672 | 1.060645 |
| käis ‘cuff’ | 3.968739 | 5.867883 | 0.676349 |
| lehekülg ‘page’ | 4,992.572 | 12.55464 | 397.6674 |
| lumi ‘snow’ | 48.14066 | 515.3855 | 0.093407 |
| maastik ‘landscape’ | 7,425.391 | 2.399504 | 3,094.552 |
| meri ‘sea’ | 6,647.012 | 671.6581 | 9.896422 |
| mis ‘what’ | 20,349.75 | 66.23035 | 307.2572 |
| müür ‘stonewall’ | 26.49384 | 79.91373 | 0.331531 |
| niit ‘meadow’ | 448.4114 | 0.012828 | 34,956.68 |
| pilt ‘picture’ | 973.61 | 78.74746 | 12.3637 |
| pink ‘bench’ | 2,018.776 | 969.819 | 2.081601 |
| puu ‘tree’ | 245.7765 | 0.257366 | 954.9691 |
| rada ‘path’ | 12,494.83 | 230.3654 | 54.23918 |
| rannakarjamaa ‘pastureland’ | 69.78827 | 4.653581 | 14.99668 |
| saar ‘island’ | 3,815.452 | 0.13163 | 28,986.24 |
| see ‘this’ | 76,558.29 | 543.5263 | 140.8548 |
| selg ‘back’ | 104.9164 | 752.4148 | 0.13944 |
| tee ‘road’ | 48,954.56 | 9,308.656 | 5.259036 |
| tool ‘chair’ | 1,762.541 | 856.5675 | 2.057679 |
| trepp ‘stairs’ | 1,142.43 | 691.6368 | 1.651777 |
| tänavanurk ‘street corner’ | 1,789.476 | 48.43804 | 36.9436 |

# Appendix E: Model comparison statistics for the models with different frequency measures

**Table A:** Results of likelihood ratio tests for the models with different frequency measures.

| Model | logLik | Chisq | Chi.Df | *p* value | Reduction in AIC |
|---|---|---|---|---|---|
| Model 0: intercept only | −1,878.1 | | | | |
| Model 1: ~ ade_form | −1,863.6 | 29.101 | 1 | 0.0000 | 27.1 |
| Model 2: ~ peal_form | −1,780.0 | 196.34 | 1 | 0.0000 | 194.4 |
| Model 3: ~ adepeal_ratio | −1,803.5 | 149.29 | 1 | 0.0000 | 147.3 |
| Model 4: ~ lemma | −1,847.0 | 62.277 | 1 | 0.0000 | 60.3 |
| Model 5: ~ ade_coll | −1,878.1 | 0.055 | 1 | 0.815 | −1.9 |
| Model 6: ~ peal_coll | −1,782.7 | 190.83 | 1 | 0.0000 | 188.9 |
| Model 7: ~ adecoll_pealcoll_ratio | −1,804.6 | 146.99 | 1 | 0.0000 | 145 |

**Table B:** Coefficients for the models with different frequency measures.

| Model | Estimate | Std. error | *z* value | *p* value |
|---|---|---|---|---|
| Model 0: intercept only | −0.66948 | 0.07696 | −8.699 | 0.0000 |
| Model 1: ~ ade_form | 0.26306 | 0.04914 | 5.354 | 0.0000 |
| Model 2: ~ peal_form | 0.71960 | 0.05522 | 13.03 | 0.0000 |
| Model 3: ~ adepeal_ratio | −0.86857 | 0.07421 | −11.705 | 0.0000 |
| Model 4: ~ lemma | 0.34514 | 0.04447 | 7.76 | 0.0000 |
| Model 5: ~ ade_coll | −0.00881 | 0.03755 | −0.24 | 0.815 |
| Model 6: ~ peal_coll | 0.46664 | 0.03674 | 12.7 | 0.0000 |
| Model 7: ~ adecoll_pealcoll_ratio | −0.34402 | 0.02980 | −11.542 | 0.0000 |

# Appendix F: Results of likelihood ratio tests

**Table A:** Results of likelihood ratio tests for Model 1: peal_form + length + mobility + complexity + synfun + trwc + (1|subject).

| Predictor | Model comparison | Chisq | Chi.Df | *p* value |
|---|---|---|---|---|
| length | Full vs. no length | 35.575 | 1 | 0.0000 |
| mobility | Full vs. no mobility | 2.943 | 1 | 0.0863 |
| complexity | Full vs. no complexity | 2.861 | 1 | 0.0908 |
| synfun | Full vs. no synfun | 1.6203 | 1 | 0.2031 |
| trwc | Full vs. no trwc | 33.27 | 1 | 0.0000 |

**Table A:** (continued)

| Predictor | Model comparison | Chisq | Chi.Df | *p* value |
|---|---|---|---|---|
| peal_form | Full vs. no peal_form | 40.938 | 1 | 0.0000 |

**Table B:** Results of likelihood ratio tests for Model 3: peal_coll + length + mobility + complexity + synfun + trwc + (1|subject).

| Predictor | Model comparison | Chisq | Chi.Df | *p* value |
|---|---|---|---|---|
| length | Full vs. no length | 45.019 | 1 | 0.0000 |
| mobility | Full vs. no mobility | 1.757 | 1 | 0.185 |
| complexity | Full vs. no complexity | 14.493 | 1 | 0.0001 |
| synfun | Full vs. no synfun | 3.704 | 1 | 0.0543 |
| trwc | Full vs. no trwc | 22.919 | 1 | 0.0000 |
| peal_coll | Full vs. no peal_coll | 69.975 | 1 | 0.0000 |

# Appendix G: Functions of the Estonian adessive case (adapted from Erelt et al. 2007: 249–251)

| Adessive function | Example sentence | Postpositional alternative | English translation |
|---|---|---|---|
| Location | *Vaas on **laual**.* | *Vaas on **laua peal**.* | ‘The vase is on the table.’ |
| Time | *Nad sõidavad **nelja-päeval** maale.* | Not attested | ‘They are driving to the country on Thursday.’ |
| State | *Jüri vaatas meid **naerul** näoga.* | Not attested | ‘Jüri looked at us with a laughing face.’ |
| Possessor | ***Maril** on kaks last.* | Not attested | ‘Mari has two children.’ (lit. ‘On Mary are two children.’) |
| Agent with finite verb forms | *See asi ununes **mul** kiiresti.* | Not attested | ‘I quickly forgot about that thing.’ |
| Instrument | *Mari mängib **klaveril** mõnd lugu.* | *Mari mängib **klaveri peal** mõnd lugu.* | ‘Mari is playing some tunes on the piano.’ |
| Manner | *Mari kuulas kikkis **kõrvul**.* | Not attested | ‘Mari listened with her ears pricked up.’ |

# References

Allen, Kachina, Francisco Pereira, Matthew Botvinick & Adele E. Goldberg. 2012. Distinguishing grammatical constructions with fMRI pattern analysis. *Brain and Language* 123(3). 174–182.

Ambridge, Ben, Julian M. Pine, Caroline F. Rowland, Franklin Chang & Amy Bidgood. 2013. The retreat from overgeneralization in child language acquisition: Word learning, morphology, and verb argument structure. *Wiley Interdisciplinary Reviews: Cognitive Science* 4(1). 47–62.

Ambridge, Ben, Julian M. Pine, Caroline F. Rowland, Daniel Freudenthal & Franklin Chang. 2014. Avoiding dative overgeneralisation errors: Semantics, statistics or both? *Language, Cognition and Neuroscience* 29(2). 218–243.

Ambridge, Ben, Amy Bidgood, Katherine E. Twomey, Julian M. Pine, Caroline F. Rowland & Daniel Freudenthal. 2015. Preemption versus entrenchment: Towards a construction-general solution to the problem of the retreat from verb argument structure overgeneralization. *PLoS One* 10(4). e0123723.

Arnon, Inbal & Morten H. Christiansen. 2017. The role of multiword building blocks in explaining L1–L2 differences. *Topics in Cognitive Science* 9(3). 621–636.

Arnon, Inbal & Uriel Cohen Priva. 2013. More than words: The effect of multi-word frequency and constituency on phonetic duration. *Language and Speech* 56(3). 349–371.

Baayen, R. Harald. 2010. Demythologizing the word frequency effect: A discriminative learning perspective. *The Mental Lexicon* 5(3). 436–461.

Baayen, R. Harald, Laura A. Anna Endresen, Anastasia Makarova Janda & Tore Nesset. 2013. Making choices in Russian: Pros and cons of statistical methods for rival forms. *Russian Linguistics* 37. 253–291.

Bartens, Raija. 1978. Synteettiset ja analyyttiset rakenteet lapin paikanilmauksissa [Synthetic and analytic constructions in saami locative expressions]. In *Suomalais-ugrilaisen Seuran toimituksia*, vol. 166. Helsinki: Suomalais-Ugrilainen Seura.

Bates, Douglas. 2014. Computational methods for mixed models. http://cran.r-project.org/web/packages/lme4/vignettes/Theory.pdf (accessed 27 March 2019).

Bates, Douglas, Martin Maechler, Ben Bolker, Steven Walker, Rune Haubo Bojesen Christensen, Henrik Singmann & Bin Dai. 2015. lme4. http://cran.r-project.org/web/packages/lme4/lme4.pdf (accessed 27 March 2019).

Bley-Vroman, Robert. 2002. Frequency in production, comprehension, and acquisition. *Studies in Second Language Acquisition* 24(2). 209–213.

Blumenthal-Dramé, Alice. 2012. *Entrenchment in usage-based theories: What corpus data do and do not reveal about the mind*. Berlin, Boston: De Gruyter Mouton.

Blumenthal-Dramé, Alice. 2016a. Entrenchment from a psycholinguistic and neurolinguistic perspective. In Hans-Jörg Schmid (ed.), *Entrenchment and the psychology of language learning how we reorganize and adapt linguistic knowledge*, 129–152. Berlin, Boston: De Gruyter Mouton.

Blumenthal-Dramé, Alice. 2016b. What corpus-based cognitive linguistics can and cannot expect from neurolinguistics. *Cognitive Linguistics* 27(4). 493–505.

Bock, J. Kathryn. 1986. Syntactic persistence in language production. *Cognitive Psychology* 18(3). 355–387.

Bod, Rens. 1998. *Beyond grammar: An experienced-based theory of language*. Stanford, CA: CSLI Publications.

Bod, Rens. 2006. Exemplar-based syntax: How to get productivity from examples. *The Linguistic Review* 23. 291–320.

Bresnan, Joan & Marilyn Ford. 2010. Predicting syntax: Processing dative constructions in American and Australian varieties of English. *Language* 86(1). 186–213.

Bresnan, Joan, Anna Cueni, Tatiana Nikitina & R. Harald Baayen. 2007. Predicting the dative alternation. In Gerlof Bouma, Irene Krämer & Joost Zwarts (eds.), *Cognitive foundations of interpretation*, 69–94. Amsterdam: Royal Netherlands Academy of Science.

Bridgwater, Emma, Aki-Juhani Kyröläinen & Victor Kuperman. 2019. The influence of syntactic expectations on reading comprehension is malleable and strategic: An eye-tracking study of English dative alternation. *Canadian Journal of Experimental Psychology* 73(3). 179.

Brysbaert, Marc, Paweł Mandera & Emmanuel Keuleers. 2018. The word frequency effect in word processing: An updated review. *Current Directions in Psychological Science* 27(1). 45–50.

Bybee, Joan. 2007. *Frequency of use and the organization of language*. Oxford: Oxford University Press.

Bybee, Joan. 2010. *Language, usage and cognition*. Cambridge: Cambridge University Press.

Caldwell-Harris, Catherine L. 2021. Frequency effects in reading are powerful–But is contextual diversity the more important variable? *Language and Linguistics Compass* 15(12). e12444.

Cappelle, Bert. 2006. Particle placement and the case for 'allostructions'. *Constructions*. https://hal.science/hal-01495786/.

Colleman, Timothy & Sarah Bernolet. 2012. Alternation biases in corpora vs. picture description experiments: DO-biased and PD-biased verbs in the Dutch dative alternation. In Dagmar Divjak & Stefan Th. Gries (eds.), *Frequency effects in language representation*, vol. 2, 87–125. Berlin, Boston: De Gruyter Mouton.

Comrie, Bernard. 1986. Markedness, grammar, people, and the world. In Fred R. Eckman, Edith A. Moravcsik & Jessica R. Wirth (eds.), *Markedness*, 85–106. New York: Plenum.

Conway, Christopher M. & Morten H. Christiansen. 2009. Seeing and hearing in space and time: Effects of modality and presentation rate on implicit statistical learning. *European Journal of Cognitive Psychology* 21(4). 561–580.

de Vega, Manuel, María J. Rodrigo, Manuel Ato, Doris M. Dehn & Beatriz Barquero. 2002. How nouns and prepositions fit together: An exploration of the semantics of locative sentences. *Discourse Processes* 34(2). 117–143.

Divjak, Dagmar. 2017. The role of lexical frequency in the acceptability of syntactic variants: Evidence from that- clauses in polish. *Cognitive Science* 41(2). 354–382.

Divjak, Dagmar, Antti Arppe & Ewa Dąbrowska. 2016. Machine meets man: Evaluating the psychological reality of corpus-based probabilistic models. *Cognitive Linguistics* 27(1). 1–33.

Erelt, Mati, Reet Kasik, Helle Metslang, Henno Rajandi, Kristiina Ross, Henn Saari, Kaja Tael & Silvi Vare. 1995. *Eesti keele grammatika I. Morfoloogia* [The Grammar of Estonian I. Morphology]. Tallinn: Eesti Teaduste Akadeemia Eesti Keele Instituut.

Erelt, Mati, Tiiu Erelt & Kristiina Ross. 2007. *Eesti keele käsiraamat* [Handbook of Estonian]. Tallinn: Eesti Keele Sihtasutus.

Evert, Stefan. 2009. Corpora and collocations. In Anke Lüdeling & Merja Kytö (eds.), *Corpus linguistics: An international handbook*, vol. 2, 1212–1248. Berlin, New York: Mouton de Gruyter.

Ferreira, Victor S. 2003. The processing basis of syntactic persistence: We repeat what we learn. Paper presented at the 44th Annual Meeting of the Psychonomic Society. Vancouver, Canada.

Ferreira, Fernanda & John M. Henderson. 1990. Use of verb information in syntactic parsing: Evidence from eye movements and word-by-word self-paced reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 16(4). 555–568.

Frank, Stefan L. & Rens Bod. 2011. Insensitivity of the human sentence-processing system to hierarchical structure. *Psychological Science* 22(6). 829–834.

Frank, Stefan L., Leun J. Otten, Giulia Galli & Gabriella Vigliocco. 2015. The ERP response to the amount of information conveyed by words in sentences. *Brain and Language* 140. 1–11.

Frost, Ram, Blair C. Armstrong, Noam Siegelman & Morten H. Christiansen. 2015. Domain generality versus modality specificity: The paradox of statistical learning. *Trends in Cognitive Sciences* 19(3). 117–125.

Gahl, Susanne. 2000. A usage-based model of aphasic sentence comprehension. UC Berkeley doctoral dissertation.

Garnsey, Susan M., Neal J. Pearlmutter, Elizabeth Myers & Melanie A. Lotocky. 1997. The contributions of verb bias and plausibility to the comprehension of temporarily ambiguous sentences. *Journal of Memory and Language* 37(1). 58–93.

Goldberg, Adele E. 2006. *Constructions at work: The nature of generalization in language*. Oxford: Oxford University Press.

Goldberg, Adele E. 2016. Partial productivity of linguistic constructions: Dynamic categorization and statistical preemption. *Language and Cognition* 8(3). 369–390.

Gries, Stefan Th. 2012. Frequencies, probabilities, and association measures in usage-/exemplar-based linguistics: Some necessary clarifications. *Studies in Language* 36(3). 477–510.

Gries, Stefan Th. 2014. Coll.analysis 3.5. A script for R to compute perform collostructional analyses (major update to handle larger corpora/frequencies).

Gries, Stefan Th. & Anatol Stefanowitsch. 2004. Extending collostructional analysis: A corpus-based perspective onalternations'. *International Journal of Corpus Linguistics* 9(1). 97–129.

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.

Hagège, Claude. 2010. *Adpositions*. Oxford: Oxford University Press.

Hale, John. 2016. Information-theoretical complexity metrics: Information-theoretical complexity metrics. *Language and Linguistics Compass* 10(9). 397–412.

Hare, Mary, Ken McRae & Jeffrey L. Elman. 2003. Sense and structure: Meaning as a determinant of verb subcategorization preferences. *Journal of Memory and Language* 48(2). 281–303.

Harrell, Frank E. 2001. *Regression modeling strategies. With applications to linear models, logistic regression and survival analysis*. New York: Springer-Verlag.

Hasher, Lynn & Rose T. Zacks. 1984. Automatic processing of fundamental information: The case of frequency of occurrence. *American Psychologist* 39(12). 1372.

Hay, Jennifer. 2001. Lexical frequency in morphology: Is everything relative? *Linguistics* 39(6). 1041–1070.

Hosmer, W. David, Jr., Stanley Lemeshow & Rodney X. Sturdivant. 2013. *Applied logistic regression*. Hoboken, NJ: John Wiley & Sons.

Jaeger, T. Florian & Neal E. Snider. 2013. Alignment as a consequence of expectation adaptation: Syntactic priming is affected by the prime's prediction error given both prior and recent experience. *Cognition* 127(1). 57–83.

Jolsvai, Hajnal, Stewart M. McCauley & Morten H. Christiansen. 2020. Meaningfulness beats frequency in multiword chunk processing. *Cognitive Science* 44(10). e12885.

Kaalep, Heiki-Jaan. 1998. Tekstikorpuse abil loodud eesti keele morfoloogiaanalüsaator. *Keel ja Kirjandus* 1. 22–29.

Kallas, Jelena & Kristina Koppel 2017. Eesti keele ühendkorpus 2017. [Estonian National Corpus 2017]. https://doi.org/10.15155/3-00-0000-0000-0000-071E7L.

Kennison, Shelia M. 2001. Limitations on the use of verb information during sentence comprehension. *Psychonomic Bulletin & Review* 8(1). 132–138.

Klavan, Jane. 2012. *Evidence in linguistics: Corpus-linguistic and experimental methods for studying grammatical synonymy* (Dissertationes Linguisticae Universitatis Tartuensis). Tartu: University of Tartu Press.

Klavan, Jane. 2020. Pitting corpus-based classification models against each other: A case study for predicting constructional choice in written Estonian. *Corpus Linguistics and Linguistic Theory* 16(2). 363–391.

Klavan, Jane. 2021. The alternation between exterior locative cases and postpositions in Estonian web texts. *Eesti ja soome-ugri keeleteaduse ajakiri/Journal of Estonian and Finno-Ugric Linguistics* 12(1). 153–188.

Klavan, Jane & Dagmar Divjak. 2016. The cognitive plausibility of statistical classification models: Comparing textual and behavioral evidence. *Folia Linguistica* 50(2). 355–384.

Klavan, Jane & Ole Schützler. 2023. The complexity principle and the morphosyntactic alternation between case affixes and postpositions in Estonian. *Cognitive Linguistics* 34(2). 297–331.

Klavan, Jane & Ann Veismann. 2017. Are corpus-based predictions mirrored in the preferential choices and ratings of native speakers? Predicting the alternation between the Estonian adessive case and the adposition peal 'on'. *Eesti ja soome-ugri keeleteaduse ajakiri/Journal of Estonian and Finno-Ugric Linguistics* 8(2). 59–91.

Klavan, Jane, Maarja-Liisa Pilvik & Kristel Uiboaed. 2015. The use of multivariate statistical classification models for predicting constructional choice in spoken, non-standard varieties of Estonian. *SKY Journal of Linguistics* 28. 187–224.

Kronenfeld, David B. 2006. Formal rules, cognitive representations and learning in language and other cultural systems. *Language Sciences* 28(4). 424–435.

Kuperman, Victor & Joan Bresnan. 2012. The effects of construction probability on word durations during spontaneous incremental sentence production. *Journal of Memory and Language* 66(4). 588–611.

Lestrade, Sander. 2010. *The space of case*. Unpublished doctoral dissertation. Nijmegen: Radboud University Nijmegen.

Lõo, Kaidi, Juhani Järvikivi, Fabian Tomaschek, Benjamin V. Tucker & R. Harald Baayen. 2018. Production of Estonian case-inflected nouns shows whole-word frequency and paradigmatic effects. *Morphology* 28(1). 71–97.

Lowder, Matthew W., Wonil Choi, Fernanda Ferreira & John M. Henderson. 2018. Lexical predictability during natural reading: Effects of surprisal and entropy reduction. *Cognitive Science* 42. 1166–1183.

MacDonald, Maryellen C. 1994. Probabilistic constraints and syntactic ambiguity resolution. *Language & Cognitive Processes* 9. 157–201.

McConnell, Kyla & Alice Blumenthal-Dramé. 2022. Effects of task and corpus-derived association scores on the online processing of collocations. *Corpus Linguistics and Linguistic Theory* 18(31). 33–76.

Mehl, Seth. 2021. What we talk about when we talk about corpus frequency: The example of polysemous verbs with light and concrete senses. *Corpus Linguistics and Linguistic Theory* 17(1). 223–247.

Newmeyer, Frederick J. 2003. Grammar is grammar and usage is usage. *Language* 79(4). 682–707.

Ojutkangas, Krista. 2008. Mihin suomessa tarvitaan sisä-grammeja? [When are sisä grams used in Finnish?]. *Virittäjä* 3. 382–400.

Perek, Florent. 2012. Alternation-based generalizations are stored in the mental grammar: Evidence from a sorting task experiment. *Cognitive Linguistics* 23(3). 601–635.

Perruchet, Pierre & Ronald Peereman. 2004. The exploitation of distributional information in syllable processing. *Journal of Neurolinguistics* 17(2–3). 97–119.

Pickering, Martin J., Matthew J. Traxler & Matthew W. Crocker. 2000. Ambiguity resolution in sentence processing: Evidence against frequency-based accounts. *Journal of Memory and Language* 43(3). 447–475.

Pierrehumbert, Janet B. 2001. Exemplar dynamics: Word frequency, lenition and contrast. *Typological Studies in Language* 45. 137–158.

Pierrehumbert, Janet B. 2006. The statistical basis of an unnatural alternation. *Laboratory Phonology* 8. 81–107.

Pinheiro, José C. & Douglas M. Bates. 2002. *Mixed-effects models in S and S-PLUS*. New York: Springer.

Potter, Christine E., Tianlin Wang & Jenny R. Saffran. 2017. Second language experience facilitates statistical learning of novel linguistic materials. *Cognitive Science* 41. 913–927.

Robenalt, Clarice & Adele Goldberg. 2015. Judgment evidence for statistical preemption: It is relatively better to vanish than to disappear a rabbit, but a lifeguard can equally well backstroke or swim children to shore. *Cognitive Linguistics* 26(3). 467–503.

Saksida, Amanda, Alan Langus & Marina Nespor. 2017. Co-occurrence statistics as a language-dependent cue for speech segmentation. *Developmental Science* 20(3). e12390.

Shetreet, Einat, Tal Linzen & Naama Friedmann. 2016. Against all odds: Exhaustive activation in lexical access of verb complementation options. *Language, Cognition and Neuroscience* 31(9). 1206–1214.

Siegelman, Noam. 2020. Statistical learning abilities and their relation to language. *Language and Linguistics Compass* 14(3). e12365.

Siegelman, Noam, Louisa Bogaerts & Ram Frost. 2017. Measuring individual differences in statistical learning: Current pitfalls and possible solutions. *Behavior Research Methods* 49. 418–432.

Smith, Nathaniel J. & Roger Levy. 2013. The effect of word predictability on reading time is logarithmic. *Cognition* 128(3). 302–319.

Snider, Neal. 2008. *An exemplar model of syntactic priming*. Ann Arbor, MI: Stanford University.

Szmrecsanyi, Benedikt, Jason Grafmiller, Benedikt Heller & Melanie Röthlisberger. 2016. Around the world in three alternations: Modeling syntactic variation in varieties of English. English World-Wide. *A Journal of Varieties of English* 37(2). 109–137.

Thiessen, Erik D. 2017. What's statistical about learning? Insights from modelling statistical learning as a set of memory processes. *Philosophical Transactions of the Royal Society B: Biological Sciences* 372(1711). 20160056.

Trueswell, John C., Michael K. Tanenhaus & Susan M. Garnsey. 1994. Semantic influences on parsing: Use of thematic role information in syntactic ambiguity resolution. *Journal of Memory and Language* 33. 285–318.

Van de Velde, Freek. 2014. Degeneracy: The maintenance of constructional networks. In Ronny Boogaart, Timothy Colleman & Gijsbert Rutten (eds.), *The extending scope of construction grammar*, 141–179. Berlin: De Gruyter Mouton.