DE GRUYTER Cognitive Linguistics 2016; 27(2): 235–268

Natalia Levshina*

# When variables align: A Bayesian multinomial mixed-effects model of English permissive constructions

DOI 10.1515/cog-2015-0054
Received May 23, 2015; revised November 25, 2015; accepted February 14, 2016

**Abstract:** This paper is a quantitative multifactorial study of the near-synonymous constructions *let* + V, *allow* + *to* V and *permit* + *to* V based on the British National Corpus. The study investigates the differences between these constructions with the help of 23 formal, semantic, social and collostructional variables. A Bayesian multinomial mixed-effects model reveals a remarkable alignment of the variables that represent different dimensions of variation, namely, the linguistic distance between the predicates, the conceptual distance between the events they represent, the distance between the speaker and the Permitter and Permittee on the animacy/entrenchment/empathy hierarchy, the social and communicative distance between the interlocutors, as well as the strength of collostructional attraction between the constructions and second verb slot fillers. The paper offers several possible explanations for this alignment from a cognitive, functional and historical perspective.

**Keywords:** permissive causation, multifactorial grammar, multinomial mixed-effects models, Bayesian statistics, iconicity, frequency

## 1 Introduction

This paper is inspired by two influential interacting trends in Cognitive Linguistics and related disciplines, which are observed on the theoretical and methodological levels. On the theoretical level, there is a tendency towards integration of cognitive, social and historical dimensions in the description of linguistic phenomena. This trend is represented, in particular, by such hybrid disciplines as social cognitive linguistics (Croft 2009) and Cognitive Sociolingusitics (Kristiansen and Dirven 2008; Geeraerts et al. 2010). One should also mention new integrative socio-cognitive theories, such as the Entrenchment-and-Conventionalization model (Schmid 2014). At the methodological level, there is an increasing interest

***Corresponding author: Natalia Levshina,** Leipzig University, Leipzig, Germany,
E-mail: natalia.levshina@uni-leipzig.de

in multivariate quantitative models of language use. This trend manifests itself in a growing number of quantitative studies that determine the factors predicting the use of functionally similar words and constructions in different languages and language varieties, such as English phrasal constructions with varying particle placement (Gries 2003), German middle field alternation (Heylen 2005), Finnish verbs of thinking (Arppe 2008), Russian verbs of trying (Divjak 2010), existential constructions in York English (Tagliamonte and Baayen 2012), clitics in Spanish (Miglio et al. 2013) and topic markers in Shanghainese (Han et al., in press), to name just a few.

The object of this study is variation of three English permissive constructions: *let* + V, *allow* + *to* V and *permit* + *to* V. Examples are provided in (1):

(1) a. *I am content to let you form your own judgment of my character*. (H84)[1]
b. *Representation 1 allows us to depict any set of pairs of coordinates*. (FNR)
c. *In this form the censor permitted the book to pass*. (B7K)

Unlike other causative constructions, such as *make* + V, *have* + V or the *into*-causative, the permissive constructions have been at the periphery of researchers' attention. We will investigate several dimensions of variation of the constructions: formal, semantic, cognitive, social and collostructional.[2] To the best of our knowledge, this is the first study that investigates how all these dimensions together are aligned in near-synonymous constructions and which offers an interpretation of this alignment from a cognitive, functional and historical perspective.

From a methodological point of view, this paper is innovative, as well: the effect of the 23 variables that represent the dimensions is tested with the help of a cutting-edge statistical technique, Bayesian mixed-effects multinomial regression analysis, implemented in the R package *RStan*. This package is based on Stan, a programming language and platform for Bayesian inference (Carpenter et al., in press). One of the main epistemological advantages of the Bayesian approach is that it allows the researcher not only to test if the null hypothesis can be rejected, as in traditional frequentist statistics, but also to test the alternative hypothesis directly by estimating the probability of parameter values given the data. This gives a fuller picture of the object of investigation and

**1** A three-character code indicates that the example is taken from the British National Corpus. The code identifies the corpus component. An absence of code means that the example has been constructed by the author.

**2** That is, related to the association between constructions and lexemes that fill in constructional slots, following the ideas of collostructional analysis in Stefanowitsch and Gries (2003) and later works.

discourages the mechanistic dichotomous decisions made on the basis of *p*-values. From a more practical perspective, this method represents a welcome addition to the toolkit for analysis of three and more near-synonyms, since the algorithms for mixed-effects multinomial models in R are not particularly well developed at the moment. More technically speaking, the Bayesian approach offers convenient solutions to notorious statistical problems, such as data sparseness in models with crossed random effects (cf. Bates et al. 2015) and complete or quasi-complete separation.

The remaining part of the paper is organized as follows. Section 2 discusses the theoretical background and presents the factors discussed in previous research. Section 3 discusses the data and the 23 variables. Section 4 presents the results of the Bayesian mixed-effects multinomial regression analysis. Finally, Section 5 offers possible cognitive, functional and historical explanations of the results.

# 2 Theoretical background

## 2.1 Permissive causation

Causation in general involves an interaction of two entities, which are called the Agonist and Antagonist in Talmy's (2000) theory of force dynamics. In "normal" causation the Antagonist overrides the Agonist's intrinsic tendency towards some action or state. An example is provided in (2a), where the mother overrides the girl's intrinsic tendency (which is to keep playing rather than brushing her teeth and going to bed). This type of causation is sometimes called factitive (Nedjalkov 1976: Ch. 3). In contrast, permissive causation is observed in situations when the Antagonist does not override the Agonist's intrinsic tendency, as in (2b).

(2) a. *The mother made the girl brush her teeth and go to bed.*
b. *The mother let the girl play in the yard.*

In what follows, the Antagonist of permissive causation (the mother) will be called the Permitter, and the Agonist (the girl) the Permittee. Note that the Permitter and Permittee are treated here as roles characterizing the corresponding constructional slots, rather than as deep semantic roles. The event expressed by the permissive verb (V1) will be called the permitting event, and the event expressed by the infinitive (V2) the permitted event.

Cross-linguistically, permissive constructions are not as well-explored as the causative constructions that express factitive causation. The former are often omitted from a discussion of causative inventories of individual languages. However, there are several studies of the English permissive constructions within the broader domain of infinitival complementation (Duffley 1992; Egan 2008), as well as some observations in general functionalist theories (e. g., Givón 1980). These studies focus mostly on the formal and semantic differences between the constructions, although other factors, such as collocational fixation and domain of use, are briefly mentioned, as well. In addition, some reference grammars and dictionaries of English point to the stylistic differences between the verbs of permission.

## 2.2 Formal and semantic differences

Duffley's study of permissive constructions (1992: Section 2.9) is a part of his larger study of the English infinitive. In his theory, the bare infinitival complement generally implies that the event expressed by the infinitive is concurrent with the matrix verb, whereas the *to*-infinitive expresses a future or potential event, which is or may be actualized as a consequence of the event expressed by the matrix verb (Duffley 1992: 88–90). The constructions with *allow* and *permit* represent permission without saying whether the permitted action was carried out or not (Duffley 1992: 83). In other words, the verbs *allow* and *permit* are non-implicative (Karttunen 1971). What is permitted or allowed, can be carried out some time later, or never. In contrast, *let* is an implicative verb, which entails that the permitted action takes place. Consider the sentences in (3). While it is possible to say (3a), (3b) sounds odd.

(3) a. *I allowed/permitted him to go, but he chose to stay.*
 b. *??I let him go, but he chose to stay.*

In other words, *let* expresses situations when the act of permission is construed as inseparable from the realization of the permitted event, while *allow* and *permit* only denote the prior condition for the permitted event (Duffley 1992: 86).

This observation can be interpreted following the more general framework of binding hierarchy, which posits an iconic correlation between the semantic integration of events and syntactic binding of predicates. According to this hierarchy, the constructions with *allow* and *permit* exhibit a lower degree of semantic and syntactic binding than the construction with *let* (Givón 1980: 357, 369).

Egan (2008) takes a different approach and investigates the conceptual difference between *let* and *allow* on the basis of Talmy's theory of force dynamics (2000: Ch. 7). More specifically, he applies Talmy's distinction between onset and extended letting. Onset letting corresponds to situations when the Permitter removes some previously existing barrier, as in (4a), whereas extended letting is observed when the Permitter does not intervene at all, as in (4b).

(4) a. *Let my people go!*
b. *Let it be.*

According to Egan's corpus study, *allow* in non-negative contexts occurs more frequently in situations on barrier-removal, or onset letting, as in (5a), whereas *let* is predominantly used in situations without a prior impingement, as in (5b).

(5) a. For it was he who allowed Jews to re-enter Britain. (BN3)
b. *Let your letter express your personality*. (EEB)

An exception is the "release" sense (Egan 2008: 225), as in (6), which is expressed predominantly by *let*:

(6) *Slowly and sadly Lucker lets my wrist drop*. (HH0)

Although all these observations may look unrelated at first sight, Egan's results in fact tie in with Duffley's (1992) account: non-imposition and releasing imply at least some spatiotemporal overlap between the events, whereas a barrier removal only provides conditions for the permitted action to take place.

In addition, Egan (2008: 215–216) shows that *let* is more frequently used with the 1$^{st}$ and 2$^{nd}$ person matrix subjects than *allow* and *permit*. In contrast, *allow* and *permit* occur more frequently with inanimate subjects. See examples in (7).

(7) a. ***I****'m going to let you go*. (BN3)
b. *The advantage of a lens is that* **it** *allows the image to be both sharp and bright*. (J52)
c. *You do not mention* <u>*a way*</u> ***that*** *permits a company to become bigger, and leaner*. (CR9)

Yet another difference is that *let* is considerably more frequently used in the imperative form, as in (8), than the other two verbs.

(8) *Let me explain*. (FRS)

## 2.3 Register and domain of use

The description of semantic differences in the previous section mostly focused on *let* as opposed to *allow* and, less systematically, *permit*. It seems that the difference between *allow* and *permit* is mostly stylistic and domain-specific. In some English dictionaries, the verb *permit* has a note "formal", i. e., suitable for formal speech in writing, but not commonly used in ordinary conversation (e. g., LDCE: 1222). Some grammars also mention that the construction *permit* + *to* V is more formal than the construction with *allow* (Leech and Svartvik 1994: 163).

*Permit* also seems to be preferred in legal discourse. According to the Longman Dictionary of Contemporary English, the first meaning of *permit* is "to allow something to happen, especially by an official decision, rule, or law" (LDCE: 1222). In a similar vein, Egan (2008: 217) observes that *permit* is more likely to be used when the Permitter is a legal directive, as in (9a), whereas *allow* is used when the Permitter can be described as some circumstances, as in (9b):

(9) a. ***The Long Term Cotton Textile Agreement of 1962 (LTA)*** *permitted the developed countries to check imports from developing countries*. (EEF)
b. *Of course*, ***improving the facilities*** *ought to allow you to charge higher prices*. (CDF)

Historically, *let* is part of the Anglo-Saxon vocabulary, whereas *allow* and *permit* are Latinate borrowings, which are more typical of formal discourse (cf. Mittwoch 1990: 125). Thus, we can expect to find a cline of formality, from *let* as the least formal verb to *allow* and finally to *permit* as the most formal one, as shown in (10):

(10) *let* [least formal] < *allow* < *permit* [most formal]

## 2.4 Collocational differences

In addition to the semantic differences mentioned in Section 2.2, Duffley (1992: 87) points out that *let* forms a tight unit with some infinitives. Such expressions are synonymous with lexical causatives, e. g., *let fall* is similar to *drop* and *let know* is similar to *inform*. Duffley writes, "[i]n these uses, one gets the impression that *let* has been dematerialized to the point of almost being a mere 'actualizer' of the infinitive event" (Duffley 1992: 87). In other words, *let* is less autonomous and more semantically bleached than *allow* and *permit*. This tight integration can be expressed iconically by word order: *let* is sometimes immediately followed by the infinitive, e. g., *let go (of)*.

## 2.5 Summary

The previous studies mention quite a few parameters of variation: semantic, formal, collocational and social. However, none of the studies examines all these parameters in a multivariate statistical analysis. The present study fills in this gap by testing a large number of formal, semantic, morphosyntactic, social and collocational variables in one multivariate model, where the effect of each variable is measured while controlling for the others.

# 3 Data set and predictors

## 3.1 Data set

The data set is a sample from the British National Corpus (XML edition). To create it, we first automatically extracted all forms of *let*, *allow* and *permit* that were followed by another verb within the context window of six words. As will be shown in Section 3.2.1, widening the context window would not yield many additional examples. Examples of adhortative *let* (e. g., *let's go*) were excluded. Since *let* + V and in some cases *allow* + *to* V cannot be used in the passive (**He was let come*; **The bird is allowed to travel long distances by the accumulated fat*) (Mittwoch 1990: 119), only the active forms of the first verb were taken into account. Next, we took all remaining 882 instances of the construction with *permit* and drew random samples of equal size for *let* and *allow*, discarding all spurious hits and replacing them by random examples until we reached the target number of observations. The total size of the data set was 2,646 examples, which were then coded for 23 variables, which are discussed in the following subsections. To speed up the coding process, we annotated the data set syntactically with the help of the Stanford Parser (Klein and Manning 2003) and extracted the information about the main slot fillers of the constructional instances. All automatic annotations were manually checked. The variables are described in Section 3.2.

## 3.2 Variables

### 3.2.1 Linguistic distance

**[Var 1] Linguistic distance.** This variable represents the linguistic distance in words between a permissive verb (V1) and the second predicate (V2).

This distance is a very crude approximation of the spatial distance between the predicates in writing and the temporal distance in spoken communication. According to previous research, distance may be relevant for the variation between the bare and *to*-infinitive (e. g., Fischer 1992b: 336; Rohdenburg 1996). Words were defined as strings of alphabetic or numeric characters separated by white spaces. In the instances of *let* + V, the distance was computed as the number of words between V1 and the bare infinitive. As for the instances of *allow* + *to* V and *permit* + *to* V, the distance was measured as the number of words between V1 and the infinitival particle *to* followed by V2. This way, in both *Let him go* and *Allow him to go* the distance would be one word. The distances were computed automatically for every example with the help of a Python script. Figure 1 represents the frequencies of the number of words between V1 and V2 in the data set. One can see that the majority of examples have only one word between the predicates. The graph also shows that the context window of six words is sufficient and widening it would not yield a substantial number of new observations.

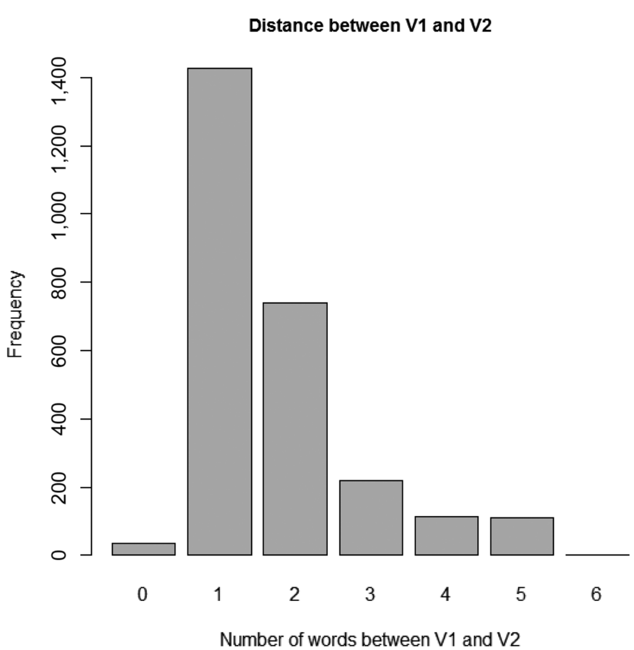


**Figure 1:** A bar plot of distances (in words) between V1 (the permissive verb) and V2 (the infinitive).

### 3.2.2 Conceptual integration of events

It is not easy to operationalize the level of conceptual integration of the causing (permitting) and caused (permitted) events without being circular. We use here two variables that have been mentioned in previous literature and which were relatively easy to code.

**[Var 2] Control of the Permittee**, which shows whether the Permittee has control over the permitted event or not. Inanimate Permittees were automatically considered lacking control. For animate Permittees, a test was applied. If it was possible to say *X let Y do Z, and Y did Z because Y chose to do so*, then the Permittee was considered to have control.[3] In some cases, it was difficult to decide, so the value was left unspecified. Compare the examples in (11), where (11a) contains a Permittee having control over the permitted event, (11b) has a Permittee lacking control, and (11c) has an undefined value:

(11) a. *Of course, improving the facilities ought to allow you to charge higher prices*. (CDF)
b. *Fight to let coma victim die* (K21)
c. *Our aim is to get them out of the chair, cut the chains and let enjoy the freedom of the sea*. (K1X)

This parameter is closely related to the integration of events and (in)directness of causation. Generally, Causees having control are associated with less direct causation and weaker integration of the causing and caused events than non-controlling ones (e. g., Haiman 1983: 784; Levshina 2011).

**[Var 3] Valency of V2**: intransitive (12a), transitive (12b) or passive (12c).

(12) a. *I allowed him to* ***go****.*
b. *I allowed him to* ***destroy*** <u>*the house*</u>*.*
c. *I allowed the house to* ***be destroyed.***

The valency is relevant because it represents the length of the causation chain. In causative constructions with an intransitive V2 the final affected entity is the Causee (here: Permittee). In constructions with a transitive V2 the Causee is only an intermediary, and the final affected entity is the object of V2. Therefore, the

3 A finer-grained classification is also possible, e. g., Divjak (2010: 127) mentions controllable, weakly controllable and non-controllable verbs.

causation is less direct when a V2 is transitive than when it is intransitive (cf. Kemmer and Verhagen 1994). In constructions with a passive infinitive, the intermediary exists, although it is either implied, as in (12c), or expressed by a prepositional phrase with *by*. The intermediary in such constructions can be regarded as even less affected than in contexts like (12b).

This variable was represented by planned orthogonal contrasts.[4] The first contrast allowed us to compare the effects of passive and active forms (both transitive and intransitive), and the second contrast focused on the difference between intransitive and transitive active V2. The reason for that was that passive forms are particularly strongly associated with formal abstract discourse, as in academic papers and official documents (Biber 1988: 112, 151–154), which means that the active – passive distinction may be also relevant for the social dimension of variation (see Section 3.2.4). Consider an example in (13).

(13) *Open-ended standards in legislation confer discretion permitting a wide variety of factors to be taken into account in adjudication*. (EB2)

### 3.2.3 Cognitive distance between the speaker (writer) and the Permitter and Permittee

**[Var 4] The semantic class of the Permitter**. This is a categorical variable with six values that can be represented as the animacy hierarchy (Silverstein 1976), which is also known as the entrenchment hierarchy (Deane 1992) or viewpoint/empathy hierarchy (DeLancey 1981). There exist different versions of such hierarchies, which may also include person and referentiality. In this study we use a purely semantic hierarchy shown in (14), which is based on Langacker (1991: 307).

(14) Speaker > Hearer > Animate > Material (Physical) Object > Abstract

**4** Planned contrasts allow the researcher to perform those comparisons that are the most important theoretically. They are widely used in factorial ANOVA. For example, a three-level categorical variable with categories A, B and C can be coded with the help of two contrasts. The first contrast will compare category A with the average of B and C, and the second will compare category B with C. Contrasts are called orthogonal if the cross-products of their coefficients sum up to zero (e. g., Levshina 2015: 185–186).

The category “Animate” included humans, animals and organizations. Although Langacker (1991: 307) and some others treat animals as a separate class, they were conflated with humans and organizations because of a very low frequency of animals in our data. The lower the Permitter on the hierarchy, the greater the cognitive distance between the speaker or writer, on the one hand, and the Permitter, on the other hand.

In a few cases, the semantic class was unspecified and labelled as “Undefined”. An example of this category is given in (15). It is not clear whether the entity that allows the parties to alter or revoke any third party right is a person or a legal document.

(15) *The position of the parties is preserved by allowing them to alter or revoke any third party right* [...] (EF3)

This variable was coded as a categorical one. The order of the classes in the hierarchy was not taken into account because there exist different versions of these hierarchies. For example, Speaker and Hearer are sometimes conflated into one class of Speech-Act Participants (SAP), as done by DeLancey (1981: 644). For this reason, it is safer to treat the classes as unordered.

**[Var 5] The semantic class of the Permittee**, with the same classes as for the Permitter, with the exception of “Undefined”.

### 3.2.4 Social and communicative distance between the Speech-Act Participants (SAP)

**[Var 6] Channel of communication**: written and spoken. The channel was coded automatically according to the meta-information provided in the corpus. This variable represents spatiotemporal distance between the SAP. In spoken communication, the SAP usually communicate in the same time and space, whereas written communication was developed in order to get one’s message across distance and time.

**[Var 7] Domain of use**. The original domain classification in the BNC is very extensive and differs for spoken and written data. Our classification is a result of merging the original domain classes on the basis of their posterior regression coefficients and overlapping credible intervals (see Section 4). The two resulting large domains, which displayed distinct behaviour with regard to the use of the permissive constructions, were called tentatively “Public” and “Personal”. The former contained written texts with the following BNC domains: social science, world affairs, commerce and finance. It also included the spoken data under the original labels “business” and “public/institutional”. The rest of the texts, which were classified roughly as “Personal”, included educational texts, imaginative prose and other text types.

**[Var 8] Imperative or non-imperative form of V1**. This variable has the values “Imperative” and “Non-imperative”, which subsumes all other tense, aspect and mood categories. The initial classification was more detailed and included finer-grained tense, aspect and mood characteristics of V1 (Present Simple, Past Simple, Perfective, Progressive, Irrealis and non-finite), but these classes did not display much difference with regard to the use of the permissive constructions, so they were conflated in a broad category “Non-imperative” as a result of the above-mentioned analytical procedure.

The imperative *let* has a broad range of functions, which do not express permission proper, but denote what Wierzbicka (2006: 183–202) calls letting of “cooperative dialogue” and “cooperative thinking”. Some of these functions are displayed in (16). One can use *let* to manage discourse and turn-taking (16a), introduce a break in discourse (16b), add a clarification or elaboration (16c), or express an offer (16d). These are not genuine requests for permission, but semantically bleached discourse elements used for politeness purposes. Using such forms implies (spatio)temporal proximity between the interlocutors and their direct contact.

(16) a. *Le let me finish*. (KBB)
b. ‘*Let me see,’ he murmured*. (ANL)
c. *Let me explain*. (FRS)
d. ‘*Sit down, Craig,’ she said softly, ‘let me fetch you some tea’*. (CKD)

**[Var 9] Verbosity, or mean sentence length** in the BNC file where every example from the data set occurred. As Haiman (1983) observed, social distance is iconically correlated with verbosity of communication.

**[Var 10] Length of V2 in characters**. The length varies from 2 (*go*) to 13 characters (*differentiate*). We expect longer verbs to appear in more formal registers.

### 3.2.5 Collostructional measures

The collostructional measures are meant to represent the degree of association between each of the three permissive constructions and the verbs that fill in the V2 slot. Slot fillers are called collexemes in collostructional analysis

(Stefanowitsch and Gries 2003). There exist a plethora of possible measures of association between collexemes and constructions (e. g., Levshina 2015: Ch. 10). For this study, we computed several popular measures that represent different aspects of relationships between a collexeme and a construction. They are based on at least some of the following frequencies, which are shown in Table 1:

**Table 1:** Frequencies required for computation of measures of association between a construction and its collexeme (slot filler).

| | Construction A | Other constructions (not A) |
|---|---|---|
| Collexeme X | *a* | *c* |
| Other collexemes (not X) | *b* | *d* |

a) frequency of collexeme X in construction A
b) frequency of all other collexemes (not X) in construction A
c) frequency of collexeme X in all other constructions (not A). By this we mean all possible uses of X in the corpus minus its frequency in construction A.
d) frequency of all other collexemes (not X) in all other constructions (not A). Following a common practice in collostructional studies of verbal constructions, it is computed as the sum frequency of all verbs in the corpus (with the exception of modals) minus a, b and c.[5]

**[Var 11] Attraction**: the proportion of collexeme X in the total frequency of construction A: a/(a + b). The term was introduced by Schmid (2000).

**[Var 12] Reliance**: the proportion of occurrences of collexeme X in construction A: a/(a + c). The term also comes from Schmid (2000).

**[Var 13] Minimum Sensitivity**: in this context, the minimum score of Attraction and Reliance. This measure, introduced by Pedersen and Bruce (1996), has been found to be the most successful corpus-based predictor of experimental reaction times by Wiechmann (2008).

**5** Since the number of constructions in a language is unknown, the computation of this cell can be regarded as theoretically problematic (cf. Schmid and Küchenhoff 2013). In practice, however, the choice of the offset value does not have much effect on the rankings of collexemes in collostructional analysis (Gries 2012).

**[Var 14] Collostructional strength**: a log-transformed *p*-value based on the Fisher exact test applied to the contingency table in Table 1, which is positive when the frequency of collexeme X in construction A is greater than one can expect by chance alone, and negative when the observed frequency in smaller than the expected frequency. To the best of our knowledge, this measure was first introduced in this form in Stefanowitsch and Gries (2005) and has been used widely in collostructional analyses of different flavours since then.[6]

**[Var 15] Δ*P* with verb as a cue**: it shows the difference between the proportion of the verb in the total uses of the construction and the proportion of the same verb in the other constructions: $a/(a + b) - c/(c + d)$. Originally, it is a psychological cue-response measure introduced by Allan (1980) and applied to constructionist studies by Ellis (2006).

**[Var 16] Δ*P* with construction as a cue**: this measure shows the difference between the proportion of the construction in the total frequency of the verb and the proportion of the same construction in the total frequency of all other verbs: $a/(a + c) - b/(b + d)$.

**[Var 17] Log-odds ratio**: this simple effect size measure for contingency tables is the log-ratio of two odds, which are expressed in a simplified form as a/b and c/d. In order to avoid division by zero in case of the verbs that only occur in a particular permissive construction and nowhere else in the corpus, a very small 0.001 correction was added to each number (*a*, *b*, *c* and *d*).

These measures were computed between the permissive construction with *let*, *allow* or *permit* observed in a given sentence and the corresponding V2.

**6** There has been some controversy around collostructional strength and collostructional analysis recently (Bybee 2010: 97–101; Schmid and Küchenhoff 2013; Küchenhoff and Schmid 2015). One of the controversial issues is the use of *p*-values, which normally serve as a measure of statistical significance (more exactly, the probability of finding the observed test statistic and more extreme values under the assumption of no association), rather than effect size (i. e., the strength of association). However, this peculiarity of collostructional strength in comparison with the other association measures is of little practical importance if the total frequency in Table 1 is kept constant (Gries 2015). On a conceptual level, one should not exclude the possibility that the speaker's certainty about the relationships between a construction and a collexeme may be more or just as relevant as his or her knowledge about the strength of their association. If one treats collostructional strength as a kind of collostructional confidence, the measure can make conceptual sense. Ideally, one should use a Bayesian measure of collostructional certainty, which might tentatively be called collostructional credibility. We leave the development of such a measure for the future.

The verb frequencies were taken from a frequency list of lemmata based on the entire corpus. The constructional frequencies were computed with the help of a Python script, which counted all instances of *let*, *allow* and *permit* with a verbal complement in the syntactically parsed version of the corpus.

Many of these measures are highly intercorrelated (cf. Schmid and Küchenhoff 2013; Levshina 2015: Ch. 10). If we add all of them to the model, the regression coefficients will be unreliable, since many solutions can fit the data fairly well (Kruschke 2011: 555). This problem is known as multicollinearity. This is why it was decided to select one measure that predicts the use of the constructions the best. We fit several Bayesian mixed-effects regression models with all variables of interest for each of these association measures and compared the models. Table 2 displays the Leave-One-Out Information Criterion (LOOIC) and Watanabe-Akaike Information Criterion (WAIC) scores for each of the models, which are popular goodness-of-fit measures in Bayesian regression analysis (cf. Vehtari et al. 2015). The smaller the score, the better the model. Both measures indicate that the model with Minimum Sensitivity fits the best. As an illustration, consider the top twelve most strongly attracted collexemes, which are observed in the pairs *let* + *go, allow* + *(to) escape, let* + *know, permit* + *(to) inspect, allow* + *(to) enter, let* + *pass, let* + *finish, let* + *touch, let* + *forget, allow* + *(to) proceed, let* + *happen* and *let* + *stay*. The majority of top scores belong to *let*, which supports the previous observations about the high degree of attraction between *let* and some infinitives. This variable will be used for subsequent multivariate analyses presented in Section 4.

**Table 2:** LOOIC and WAIC for different collostructional measures.

| Collostructional measure | LOOIC | WAIC |
|---|---|---|
| Attraction | 3,531.0 | 3,526.1 |
| Reliance | 3,848.5 | 3,843.6 |
| Minimum Sensitivity | 3,469.1 | 3,466.6 |
| Collostructional strength | 3,798.7 | 3,796.3 |
| Δ*P* with verb as a cue | 3,782.6 | 3,779.6 |
| Δ*P* with construction as a cue | 3,849.5 | 3,846.1 |
| Log-odds ratio | 3,521.2 | 3,516.2 |

### 3.2.6 Other variables

This section describes the remaining variables, which were not directly related to any of the dimensions described above.

**[Var 18] Polarity**: positive or negative. The latter is operationalized as the presence of negative particles, pronouns or adverbs in the simple clause with the permissive construction, as in (17):

(17) *We are **not** going to let them drive us away*. (CBF)

**[Var 19] Coreferentiality**: the presence or absence of coreferentiality between the Permitter and other participants of the causative situation, as in (18):

(18) a. *But he knew that he had one more duty to perform before **he** allowed **himself** to succumb to his craving for rest*. (EFW)
b. ***We** are not going to let them drive **us** away*. (CBF)

**[Var 20] Possession**: presence of absence of grammatical possession relationship between the Permitter as the possessor and another participant as the possessee, formally marked by the possessive case or a possessive pronoun, as in (19):

(19) *A man **who** allowed **his** dogs to walk around without collar and lead says he'll appeal after being ordered to pay almost nine hundred pounds*. (K1E)

Variables 18–20 were included because they had been shown to play a role in variation of constructions that denote more and less direct causation in other languages (Levshina 2011, Forthcoming.).

**[Var 21] Semantics of V2**: non-mental and mental. The classification of caused events into perceivable (i.e., non-mental) and non-perceivable (i.e., mental) has been shown to play a role in the variation between the bare and *to*-infinitive at an earlier historical stage (Fischer 1995). Examples of mental verbs are verbs of perception (*see*), mental state (*know*) and emotion (*love*). A finer-grained classification like Levin's (1993) would be difficult to apply here because of a large number of classes. Using that classification would create data sparseness and negatively affect the quality of the statistical model.

**[Var 22] Log-transformed frequency of V2.** This variable was included because a first informal inspection of the infinitives suggested that the collexemes of *allow* and especially *permit* included a number of rare verbs.

**[Var 23] *Horror aequi***: the presence of another permissive verb (*let*, *allow*, *permit* or *enable*) in the left context within the same sentence. *Horror aequi* is a tendency to avoid repetition of identical elements (Rohdenburg 2003).[7] The reason for considering this variable is to take into account the choice of a particular permissive construction for stylistic purposes. An example is (20), where the context previous to *permit* contains the verb *allow*:

(20) *The 'Schools tor Mothers' and 'Babies' Welcomes', set up by volunteers from 1908 onwards, set out to teach women to breast feed their children, so as to avoid the problem of contaminated milk and unhygienic feeding bottles; to follow a strict feeding schedule; not to allow dummies; not to use inflammable flannelette clothing and not to* ***permit*** *the infant to sleep with its parents for fear of suffocating*. (GUW)

As recommended by Gelman et al. (2014), all numeric variables were scaled and centered around zero. All categorical variables were represented as sum contrasts.

# 4 Bayesian mixed-effects multinomial model

## 4.1 The main distinctive features of Bayesian statistics

If we want to be able to say something about the language in general based on a sample of observations from a corpus, we need inferential statistics. At present, the most popular type of inferential statistics is frequentist. There exist numerous algorithms for frequentist inference. Bayesian inference is becoming popular only now, but it offers some advantages in comparison with frequentist statistics. While frequentist statistics only allows one to test whether the null hypothesis can be rejected, Bayesian statistics enables one both to test the null hypothesis and to estimate the probability of specific parameter values given the data.

A distinctive feature of Bayesian statistics is the use of so-called priors. These are the prior beliefs in the probability of some parameters before the

7 Rohdenburg (2003) uses this principle to explain why the *to*-infinitive tends to be avoided immediately after a governing *to*-infinitive (e. g., *to try to do*). We have considered the possibility that the presence or absence of *to* before V1 might be a relevant predictor, but the data set contains only four instances of *to let/allow/permit* followed immediately by V2, which is not sufficient for a statistical analysis.

data are taken into account. After the data are taken into account, the model returns the posterior probabilities of specific parameter values. These posterior probabilities depend on both the prior beliefs and the data, whereas the results of a frequentist model depend only on the data. However, in the Bayesian approach, one can also provide non-informative priors, which will result in posteriors that are influenced only by the data, as in frequentist statistics. In this study, we use uniform priors on the intercept and slopes.[8]

Another particular characteristic of this approach is the use of a Markov Chain Monte Carlo (MCMC) algorithm in order to approximate the posterior distribution. The algorithm generates representative random values from this distribution and then estimates the posterior probabilities from those representative values. This walk through the parameter space can be quite time-consuming and requires a powerful computer. In this paper, we employ the No U-Turn Sampler (NUTS), which is an extension of the Hamiltonian Monte Carlo algorithm implemented in Stan.

The model described below was fit with two thousand iterations per chain. Four Markov chains were sampled. We also used a warm-up (or burn-in) period of 1,000 iterations in each chain, that is, we removed the data based on the first 1,000 iterations in order to correct the initial sampling bias.

## 4.2 Model structure and diagnostics

The response variable was the use of the permissive construction with *let*, *allow* or *permit* in a given context. Thus, there are three possible outcomes. Models with multiple categorical outcomes are called multinomial. In our model, *let* was used as the reference level. Therefore, we obtained two sets of coefficients, one for *allow* compared with *let*, and the other for *permit* compared with *let*.

The model is mixed, i. e., it contains both fixed effects (the above-mentioned semantic, social and other variables) and random effects (random intercepts,

---

**8** In fact, we have tried different priors, non-informative (uniform) and weakly informative ones (more exactly, normal distribution[0, 5], Cauchy[0, 5] and Student's *t*-distribution with two degrees of freedom). The results demonstrate that the choice between these priors has no influence on the posteriors probabilities, with the exception of *t*(2), which produces somewhat less extreme values. However, even in that case the difference is very small. Although we are not using them in the final model reported here, weakly informative priors may be helpful when there is complete and quasi-complete separation, which arises when a linear combination of predictors is perfectly predictive of the outcome. The use of such priors makes the coefficient estimates numerically stable (see an example in Gelman et al. 2014).

more exactly), represented by the infinitives that fill in the V2 slot and the BNC files. There were 756 unique verbs that fill in the V2 slot in the data set. Most of them (446 verbs) were *hapax legomena*, but some verbs were highly frequent, e. g., *go* (134 occurrences), *know* (103), *take* (74), *have* (66) and *do* (62). We considered only the infinitives with frequency greater than 5 as individual factor values. All low-frequency infinitives were conflated and coded as “Other”. The total number of unique BNC files was 1,522. Most of them (930) occurred only once. We conflated all files with the frequency of 5 and less in the “Other” category, as well.

The model that is reported below contains all seventeen variables of interest as fixed effects. A more parsimonious model with only those predictors whose 95 % credible intervals do not include zero (see Section 4.3.1) reveals highly similar results. The model does not contain interaction terms. Interactions are observed when the effect of two or more variables on the response is not additive. Particularly important are cross-over interactions, when the effect of variable X is opposite for different values of variable Y, which may undermine the conclusions based on the model with main effects only. A manual check of pairwise possible interactions between the variables has revealed a few interactions, which nuance some findings. However, they do not invalidate the main conclusions made on the basis of the main-effect-only model, and will not be reported due to the lack of space.

It is also important to evaluate the overall predictive power of the model. To do so, we computed a prediction accuracy measure based on a comparison between the mean predicted probabilities of *allow*, *let* and *permit* for every data point and the actual construction that was used in the given context. The probabilities were transformed into categorical choices by choosing the construction with the maximum probability. The predicted and observed outcomes were then cross-tabulated and the number of correct predictions was computed, which was 74 % of the total number of observations in the data set. With the baseline at 33.3 % (i. e., a situation if the three outcomes were assigned randomly), this is a clear improvement.

There are a few specific things that should be taken into account when fitting a Bayesian model. Importantly, there should be no strong autocorrelation between successive draws in Markov chains. The chains should also converge to the posterior distribution, i. e., reach stationarity. Finally, the chains should mix well, that is, they should traverse the posterior distribution quickly. Figure 2 shows trace plots that reveal how well the four Markov chains plotted on top of one another mix and converge for all predictor estimates in the model after the burn-in period. If a trace plot looks like a “fat, hairy caterpillar”, which is not bending in any direction, everything is fine (Lunn et al. 2013: 73–74). Another diagnostic tool is the *R-hat*


**Figure 2:** Mixing and convergence of four Markov chains for the predictor parameters.

statistic, which represents the ratio of between-chain variance to within-chain variance for every parameter (Sorensen et al. in preparation). This statistic should be very close to 1. This was the case in our model. Thus, we can conclude that the Markov chains have converged.

## 4.3 Interpretation of regression analysis results

### 4.3.1 Credible intervals and marginal posterior probabilities of predictors. The effect of linguistic distance

The Stan algorithm returns 4,000 posterior estimates of each regression parameter (1,000 estimates in each of the four chains). These probability distributions can be represented in a histogram, as in Figure 3, which displays the distributions of the effect of linguistic distance between V1 and V2 on the log-transformed odds of *allow* vs. *let* (left) and *permit* vs. *let* (right). The estimates are log-odds ratios (not to be confused with log-odds ratios as a collostructional measure). A positive log-odds ratio shows that the odds of *allow or permit* against *let* increase with the linguistic distance, whereas a negative value means that their odds decrease and the odds of *let* increase. From the posterior distributions one can compute the posterior means, which are displayed as dots

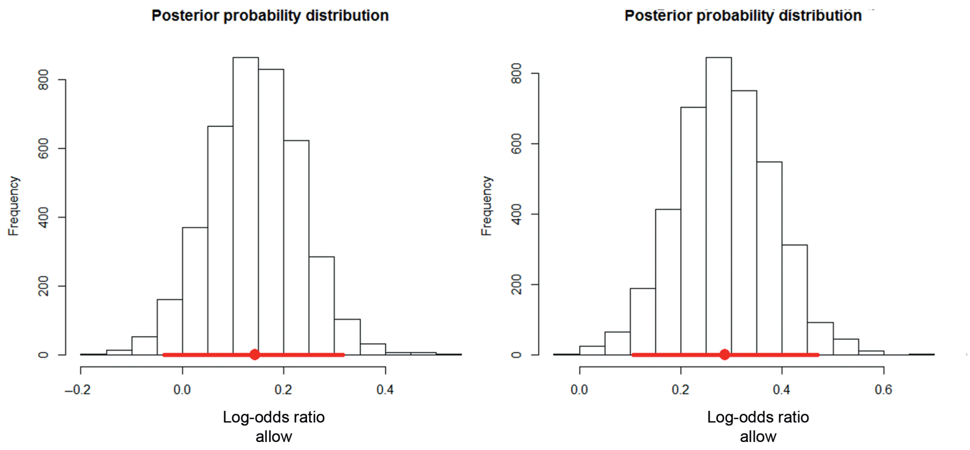


**Figure 3:** Posterior probability distributions of the estimates of linguistic distance for *allow* (left) and *permit* (right). The dots represent the posterior means, and the lines indicate the 95 % credible intervals.

in Figure 3, as well as the 95 % credible intervals, which correspond to the region between the 2.5th and the 97.5th percentile, where the 95 % of the posterior probability density lies.[9]

Unlike frequentist statistics, Bayesian inference does not employ *p*-values for null hypothesis significance testing. Instead, one can use 95 % credible intervals and assume that the effect of a variable is non-zero if the 95 % credible interval does not include zero. One can see in Figure 3 that the 95 % credible interval of *allow* includes zero, and that of *permit* does not. This means that we can be sufficiently confident that the effect is different from zero only in the case of *permit*.

Moreover, one can draw probabilistic inferences from the posterior probability distribution and calculate the posterior probability of a parameter taking a particular range of values, e. g., being positive or negative (cf. Sorensen et al., In preparation). In our case, the probability that the coefficient of *permit* is positive is 99.95 %. This is the proportion of the posterior values that are greater than zero. Only 0.05 % of the values are smaller than zero. In the case of *allow*, the probability that the coefficient is positive is 94.4 %, and that it is negative is 5.6 %. That is, even though the 95 % equal-tailed credible interval contains zero, we still see that there is a high probability of the coefficient being positive. This information enriches the results of a statistical analysis and can be used for generation of future research hypotheses (cf. Vasishth et al. 2013).

The analysis thus reveals that there are on average more words between *permit* and *allow* and the *to*-infinitive than there are between *let* and the bare infinitive. Moreover, the difference between *permit* and *let* is greater than that between *allow* and *let*.

### 4.3.2 Conceptual integration of events

This section describes the results of the regression analysis for the two variables that represented conceptual integration of the permitting and permitted events. The posterior means, 95 % credible intervals and probabilities of the coefficients being above and below zero are shown in Table 3. In the case of categorical variables, zero represents the grand mean (i. e., the mean of means).

---

**9** 95 % credible intervals are conceptually different from 95 % confidence intervals in frequentist statistics. The notion of a credible interval is more intuitive than the notion of a confidence interval, which has produced a lot of misunderstanding. A 95 % credible interval means that the probability of the parameter of interest being within the interval bounds is 95%. A 95 % confidence interval, in contrast, can be seen as the result of a statistical procedure that generates the intervals containing the true value of the parameter 95 % of the time.

**Table 3:** Permittee having control [Var 2] and Valency of V2 [Var 3]: posterior means and other information.

| Parameter | Posterior mean | Lower boundary of 95 % CI | Upper boundary of 95 % CI | *P* (*β* < 0) | *P* (*β* > 0) |
|---|---|---|---|---|---|
| Permittee having control: allow | 0.387 | 0.041 | 0.739 | 1.3 % | 98.7 % |
| Permittee having control: permit | 0.575 | 0.185 | 0.974 | 0.2 % | 99.8 % |
| Transitive V2: allow | 0.116 | −0.06 | 0.284 | 9.3 % | 90.7 % |
| Transitive V2: permit | 0.136 | −0.048 | 0.316 | 7.8 % | 92.2 % |
| Passive V2: allow | 0.392 | 0.181 | 0.614 | 0 % | 100 % |
| Passive V2: permit | 0.461 | 0.24 | 0.699 | 0 % | 100 % |

From the information presented in Table 3 one can conclude that Permittees having control increase the odds of both *allow* and *permit* in comparison with *let*. As for transitive V2, the 95 % credible intervals, which include zero, do not give us certainty that the effect is truly different from zero, but the probabilities of observing a positive effect are larger than 90 % both for *allow* and *permit*. Finally, the odds of both *allow* and *permit* increase when the infinitive is passive. All this means that a greater conceptual distance between the permitting and permitted events increases the odds of *allow* and *permit*. Conversely, a greater integration of events increases the chances of *let*. This is evidence in support of Givón's (1980) binding hierarchy, which also indirectly corroborates the conclusions made by Duffley (1992) and Egan (2008).

#### 4.3.3 Cognitive distance between the speaker (writer) and the Permitter and Permittee

Table 4 displays the posterior means and other information for the semantic classes of the Permitter on the hierarchy of animacy, entrenchment, etc. For abstract Permitters, which are found at the lower end of the hierarchy and which have the weight of −1 in the sum contrasts, the posterior means can be computed from the other coefficients by summing up all other coefficients multiplied by −1. The computed values are 1.76 for *allow* and 2.064 for *permit*. The relative magnitude of the posterior means in Table 4 suggests the following simplified hierarchy, which holds both for *allow* and *permit*:

(21) SAP (Speaker and Hearer) > Animate > Inanimate (Material Objects/ Abstract)

**Table 4:** The position of the Permitter on the animacy hierarchy [Var 4]: posterior means and other information.

| Parameter | Posterior mean | Lower boundary of 95 % CI | Upper boundary of 95 % CI | *P* (*β* < 0) | *P* (*β* > 0) |
|---|---|---|---|---|---|
| Permitter = Speaker: allow | −1.38 | −1.946 | −0.836 | 100 % | 0 % |
| Permitter = Speaker: permit | −1.856 | −2.644 | −1.145 | 100 % | 0 % |
| Permitter = Hearer: allow | −2.641 | −3.562 | −1.818 | 100 % | 0 % |
| Permitter = Hearer: permit | −1.583 | −2.322 | −0.818 | 100 % | 0 % |
| Permitter = Animate: allow | −0.95 | −1.329 | −0.575 | 100 % | 0 % |
| Permitter = Animate: permit | −1.085 | −1.492 | −0.699 | 100 % | 0 % |
| Permitter = Material Object: allow | 2.209 | 1.314 | 3.291 | 0 % | 100 % |
| Permitter = Material Object: permit | 2.043 | 1.036 | 3.164 | 0 % | 100 % |
| Permitter = Undefined: allow | 1.002 | 0.178 | 1.931 | 0.8 % | 99.2 % |
| Permitter = Undefined: permit | 0.417 | −0.514 | 1.42 | 20.1 % | 79.9 % |

The odds of *allow* and *permit* increase as one goes down this hierarchy, and, conversely, the likelihood of *let* increases as ones goes up the hierarchy. The 95 % credible intervals do not contain zero, except for the category "Undefined".

Table 5 presents the results for the semantic classes of the Permittee. The posterior means for abstract Permittees were 0.887 (*allow*) and 0.437 (*permit*). After examining the coefficients, we can conclude that the original hierarchy is partly observed for *allow* (with the exception of the order of Material Objects and Abstract Entities), but not for *permit*, where the Permittee = Speaker has a positive mean. Moreover, some of the 95 % credible intervals of both

**Table 5:** The position of the Permittee on the animacy hierarchy [Var 5]: posterior means and other information.

| Parameter | Posterior mean | Lower boundary of 95 % CI | Upper boundary of 95 % CI | *P* (*β* < 0) | *P* (*β* > 0) |
|---|---|---|---|---|---|
| Permittee = Speaker: allow | −1.481 | −2.081 | −0.908 | 100 % | 0 % |
| Permittee = Speaker: permit | 0.392 | −0.196 | 0.98 | 9.3 % | 90.7 % |
| Permittee = Hearer: allow | −0.311 | −0.918 | 0.259 | 85.4 % | 14.6 % |
| Permittee = Hearer: permit | −1.674 | −2.524 | −0.905 | 99.97 % | 0.03 % |
| Permittee = Animate: allow | −0.063 | −0.423 | 0.298 | 63.4 % | 36.6 % |
| Permittee = Animate: permit | 0.159 | −0.252 | 0.572 | 21.6 % | 78.4 % |
| Permittee = Material Object: allow | 0.967 | 0.441 | 1.523 | 0 % | 100 % |
| Permittee = Material Object: permit | 0.686 | 0.076 | 1.316 | 1.3 % | 98.7 % |

*allow* (Permittee = Hearer and Permittee = Animate Being) and *permit* (Permittee = Speaker and Permittee = Animate Being) include zero. Thus, the evidence of the hierarchy is much weaker with Permittees than with Permitters.

### 4.3.4 Social and communicative distance between speech-act participants

The results in Table 6 suggest that both *allow* and *permit* are preferred in written texts, in communication on "public" topics (i. e., on social, political and business issues), in verbose style with longer sentences and in combination with longer infinitives. In addition, *permit* is dispreferred in imperative contexts, but *allow* does not differ much from *let* in that respect. Also recall that the odds of *permit* and *allow* increase when the infinitive is passive (see Section 4.3.2). From this we can conclude that *allow* and especially *permit*, which has more extreme estimates, are more associated with situations that involve a social and communicative distance between the SAP than *let*.

**Table 6:** Channel [Var 6], domain [Var 7], imperative V1 [Var 8], mean sentence length [Var 9] and length of V2 [Var 10]: posterior means and other information.

| Parameter | Posterior mean | Lower boundary of 95 % CI | Upper boundary of 95 % CI | $P(\beta < 0)$ | $P(\beta > 0)$ |
|---|---|---|---|---|---|
| Written channel: allow | 0.285 | 0.004 | 0.569 | 2.4 % | 97.6 % |
| Written channel: permit | 1.023 | 0.556 | 1.521 | 0 % | 100 % |
| Public domain: allow | 0.369 | 0.191 | 0.549 | 0 % | 100 % |
| Public domain: permit | 0.587 | 0.39 | 0.784 | 0 % | 100 % |
| Imperative V1: allow | 0.027 | −0.492 | 0.576 | 47.1 % | 52.9 % |
| Imperative V1: permit | −0.962 | −1.48 | −0.44 | 99.98 % | 0.02 % |
| Mean sentence length: allow | 0.376 | 0.19 | 0.563 | 0 % | 100 % |
| Mean sentence length: permit | 0.57 | 0.37 | 0.777 | 0 % | 100 % |
| V2 length: allow | 0.535 | 0.3 | 0.769 | 0 % | 100 % |
| V2 length: permit | 0.544 | 0.312 | 0.791 | 0 % | 100 % |

### 4.3.5 Collostructional fixation

The results for the collostructional variable Minimum Sensitivity are displayed in Table 7. They demonstrate that greater collocational fixation decreases the odds of *allow* and *permit* against *let*. For *permit*, the effect is particularly strong. This means that the construction with *let* has the strongest association

**Table 7:** Minimum Sensitivity [Var 13]: posterior means and other information.

| Parameter | Posterior mean | Lower boundary of 95 % CI | Upper boundary of 95 % CI | *P* (*β* < 0) | *P* (*β* > 0) |
|---|---|---|---|---|---|
| Minimum Sensitivity: allow | −0.648 | −0.858 | −0.438 | 100 % | 0 % |
| Minimum Sensitivity: permit | −2.734 | −3.1 | −2.365 | 100 % | 0 % |

with its V2 slot fillers, followed by *allow* and then by *permit*. This is in accordance with Duffley's remarks about the high degree of association between *let* and some infinitives.

### 4.3.6 Other variables

This section discusses the remaining six variables. Their estimates are shown in Table 8. All 95 % credible intervals contain zero. However, some of the parameters have very high probabilities of being less or greater than zero. In particular, *permit* tends to be more preferred in negative contexts, without possession markers and with low-frequency V2 than *let*, whereas *allow* is more likely to be used in coreferential contexts.

**Table 8:** Polarity [Var 18], coreferentiality [Var 19], possession [Var 20], semantic class of V2 [Var 21], log-transformed frequency of V2 [Var 22] and *horror aequi* [Var 23]: posterior means and other information.

| Parameter | Posterior mean | Lower boundary of 95 % CI | Upper boundary of 95 % CI | *P* (*β* < 0) | *P* (*β* > 0) |
|---|---|---|---|---|---|
| Negative polarity: allow | −0.035 | −0.231 | 0.16 | 63.5 % | 36.5 % |
| Negative polarity: permit | 0.207 | −0.008 | 0.425 | 2.9 % | 97.1 % |
| Coreferentiality: allow | 0.189 | −0.05 | 0.432 | 6.2 % | 93.8 % |
| Coreferentiality: permit | 0.009 | −0.275 | 0.293 | 48.2 % | 51.8 % |
| Possession: allow | 0.046 | −0.172 | 0.274 | 34.9 % | 64.1 % |
| Possession: permit | −0.276 | −0.551 | 0.005 | 97.3 % | 2.7 % |
| Non-mental V2: allow | 0.139 | −0.273 | 0.531 | 25.7 % | 74.3 % |
| Non-mental V2: permit | 0.285 | −0.175 | 0.747 | 11.5 % | 88.5 % |
| V2 log-frequency: allow | 0.038 | −0.214 | 0.298 | 38.7 % | 61.3 % |
| V2 log-frequency: permit | −0.233 | −0.503 | 0.048 | 95 % | 5 % |
| *Horror aequi*: allow | −0.225 | −0.948 | 0.578 | 72 % | 18 % |
| *Horror aequi*: permit | 0.322 | −0.365 | 1.107 | 19.6 % | 80.4 % |

## 5 Summary and discussion

The Bayesian regression analysis has revealed substantial differences between the constructions. The previous observations about the semantic, sociolinguistic and collostructional differences between the constructions have been largely confirmed, and new differences have been found.

The main result of the statistical analyses is a remarkable alignment of several literal and metaphoric distances in the sense that smaller distance increases the odds of *let* and greater distance increases the odds of *allow* and *permit*.

A. **Linguistic distance** between V1 and V2. The more words there are between the predicates, the higher the chances of *allow* and *permit*. This distance does not take into account the difference between the infinitives taken by the permissive verbs, i. e., the bare infinitive taken by *let* and the *to*-infinitive taken by *allow* and *permit*. The use of the particle *to* accounts thus for even greater linguistic distance between the predicates.

B. **Conceptual distance** between the permitting and permitted events expressed by V1 and V2. This distance is captured by the presence of the autonomous Permittee who has control over the permitted event. This feature tends to increase the odds of both *allow* (with only borderline certainty that the effect is non-zero) and *permit* against *let*. In addition, longer causal chains with an intermediary, which are expressed by constructions with transitive active and passive V2, tend to be more associated with *allow* and *permit*.

C. **Cognitive distance** between the speaker and the Permitter and (only in the case of *allow*) Permittee on the animacy hierarchy, which has also been interpreted as the hierarchy of entrenchment, viewpoint or empathy. In general, the further from the speaker the participants are on this hierarchy, the higher the odds of *allow* and *permit* against *let*.

D. **Communicative** and **social distance** between the interlocutors. The odds of *allow* and *permit* increase when communication is written, covers public topics (namely, business, economy and politics), and does not involve immediate interaction (as in the case of the imperative *let*). This distance is also mirrored in the length of the infinitives and mean sentence lengths as indicators of verbosity and formality.

E. **Collostructional distance**, i. e., loose association between V1 and V2, as the inverse of collostructional fixation expressed by Minimum Sensitivity. The looser the association, the higher the chances of *allow* and *permit*.

Notably, *permit* tends to have more extreme posterior means across these dimensions than *allow*, with the exception of the variables related to cognitive distance between the speaker (writer) and the participants. This means that *permit* more than *allow* differs from *let* with regard to most dimensions.

Thus, we have found a remarkable alignment of different kinds of literal and figurative distances (or conversely, proximities). A crucial question is how to explain these results. The alignment of the social and cognitive dimensions is easy to explain. According to Biber (1988) and other studies of register variation, informal involved spoken communication is characterized by frequent reference to Speech-Act Participants (*I* and *you*) and therefore smaller cognitive distance between the speaker and the referents. More direct letting with closer conceptual integration of the events may also be more typical and salient in informal communication than less direct letting with loosely integrated events. Finally, the level of formality and verbosity may explain the differences in the linguistic distance between V1 and V2.

The formal variation of the constructions (i. e., the presence or absence of the particle *to*) and its correlation with the other dimensions require a more detailed discussion. There are at least three cognitive and functional principles that can be useful for that purpose. One of them is the principle of iconicity, more precisely, iconicity of cohesion (Haiman 1983; Haspelmath 2008). Haspelmath formulates this principle succinctly as follows: "Meanings that belong together more closely semantically are expressed by more cohesive forms" (Haspelmath 2008: 2).

This can explain the fact that *allow* and *permit*, which are followed by the *to*-infinitive, are more preferred in the situations of less direct permissive causation than *let*, which takes the bare infinitive. Moreover, iconicity can also explain the correspondence between social distance and length of linguistic forms (Haiman 1983: 801), which is expressed in this study by the length of V2 and the type of the infinitive (bare or with the particle *to*).

The second possible explanation involves the principle of economy (Haiman 1983; Du Bois 1985; Haspelmath 2008). According to this principle, the more frequent constructions should also be more compact. Indeed, *let* + V is the most frequent construction, and *permit* + *to* V is the least frequent one, as shown in Table 9. These frequencies were automatically extracted with the help of the syntactic annotation. The difference is not that striking in the entire BNC, but when we look only at the informal conversations, the frequency contrast between *let* + V and the two other constructions becomes much more obvious. Since we believe that everyday spoken communication has a greater impact on language change than written texts, these frequencies may be more telling. According to the principle of economy, *let* + V as the most frequent construction should also be more formally compact than the other ones, which is exactly the case.

**Table 9:** Frequencies of permissive constructions in the BNC (only active forms of V1).

| | *let* + *V* | *allow* + *V* | *permit* + *V* |
|---|---|---|---|
| Entire BNC | 10,717 | 9,358 | 882 |
| Informal conversations only | 125 | 1 | 0 |

However, the reducing effect of frequency is usually found at the level of absolute rather than relative frequencies (cf. Croft [2008], who uses this as an argument against Haspelmath [2008]). Still, one can extend the principle of economy to near-synonymous constructions if one applies Horn's idea of pragmatic division of labour (Horn 1984) expressed in his well-known Q and R Principles ("Say as much as you can" and "Say no more than you must", respectively). These principles, in their turn, are based on Zipf's (1949: 19–23) Principle of Least Effort, which manifests itself in two opposing forces of the speaker's and auditor's economy. The Q Principle determines that more complex and/or prolix linguistic expressions receive less probable or salient interpretations, whereas the R principle is responsible for the association between less complex expressions and more typical interpretations. These form-meaning correspondences may become conventionalized, so that the less complex constructional schema is associated with more typical meanings, and the more complex one with more marginal ones. The result is an efficient form-meaning mapping, characterized by an equilibrium between the Q and R principles, as well as between the speaker's and hearer's efforts.

Finally, the third candidate is the principle of cognitive complexity: "In the case of more or less explicit grammatical options the more explicit one(s) will tend to be favoured in cognitively more complex environments" (Rohdenburg 1996: 151). The more words there are between V1 and V2, the more difficult it is to recognize V2 as part of the construction. This can explain why *allow* and *permit*, which are associated with greater linguistic distance between V1 and V2, are used with the *to*-infinitive.

All three theories provide plausible explanations of the formal differences between the constructions. A totally new perspective opens up, however, when one takes into account historical evidence. The construction with *let*, which is related to the highly similar Germanic constructions (the German construction *lassen* + Vinf, Dutch *laten* + Vinf, Swedish *låta* + Vinf, etc.), already existed in Old English with the meaning "cause, allow X to do something" (Fischer 1992a). In Middle English, it became highly grammaticalized. In contrast, the verbs *allow* and *permit* were borrowed from French or directly from Latin only in Late Middle

English. At first, they were used only with nominal arguments, e. g., in the ditransitive construction, but gradually began to appear with infinitival complements, as well. According to the Helsinki Corpus of English Texts (1991), the first instances of these constructions are found only in the late sixteenth – early seventeenth centuries.

Throughout all these periods, the permissive *let* was, with only a few exceptions, used with the bare infinitive (Fischer 1992a), which was also the default type of the infinitive in Old English. In Middle English, the *to*-infinitive became more widespread, both quantitatively and qualitatively, whereas the bare infinitive was left only in a few specific constructions (Los 2005). The construction with *let* and the bare infinitive is one of these constructions, which may have been preserved due to its high frequency, similar to modal verb constructions (Bybee 1985). In present-day English, only a closed class of auxiliaries and semi-auxiliaries can be used with the bare infinitive.

Does this mean that the alignment of form and function is simply a more or less random result of the historical process (cf. Cristofaro 2012)? This conclusion seems premature. First, the speakers' use and acquisition of the constructions may well be based on principles that are different from the processes that have led to the emergence of the constructions in their present form (Cristofaro 2012: 665), so that the cognitive principles of iconicity, economy and complexity may still play a role. Next, language change, and grammaticalization in particular, is determined by cognitive processes, which involve iconicity, economy and other principles (e. g., Bybee 2003; Fischer 2008). Under certain conditions, the status of V1 can change over time towards auxiliary. It is not excluded, in principle, that *allow* and *permit* can become quasi-auxiliaries in the future and therefore lose *to*, similar to what is currently happening to the verb *help* (Mair 2008: 135–140). Finally, the diachronic explanation alone cannot account for the overwhelming cross-linguistic evidence of iconic and economic relationships in language, e. g., the form-meaning parallelism in the binding hierarchy (Givón 1980 and later work) or the form-frequency correlation in the coding of the causal-noncausal alternation (Haspelmath et al. 2014). All these correspondences are highly unlikely to be a mere coincidence.

More research is needed in order to test whether the suggested factors (and which of them) have a causal effect on the development and use of the English permissive constructions. One should not exclude the possibility that all these factors operate together (e. g., Croft 2008: 55–56). In any event, the process of finding an explanation should involve multiple factors, in the same way it has become the standard for descriptive models of language variation in Cognitive Linguistics and sociolinguistics. We also believe that finding new examples and counter-examples of such alignment (or lack thereof) between the dimensions in different constructions from typologically diverse languages and language varieties will help us explain the remarkable alignment in the English permissive construction and provide new insights about the relationships between language, social communication and cognition.

**Acknowledgments:** The author would like to thank Associate Editor Dagmar Divjak for her excellent observations and wise suggestions, Shravan Vasishth for his invaluable advice on the philosophy and practice of Bayesian statistics, and two anonymous reviewers for their insightful comments. Most work reported here was performed as part of a project funded by a grant from the Belgian research foundation F.R.S. – FNRS. All usual disclaimers apply.

## References

Allan, Lorraine G. 1980. A note on measurement of contingency between two binary variables in judgment tasks. *Bulletin of the Psychonomic Society* 15. 147–149.

Arppe, Antti. 2008. *Univariate, bivariate and multivariate methods in corpus-based lexicography – A study of synonymy*. Helsinki: University of Helsinki dissertation.

Bates, Douglas, Reinhold Kliegl, Shravan Vasishth & Harald Baayen. 2015. Parsimonious mixed models. arXiv:1506.04967 [stat]. http://arxiv.org/abs/1506.04967 (accessed 31 October 2015).

Biber, Douglas. 1988. *Variation across speech and writing*. Cambridge: Cambridge University Press.

British National Corpus, version 3 (BNC XML Edition). 2007. Distributed by Oxford University Computing Services on behalf of the BNC Consortium. http://www.natcorp.ox.ac.uk/

Bybee, Joan L. 1985. *Morphology: A study of the relation between meaning and form*. Amsterdam & Philadelphia: John Benjamins.

Bybee, Joan L. 2003. Cognitive processes in grammaticalization. In Michael Thomasello (ed.), *The new psychology of language. Vol. II*, 145–167. Mahwah, NJ: Lawrence Erlbaum.

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

Carpenter, Bob, Andrew Gelman, Matt Hoffman, Daniel Lee, Ben Goodrich, Michael Betancourt, Marcus Brubaker, Jiqiang Guo, Peter Li &Allen Riddell. In press. Stan: A probabilistic programming language. *Journal of Statistical Software*.

Cristofaro, Sonia. 2012. Cognitive explanations, distributional evidence, and diachrony. *Studies in Language* 36(3). 645–670.

Croft, William. 2008. On iconicity of distance. *Cognitive Linguistics* 19(1). 49–57.

Croft, William. 2009. Toward a social cognitive linguistics. In Vyvyan Evans & Stéphanie Pourcel (eds.), *New directions in cognitive linguistics*, 395–420. Amsterdam & Philadelphia: John Benjamins.

Deane, Paul D. 1992. *Grammar in mind and brain*. Berlin & New York: Mouton de Gruyter.

DeLancey, Scott. 1981. An interpretation of split ergativity and related patterns. *Language* 57(3). 626–657.

Divjak, Dagmar. 2010. *Structuring the Lexicon: A clustered model for near-synonymy*. Berlin: De Gruyter Mouton.

Du Bois, John. 1985. Competing motivations. In John Haiman (ed.), *Iconicity in syntax*, 343–365. Amsterdam & Philadelphia: John Benjamins.
Duffley, Patrick J. 1992. *The English infinitive*. London &New York: Longman.
Egan, Thomas. 2008. *Non-finite complementation: A usage-based study of infinitive and -ing clauses in English*. Amsterdam & New York: Rodopi.
Ellis, Nick. 2006. Language acquisition as rational contingency learning. *Applied Linguistics* 27(1). 1–24.
Fischer, Olga. 1992a. Syntactic change and borrowing: The case of the accusative-and-infinitive construction in English. In Marinel Gerritsen & Dieter Stein (eds.), *Internal and external factors in syntactic change*, 17–88. Berlin: Mouton de Gruyter.
Fischer, Olga. 1992b. Syntax. In Norman Blake (ed.), *The Cambridge history of the English language*, Volume II: 1066–1476, 207–408. Cambridge: Cambridge University Press.
Fischer, Olga. 1995. The distinction between bare and *to*-infinitival complements in late Middle English. *Diachronica* 12. 1–30.
Fischer, Olga. 2008. On analogy and the motivation for grammaticalization. *Studies in Language* 32(2). 336–382.
Geeraerts, Dirk, Gitte Kristiansen & Yves Peirsman (eds.). 2010. *Advances in cognitive sociolinguistics*. Berlin & New York: Mouton de Gruyter.
Gelman, Andrew, John B. Carlin, Hal S. Stern, David B. Dunson, Aki Vehtari & Donald B. Rubin. 2014. *Bayesian Data Analysis*, 3rd edn. Boca Raton, FL: CRC Press.
Givón, Talmy. 1980. The binding hierarchy and the typology of complements. *Studies in Language* 4(3). 333–377.
Gries, Stefan Th. 2003. *Multifactorial analysis in corpus linguistics: A study of particle placement*. New York & London: Continuum.
Gries, Stefan Th. 2012. Frequencies, probabilities, association measures in usage-/exemplar-based linguistics: Some necessary clarifications. *Studies in Language* 36(3). 477–510.
Gries, Stefan Th. 2015. More (old and new) misunderstandings of collostructional analysis: On Schmid & Küchenhoff. *Cognitive Linguistics* 26(3). 505–536.
Haiman, John. 1983. Iconic and economic motivation. *Language* 59(4). 781–819.
Han, Weifeng, Antti Arppe & John Newman. In press. Topic marking in a Shanghainese corpus: From observation to prediction. *Corpus Linguistics and Linguistic Theory*.
Haspelmath, Martin. 2008. Frequency vs. iconicity in explaining grammatical asymmetries. *Cognitive Linguistics* 19(1). 1–33.
Haspelmath, Martin, Andreea Calude, Michael Spagnol, Heiko Narrog & Elif Bamyaci. 2014. Coding causal–noncausal verb alternations: A form–frequency correspondence explanation. *Journal of Linguistics* 50. 587–625. DOI 10.1017/S0022226714000255.
Helsinki Corpus of English Texts. 1991. Department of Modern Languages, University of Helsinki. Compiled by Matti Rissanen (Project leader), Merja Kytö (Project secretary); Leena Kahlas-Tarkka, Matti Kilpiö (Old English); Saara Nevanlinna, Irma Taavitsainen (Middle English); Terttu Nevalainen, Helena Raumolin-Brunberg (Early Modern English)
Heylen, Kris. 2005. A quantitative corpus study of German word order variation. In Stephan Kepser & Marga Reis (eds.), *Linguistic evidence: Empirical, theoretical and computational perspectives*, 241–264. Berlin & New York: Mouton de Gruyter.
Horn, Laurence R. 1984. Toward a new taxonomy for pragmatic inference: Q-based and R-based implicature. In Deborah Schiffrin (ed.), *Meaning, form, and use in context: Linguistic applications*, 11–42. Washington, DC: Georgetown University Press.

Karttunen, Lauri. 1971. Implicative verbs. *Language* 47(2). 340–358.
Kemmer, Susanne & Arie Verhagen. 1994. The grammar of causatives and the conceptual structure of events. *Cognitive Linguistics* 5. 115–156.
Klein, Dan & Christopher D. Manning. 2003. Accurate unlexicalized parsing. In *Proceedings of the 41th Annual Meeting of the Association for Computational Linguistics*.
Kristiansen, Gitte & René Dirven (eds.). 2008. *Cognitive sociolinguistics: Language variation, cultural models and social systems*. Berlin & New York: Mouton de Gruyter.
Kruschke, John K. 2011. *Doing Bayesian data analysis: A tutorial with R and BUGS*. Oxford: Elsevier.
Küchenhoff, Helmut & Hans-Jörg Schmid. 2015. Reply to "More (old and new) misunderstandings of collostructional analysis: On Schmid & Küchenhoff" by Stefan Th. Gries. *Cognitive Linguistics* 26(3). 537–547.
Langacker, Ronald W. 1991. *Foundations of cognitive grammar, Vol. II, Descriptive application*. Stanford, CA: Stanford University Press.
LDCE Longman Dictionary of Contemporary English: New Edition (2003). Harlow: Longman.
Leech, Geoffrey & Jan Svartvik. 1994. *A communicative grammar of English*, 2nd edn. London: Longman.
Levin, Beth. 1993. *English verb classes and alternations: A preliminary investigation*. Chicago: University of Chicago Press.
Levshina, Natalia. 2011. *Doe wat je niet laten kan* [Do what you cannot let]: *A usage-based study of Dutch causative constructions*. Leuven: University of Leuven dissertation.
Levshina, Natalia. 2015. *How to do Linguistics with R: Data exploration and statistical analysis*. Amsterdam & Philadelphia: John Benjamins.
Levshina, Natalia. Forthcoming. Why we need a token-based typology: A corpus-based study of analytic and lexical causatives in fifteen European languages. *Folia Linguistica*.
Los, Bettelou. 2005. *The rise of the to-infinitive*. Oxford: Oxford University Press.
Lunn, David, Christopher Jackson, Nicky Best, Andrew Thomas & David Spiegelhalter. 2013. *The BUGS book: A practical introduction to Bayesian analysis*. Boca Raton, FL: CRC Press.
Mair, Christian. 2008. *Twentieth-century English: History, variation and standardization*. Cambridge: Cambridge University Press.
Miglio, Viola G., Stefan Th. Gries, Michael J. Harris, Eva M. Wheeler & Raquel Santana-Paixão. 2013. Spanish *lo(s)-le(s)* clitic alternations in psych verbs: A multifactorial corpus-based analysis. In Jennifer C. Amaro, Gillian Lord, Ana de Prada Pérez & Jessi E. Aaron (eds.), *Selected Proceedings of the 16th Hispanic Linguistics Symposium*, 268–278. Somerville, MA: Cascadilla Proceedings Project.
Mittwoch, Anita. 1990. On the distribution of bare infinitive complements in English. *Journal of Linguistics* 26. 103–131.
Nedjalkov, Vladimir P. 1976. *Kausativkonstruktionen*. Tübingen: TBL.
Pedersen, Ted & Rebecca Bruce. 1996. What to infer from a description. Technical Report 96-CSE-04, Southern Methodist University, Dallas, TX.
R Core Team. 2015. *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. http://www.R-project.org/ (accessed 23 November 2015)
Rohdenburg, Günther. 1996. Cognitive complexity and increased grammatical explicitness in English. *Cognitive Linguistics* 7(2). 149–182.

Rohdenburg, Günther. 2003. *Horror aequi* and cognitive complexity as factors determining the use of interrogative clause linkers. In Günther Rohdenburg & Britta Mondorf (eds.), *Determinants of grammatical variation in English*, 205–250. Berlin & New York: Mouton de Gruyter.

Schmid, Hans-Jörg. 2000. *English abstract nouns as conceptual shells. From corpus to cognition*. Berlin & New York: Mouton de Gruyter.

Schmid, Hans-Jörg. 2014. Lexico-grammatical patterns, pragmatic associations and discourse frequency. In Thomas Herbst, Hans-Jörg Schmid & Susen Faulhaber (eds.), *Constructions, collocations, patterns*, 239–295. Berlin & Boston: De Gruyter Mouton.

Schmid, Hans-Jörg & Helmut Küchenhoff. 2013. Collostructional analysis and other ways of measuring lexicogrammatical attraction: Theoretical premises, practical problems and cognitive underpinnings. *Cognitive Linguistics* 24(3). 531–577.

Silverstein, Michael. 1976. Hierarchy of features and ergativity. In R. M. W. Dixon (ed.), *Grammatical categories in Australian languages*, 112–171. Canberra: Australian National University.

Sorensen, Tanner, Sven Hohenstein & Shravan Vasishth. In preparation. Bayesian Linear Mixed Models using Stan: A tutorial for psychologists, linguists, and cognitive scientists. http://www.ling.uni-potsdam.de/~vasishth/statistics/BayesLMMs.html (accessed 22 November 2015).

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

Stefanowitsch, Anatol & Stefan Th. Gries. 2005. Covarying collexemes. *Corpus Linguistics and Linguistic Theory* 1(1). 1–43.

Tagliamonte, Sali A., & R. Harald Baayen. 2012. Models, forests and trees of York English: Was/were variation as a case study for statistical practice. *Language Variation and Change* 24(2). 135–178.

Talmy, Leonard. 2000. *Toward a cognitive semantics*. Cambridge, MA: MIT Press.

Vasishth, Shravan, Zhong Chen, Qiang Li & Guelian Guo. 2013. Processing Chinese relative clauses: Evidence for the subject-relative advantage. *PLoS ONE* 8(10). 1–14. http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0077006 (accessed 21 November 2015).

Vehtari, Aki, Andrew Gelman & Jonah Gabry. 2015. Efficient implementation of leave-one-out cross-validation and WAIC for evaluating fitted Bayesian models. arXiv preprint arXiv:1507.04544.

Wiechmann, Daniel. 2008. On the computation of collostruction strength. *Corpus Linguistics and Linguistic Theory* 4(2). 253–290.

Wierzbicka, Anna. 2006. *English: Meaning and culture*. Oxford: Oxford University Press.

Zipf, George K. 1949. *Human behavior and the principle of least effort*. Cambridge, MA: Addison-Wesley Press.