Communicative efficiency and probabilistic grammar: Bayesian mixed-effect regression models of *help* + (*to*) Infinitive in varieties of web-based English

Abstract

This study focuses on *help* followed by the bare or *to*-infinitive in seven varieties of web-based English from Australia, Ghana, Great Britain, Hong Kong, India, Jamaica and the USA. It investigates the role of information content, which depends on the predictability of the infinitive given *help* and the other way round, in the user's choice between the constructional variants, in addition to various other factors known from the literature, such as register, minimization of cognitive complexity and avoidance of identity (*horror aequi*). The probabilistic constraints are tested in a series of mixed-effects Bayesian logistic regression models. The results indicate that the *to*-infinitive is particularly well represented in contexts with high information content. More specifically, if the expectation that a given infinitive is used with *help*, and not in another construction, is low, there are greater chances that the speaker will prefer the marked form of the infinitive. This tendency, which can be interpreted as communicatively efficient behaviour, is observed in all seven varieties.

Keywords: communicative efficiency, Bayesian regression, *help*, information density, complexity, *horror aequi*

1. Introduction

This paper focuses on assistive causation constructions, which consist of the verb *help* followed by the infinitive with or without *to*, as in (1):

(1) a. *Mary helped John to cook the dinner.*

b. *Mary helped John cook the dinner.*

The construction *help* + (*to*) Infinitive is a rare case when this choice is possible in Present-Day English. Different factors have been proposed to explain the difference between the constructions. Some of them are related to the universal functional principles of iconicity, minimization of cognitive complexity and avoidance of identity (also known as *horror aequi*). Other factors include register, morphological form and the presence or absence of the Helpee. Lohmann's (2011) quantitative study of *help* in British English showed that the variation is multifactorial and probabilistic.

The present paper takes a new direction in this discussion and investigates to what extent the use or omission of *to* before the infinitive can be explained by the speaker's tendency to present more predictable information by less coding material and shorter forms, and less predictable information by more coding material and longer forms (e.g. Levy & Jaeger 2007; Jaeger 2010). This tendency can be seen as a manifestation of the speaker's bias towards efficient, or economical communication (e.g. Haiman 1983). Importantly, the effect of predictability is tested when the other relevant factors, which are known from previous research, are controlled for.

The present paper investigates the role of all these factors in the use of the bare and *to*-infinitive after *help* in seven varieties of online English from Australia, Ghana, Great Britain, Hong Kong, India, Jamaica and the USA, using the data from the GloWbE corpus (Davies 2013). One of the central questions of this study is whether the relationships of predictability between *help* and the infinitive have a similar effect in all varieties or there is substantial cross-lectal variation.

Methodologically, this paper employs Bayesian regression analysis, which is still novel in linguistics (but see one of the first attempts in Author xxxx). Bayesian regression allows the researcher to test directly the research hypothesis: namely, that the variable representing predictability or any other explanatory factor has an effect on the presence or absence of the particle *to*. Bayesian regression is a perfect match for probabilistic grammar because it returns the probability of a variable having an expected effect on the outcome. Such probabilities can be easily compared cross-lectally. They also allow us to study a continuum of credibility without forcing us to make binary decisions based on *p*-values. Moreover, one can quantify the (dis)similarity between variety-specific models by measuring the degree of overlap between the posterior distributions of regression estimates.

The paper is organized as follows. Section 2 introduces the main factors that have been discussed in previous research. In Section 3, I focus on communicative efficiency and the information-theoretic measures used in the present study. Section 4 describes the data source and the process of data extraction. In Section 5 one can find the variables that are tested in this study. Section 6 introduces the method (Bayesian logistic regression) and reports the results of the quantitative analyses. Finally, a discussion of the findings is offered in Section 7.

2 Previous research: universal functional principles and more

2.1 Principle of iconicity

Iconicity is the correspondence between linguistic form and function. There exist many types of iconic relationships at all levels of language structure, from phonology and orthography to morphology and syntax. For our case study, the most relevant type of iconicity is the correspondence between formal and conceptual distance. As formulated by Haiman (1983: 782), "[t]he linguistic distance between expressions corresponds to the conceptual distance between them." With regard to *help* + (*to*) Infinitive, one can say that the formal distance between *help* and the infinitive is greater when the latter is preceded by the particle *to*. In addition, iconicity of independence or autonomy may also be relevant (cf. Bybee 1985): the events that are more integrated conceptually are also more integrated formally. As for *help*, the formal integration is weaker in the case of the *to*-infinitive, which occurs in a wide range of constructions, and stronger in the case of the bare infinitive, which is very restricted and occurs primarily as a complement to auxiliary and modal verbs and with supportive *do* (Huddleston & Pullum 2002: 1174).

On the semantic side, conceptual proximity or dependence is more difficult to define. It can mean a number of different things, for example, spatio-temporal integration of the events, the degree of control and agentivity of the participants, etc. (Givón 1990: Section 13.2). With regard to *help*, it has been proposed that the variant with the bare infinitive designates a more active involvement of the Helper in carrying out the event expressed by the infinitival complement (Dixon 1991: 199). Consider the following examples:

(2) a. *John helped Mary eat the pudding* (he ate half).

b. *John helped Mary to eat the pudding* (by guiding the spoon to her mouth, since she was still an invalid). (Dixon 1991: 199)

When *to* is omitted, as in (2a), the sentence is likely to describe a cooperative effort where Mary and John eat the pudding together; when *to* is included, as in (2b), the sentence means that John acts as a facilitator for Mary, who actually ate the pudding herself (Dixon 1991: 199; 230). Similarly, Duffley (1992: Section 2.3) suggests that the use of the *to*-infinitive evokes help as a condition that enables the Helpee to bring about the event denoted by the infinitive. Yet, many researchers have questioned this interpretation: there are numerous contexts and examples where this distinction cannot be traced (e.g. Huddleston & Pullum 2002: 1244).

It has also been suggested that animate Helpers have a potentially greater involvement in the event (Lind 1983). Indeed, Lohmann (2011) finds that animate Helpers have higher odds of the bare infinitive than inanimate Helpers, which can be regarded as evidence in support of the iconicity principle.

2.2 Principle of (minimization of) cognitive complexity

This principle says, "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 *help* and the infinitive, the more difficult it is to recognize the latter as part of the construction. Consider an example of a complex environment in (3). The variant with the bare infinitive in (3b) is barely acceptable.

(3) a. *I helped them as well as I could to cook the dinner.*

b. *?? I helped (them) as well as I could wash up.* (Rohdenburg 1996: 159).

Therefore, the longer the distance, the more likely it is that the infinitive will be marked by the particle *to* (see also Lohmann 2011).

2.3 Principle of avoidance of identity, or *horror aequi*

Horror aequi is a widespread tendency to avoid repetition of identical elements (Rohdenburg 2003). This idea is also known as the Obligatory Contour Principle, which has been first formulated for phonology (Leben 1973), but has been used to explain different phenomena at all linguistic levels since then (e.g. omission of optional *that* in Walter & Jaeger 2008). Rohdenburg uses *horror aequi* to explain why the *to*-infinitive tends to be avoided immediately after a governing *to*-infinitive (e. g. *to try to do*). When the verb *help* is itself preceded by *to*, the following infinitive is usually without *to* (Biber et al. 1999: 737). See an example in (4):

(4) *Sorry, but how is this supposed <u>to</u>* ***help answer*** *the question?* (Great Britain, general, 303502) [1]

This hypothesis was confirmed by Lohmann (2011), who also finds an interaction between this factor and complexity (see Section 2.2). The more words there are between *help* and the infinitive, the weaker the influence of *horror aequi*.

2.4 Other factors

- Register: The shorter variant with the bare infinitive is considered to be less formal than the one with the marked infinitive (e.g. Rohdenburg 1996: 159; see also Biber et al. 1999: 736–737).

[1] The annotation means that the sentence is taken from the GloWbE corpus, British general subcorpus, website ID 303502.

- Dialect and time: It has been observed that American English has a particularly strong preference for the variant with the infinitive, although the bare infinitive is more common than the *to*-infinitive in both British and American varieties (e.g. Biber et al. 1999: 735). In addition, the bare infinitive has been gradually replacing the *to*-infinitive in the constructions with *help* in both varieties, so that one can speak of a parallel diachronic development (Mair 2002). As shown in a corpus study by Rohdenburg (2009: 318-319), the infinitive marker *to* was dropped very rarely in British and American English with the authors born to the end of the 18th century, but there was a significant increase in the drop of the marker by the end of the 19th century. This tendency continued in American English also in the 20th century, with British speakers following the suit with some delay, which supports Mair's (2002) claim of Americanization *cum* grammaticalization of *help*.
- Inflectional form: Lohmann (2011) observes that the form *helping* tends to be more frequently used with the *to*-infinitive in British English than the other inflectional forms of *help*. According to Rohdenburg (2009: 317), the effect of *helping* has an analogy with *daring* and *needing*, which differ from all forms of *dare* and *need* by being virtually always associated with marked infinitives. In addition to that, there is a weakly significant preference of the third person singular form *helps* for the *to*-infinitive in comparison with the base form (Lohmann 2011).
- Presence of the Helpee: Biber et al. (1999: 735) show that the bare infinitive is particularly dominant in the pattern *help* + NP + infinitive clause, which is also supported in Lohmann (2011).

3 Communicative efficiency and information theory

Communicative efficiency can be achieved by many ways, from choosing an appropriate politeness marker, to omission of redundant information. Iconicity and minimization of cognitive complexity (see Section 2) can also be regarded as devices for maximization of communicative efficiency. In the centre of the present discussion, however, is a specific case when less predictable elements, which carry more information, get more formal coding, and more predictable elements, which carry less information, get less coding. A well-known manifestation of this principle is the hypothesis of Uniform Information Density (see Levy & Jaeger 2007; Jaeger 2010). The hypothesis states that information tends to be distributed uniformly across the speech signal.[2]

A crucial question is how to measure predictability. In information-theoretic studies, which go back to Shannon's (1948) seminal work, one usually speaks about contextual (or rather co-textual) predictability, defined as the conditional probability of a unit given its immediate context, e.g. *n* words on the right or left. From this probability one can compute information content (also known as surprisal or informativity) of the unit in question. The less predictable a unit is from its context, the more informative (surprising) it is. There is ample evidence that more expected words, syllables or phonemes are more likely to undergo length reduction and loss of articulatory detail than less expected ones (e.g. Jurafsky et al. 2001; Aylett & Turk 2006; Bell et al. 2009; Mahowald et al. 2013).

Of particular relevance for the present study are the information-theoretic studies of grammatical alternations with optional elements. For example, Jaeger (2010) demonstrates that the omission of the complementizer *that*, as in *I know (that) he was at home yesterday*, is more likely when the presence of a complement clause is highly predictable from the matrix verb, e.g. *know*, *guess*, *think* or *say*. In another study, Jaeger (2011) finds that the so-called *Whiz*-deletion, as in the example *The smell (that is) released by a pig or chicken farm is indescribable*, depends on how much information is carried by the onset of the relative clause, as well as on the predictability of a relative clause given the specific noun. For a review of other studies, see Jaeger and Buz (In press).

As far as *help* + (*to*) Infinitive is concerned, one can hypothesize that the particle *to* will tend to be used if the context is more informative. Informative context is defined in the present study in two ways: a) based the predictability of the infinitive given HELP and b) based on the predictability of HELP given the infinitive.[3] The use of these two measures has been inspired by their existing analogues in the usage-based constructional approaches. These analogues are known as Attraction, i.e. the probability of a word filling a particular

[2] Although the mechanisms that underlie this probabilistic reduction are still a matter of debate, it seems that these effects, especially the omission of grammatical elements, are more likely to be explained by a tendency to maximize efficiency of message transmission than by maximization of production ease alone (cf. Jaeger and Buz, In press).

[3] HELP in small caps represents the construction with *help* + Infinitive as a whole, whereas *help* in italics stands for the lexeme.

constructional slot in the construction, and Reliance, i.e. the probability of a construction given its lexical slot filler (Schmid 2000).

4 Corpus and the procedure of data extraction

The data used in the present study come from the Corpus of Global Web-based English (GloWbE) created by Davies (2013). This large corpus contains 1.9 billion words and represents online English from twenty countries. For this case study, seven geographic varieties were chosen from different parts of the world: Australia, Ghana, Great Britain, Hong Kong, India, Jamaica and the USA. The choice for this corpus was motivated primarily by its size. One needs large corpora in order to compute reliable information-theoretic measures, especially if the construction of interest is not very frequent. I used a part of the corpus with eighteen million words per country, nine million from the 'General' subcorpus and nine million from the 'Blog' subcorpus.

The data extraction procedure was as follows. First, I used a Python script to collect all instances of *help* in any inflectional form followed by an infinitive somewhere in the sentence. If there were finite verb forms, clause-combining conjunctions like *because*, or subject pronouns like *I*, *he* and *she* between *help* and the infinitive, the instance was discarded. A quality check based on one hundred manually extracted examples from five subcorpora revealed that this approach was quite successful in recognizing the instances of the construction, with recall of 86% and precision of 93%. Only active uses were collected because the bare infinitive can be used only in active sentences (Huddleston and Pullum 2002: 1244), as shown in (5):

(5) a. *John was helped to cook the dinner.*

b. *??John was helped cook the dinner.*

The spelling variants of the verbs were normalised, so that the pairs like *maximize* and *maximise*, *fulfil* and *fulfill* were treated as one word.

In spite of the procedure of corpus cleaning performed by the corpus compilers, there were still quite a few duplicate sentences in the data. They were removed with the help of a script. Another problem were nonsense sentences, which were probably machine-generated or contained advertising information (cf. similar problems reported in Mair 2015: 31–32). However, they were not numerous and were removed during the process of variable coding.

After the data collection and cleaning, I obtained the frequencies shown in Table 1. Since the sizes of the subcorpora were identical (18 million words), the 'raw' frequencies are directly comparable between the varieties. One can see that Hong Kong has the highest total frequency of the constructions, and Jamaica the lowest. However, the differences are not large. As for the relative frequencies of the variants, the USA subcorpus displays the highest relative frequency of *help* followed by the bare infinitive (almost 85%), whereas the Jamaican subcorpus has the lowest one (only around 60%). Still, the variant with the bare infinitive is the more frequent one in all countries.

Country	***help* + bare Inf (%)**	***help* + *to* Inf (%)**	**Total**
Australia	4556 (76%)	1438 (24%)	5994 (100%)
Ghana	4980 (71.6%)	1971 (28.4%)	6950 (100%)
Great Britain	4153 (70%)	1782 (30%)	5937 (100%)
Hong Kong	5528 (72.6%)	2086 (27.4%)	7614 (100%)
India	5232 (72.7%)	1968 (27.3%)	7200 (100%)
Jamaica	3497 (60.4%)	2291 (39.6%)	5788 (100%)
USA	5124 (84.6%)	934 (15.4%)	6058 (100%)

Table 1: Absolute and relative frequencies of *help* + (*to*) Infinitive in seven countries.

The next section describes predictors for regression analysis, which represent the factors mentioned in Sections 2 and 3. One variable has not been taken into account, namely, animacy of the Helper, because it was very difficult to automate the annotation procedure. The parser returned very poor results due to high complexity of the syntactic structures (e.g.

when *help* was itself part of an infinitival clause). Note that the effect of animacy in Lohmann's (2011) study was rather weak. Two new variables were added: a contrast between animate and inanimate Helpees, and transitivity of the infinitive.

5 Predictors for regression analysis

5.1 Information-theoretic variables

This study tests the following information-theoretic variables:

a) information content of the infinitive given the construction, defined as the negative log-transformed conditional probability of the infinitive (with or without *to*) given the construction with *help*: –log P(Verb |HELP). This conditional probability is computed as the number of occurrences of a given infinitive with HELP divided by the total frequency of the construction with *help* in the relevant subcorpus. In corpus-based constructional studies this probability is known as *Attraction* (Schmid 2000). The more frequently a verb is used in the construction with *help* in comparison with the other verbs, the lower the information content;

b) information content of the construction given the infinitive, defined as the negative log-transformed conditional probability of the construction with *help* (with or without *to*) given the infinitive: –log P(HELP|Verb). This conditional probability, which is also known as *Reliance* (Schmid 2000), is computed by dividing the number of occurrences of a given infinitive with HELP by the total frequency of the verb in the subcorpus in all forms. The more frequently a verb is used in the construction with *help* in comparison with the other uses of the same verb, the lower the information content.

5.2 Variable representing cognitive complexity

This principle is represented by linguistic distance, which was measured as the number of words between the wordform of *help* and the infinitive (the particle *to* was not counted). For example, the sentence in (6) has the distance of four words.

(6) *I worked at Airbus before going into private equity in 2001,* ***helping*** *a European family office* ***to diversify*** *their investment portfolio.* (Hong Kong, blog, 3581048)

Although there are different ways of defining syntactic complexity, such as counting the number of syntactic nodes and quantifying the level of embeddedness, word counts serve as a good proxy for the more sophisticated measures (Szmrecsanyi 2004). This is why, following Lohmann (2011), I will use simple word counts, too.

5.3 Variable representing *horror aequi*

This factor is represented by the variable which reflects the presence of the particle *to* immediately before *help*, as in (7):

(7) *The Plate-Inversion protocol, and this post are two simple hacks* ***to help*** *you get started.* (India, blog, 3388613)

This is a binary variable, with the values "Yes" and "No".

5.4 Other variables

- Formality is represented by the average word length in the website text where a given instance of *help* was attested. The greater the average word length, the more formal the text. This operationalization is based on Biber's (1988) multidimensional analysis of register variation. He found, in particular, that longer word forms, alongside the type-token ratio and the relative frequency of nouns and adjectives, contribute strongly to the negative pole of the first factor or dimension, which is interpreted as “Involved vs. informational production” and has conversations and academic texts at its extremes. The use the mean word length is purely practical. Many texts in the corpus are very short and could not provide reliable relative frequencies for the lexico-grammatical categories required for a full-fledged multidimensional analysis.

- Morphological form: the inflectional form of the main verb with the values *help*, *helps*, *helped* and *helping*.

- Properties of the Helpee with the following values: zero (i.e. no explicit Helpee), animate (including organizations, countries and animals) and inanimate (all the rest). Examples are given in (8).

(8) a. *These bumps and turns will only help contribute towards a relationship*. [Zero] (Ghana, general, 1259905)

b. *It provides a systematic approach to helping* ***people*** *defeat dyslexia and related reading problems*. [Animate] (Great Britain, blog, 3058500)

c. *Five habits to help* ***your mind*** *get fit* [Inanimate] (Great Britain, blog, 3036513)

Zero Helpees are expected to be more often used with the *to*-infinitive than overt Helpees, following the previous findings (see Section 2.4). The values were assigned automatically if the Helpees were animate personal pronouns (i.e. *me, you*, *her, him* and *us*), and if the linguistic distance (see below) between *help* and the infinitive was zero. All other contexts were annotated manually. I also used user-defined contrasts, where animate Helpees were contrasted with inanimate ones, and both were compared with zero.

- Valency of the infinitive, with the following values: intransitive (including copulas), transitive (including ditransitives and very rare passives) or followed by a clause. Examples are shown in (9).

(9) a. *May God help nations* ***to live*** *together in peace.* [Intransitive]

(Ghana, blog, 3621705)

b. *Grow your business by helping your clients* ***grow*** *theirs* [Transitive] (Great Britain, blog, 3027910)

c. *Everyone has something to offer and it's about helping people* ***believe*** *they play an integral part in the workplace.* [Clause] (Great Britain, general, 416004)

In order to code this variable, the sentences were first parsed syntactically with the help of Stanford Parser (Klein & Manning 2003). The contexts were then manually checked, and the category 'Clause' was added manually.

6 Bayesian logistic regression with mixed effects: characteristics and results

6.1 Bayesian logistic regression: characteristics of the models

To test the effect of the predictors on the use of bare and *to*-infinitives, I used Bayesian mixed-effects logistic regression. For this purpose, I employed Stan, a programming language and platform for Bayesian inference (Stan Development Team 2015) and the R interface to Stan implemented in the R package *brms* (Bürkner, In press).

Seven Bayesian logistic regression models were fitted, one for each variety. The response variable was the use of the bare or *to*-infinitive. The predictors described in Section 5 were treated as fixed effects. The individual websites and the verbs that fill in the infinitive slot were treated as random effects (random intercepts). The websites and infinitives with the frequency less than five were conflated in one group. Sum contrasts were used with all categorical and binary variables, so that zero represents the grand mean (i.e. the unweighted mean of means) of the categories. The continuous variables were centred. Two interaction terms were modelled: the interaction between linguistic distance and the *horror aequi* variable, which was found to be significant by Lohmann (2011), and an interaction between the two information-theoretic measures. Pairwise interactions between the information-theoretic measures and the other factors were tested separately. They are discussed in Section

6.2.5. The discriminating power of the models was acceptable (with the concordance index *C* ranging from 0.73 to 0.79).

The Bayesian approach has a number of epistemological advantages in comparison with the traditional frequentist approach. Most importantly, the researcher can directly test the alternative hypothesis by estimating the probabilities of parameter values given the data. These probabilities can then be easily compared cross-lectally. They are called posterior probabilities because they are computed after the data have been taken into account. They also depend on prior probabilities, or priors, which represent the researcher's prior beliefs in the probability of some parameters before the data are taken into account. Some frequentist statisticians consider the use of priors too subjective. However, if one provides non-informative priors, as was done in the present study, this will result in posteriors that are influenced only by the data, as in frequentist statistics. For more information about the technical details of Bayesian modelling, one can be referred to Kruschke (2011). In what follows, I focus on the results.

Crucially, the algorithm returns 4000 posterior estimates of each regression parameter (1000 estimates in four Markov chains per each model). These probability distributions can be represented in a histogram which displays our posterior beliefs after the data have been taken into account. An example is provided in Figure 1. It shows the effect of zero Helpee as opposed to an explicit Helpee on the form of the infinitive in the Ghanaian variety. The numeric values on the horizontal axis are the log-odds ratios of the effect of the predictor on the response. A positive log-odds ratio means that the odds of the *to*-infinitive increase if the Helpee is not expressed, whereas a negative value means that the odds of the *to*-infinitive decrease (and, conversely, the odds of the bare infinitive increase). From the posterior distribution one can compute the posterior mean, which is displayed as a dot in Figure 1, as well as 95% highest density intervals (HDIs, or highest posterior density intervals, or HPDs, or credible intervals, as they are also referred to), which show the region between the 2.5% and the 97.5% percentiles, where the 95% of the posterior probability density lies.[4] HDIs span the most believable posteriors. If one has to make a categorical judgment of the type "Does the variable increase the chances of one or the other outcome?", one can use this criterion. If an HDI does not include zero, one can say that the effect is credibly nonzero

[4] Note that 95% HDIs of posteriors, or credible intervals, are conceptually different from 95% confidence intervals in frequentist statistics (see Kruschke 2011: 277–281).

(Kruschke 2011). The 95% HDI in the example includes zero. This means that we do not have the required degree of certainty that this variable has an effect.

The posterior distribution can also help us assess the probability of observing the positive and negative effect of a given predictor on the chances of the *to*-infinitive by computing the proportions of the posteriors that are greater and less than zero. In our example, the proportion of the posteriors greater than zero is 96.5%, whereas the proportion of the posteriors less than zero is only 3.5%. Even though we cannot say that the effect is credibly nonzero using the 95% HDI as the criterion, we still find a high probability of observing a positive effect. If we used only the frequentist approach and *p*-values, we would probably ignore this important information (cf. Vasishth et al. 2013).

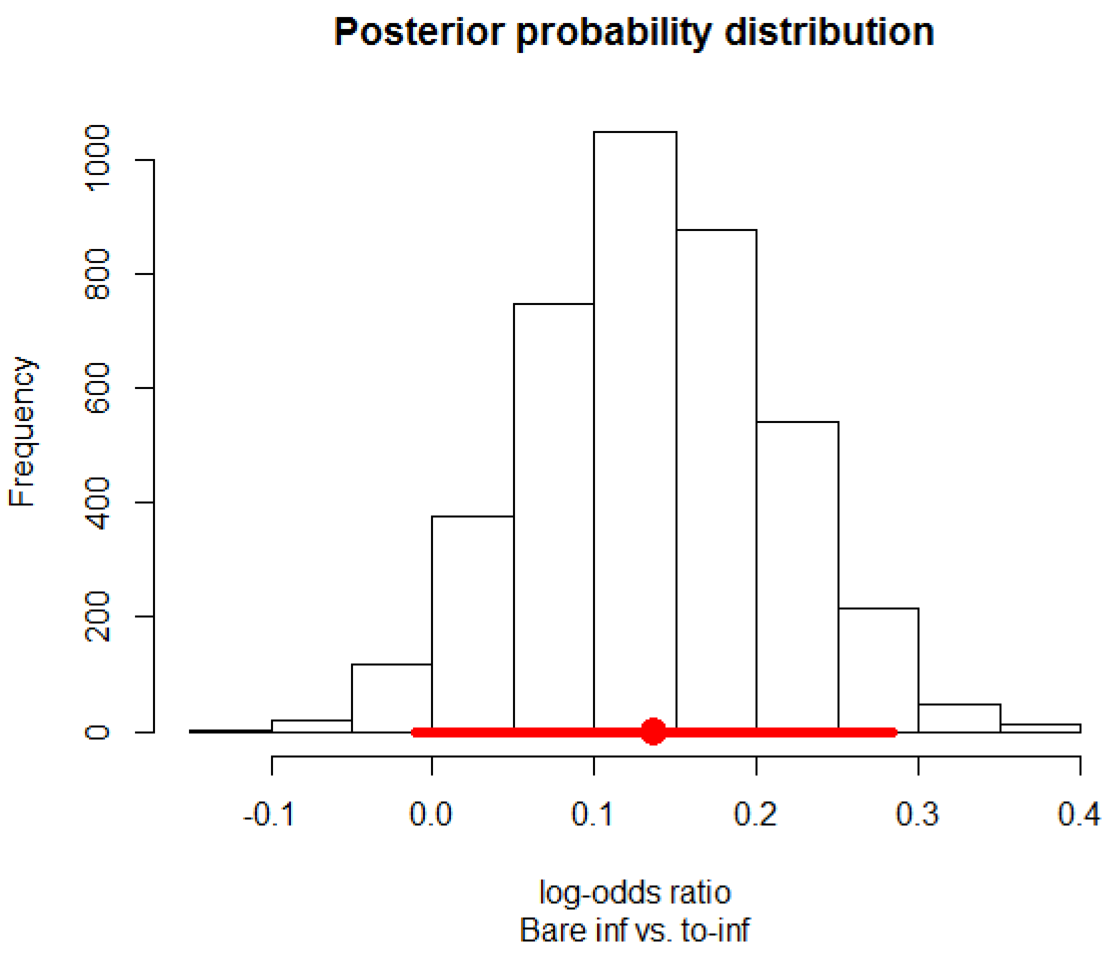

Figure 1: Posterior probability distribution of the effect of zero Helpee (vs. overt one) on the presence or absence of *to* in the Ghanaian variety.

The posterior distributions also allow us to measure the (dis)similarity between the varieties by computing the degree of overlap between each pair of varieties and then taking the average. Consider an illustration in Figure 2, which shows the posterior distributions of the effect of average word length in Great Britain and Jamaica (in log-odds ratios). Here, the overlap is about 20%. The greater the overlap, the more similar the varieties with regard to the effect of a given variable. The overlaps were computed with the help of the R package *overlapping* (Pastore 2017).

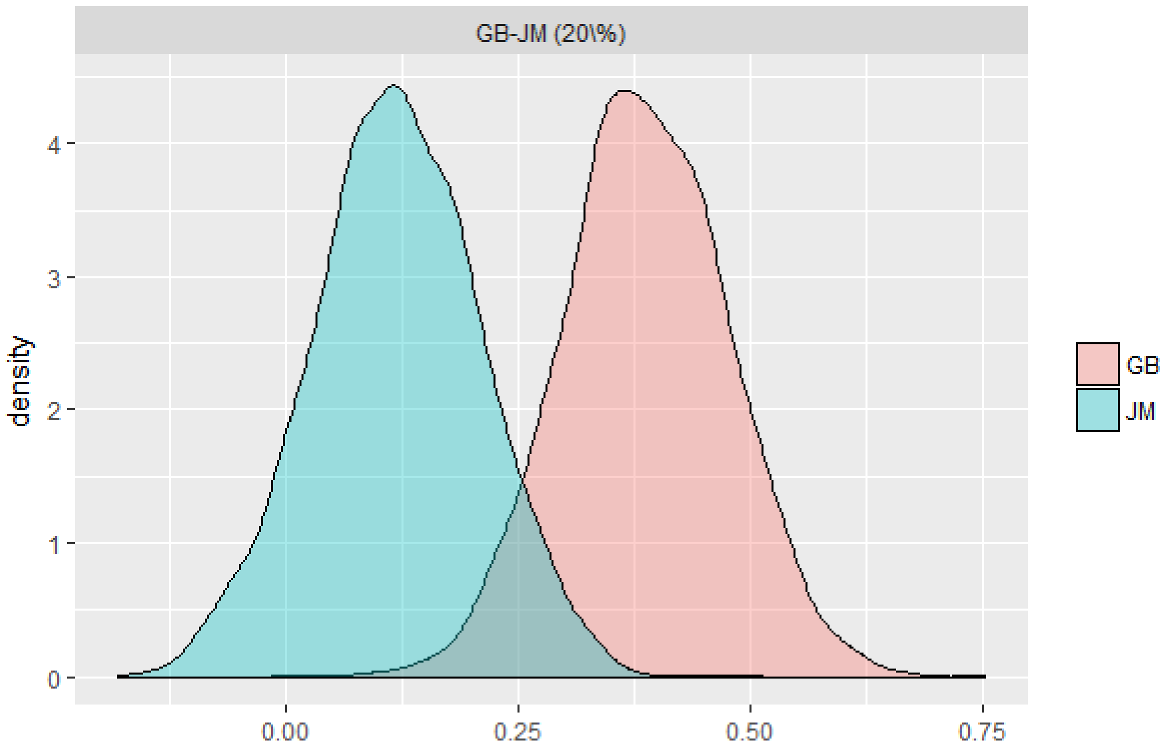

Figure 2: Overlap between two posterior distributions.

6.2 Results of Bayesian modelling

6.2.1. Information-theoretic variables

The results for the information content of the infinitive given HELP are reported in Table 2. If a value is positive, this means that an increase in this variable also increases the chances of the *to*-infinitive. The 95% HDIs, which lie between the percentiles 2.5% and 97.5%, are also shown in the table. The last column shows the probability of the effect being positive. Recall that we expect more informative, or surprising infinitives to increase the chances of the marked form. The table shows no evidence in support of this hypothesis. All HDIs include zero. Only the USA data reveal some tendency in that direction, with the posterior probability of this variable having an effect 90.6%. The average overlap between all pairs of the posterior distributions representing the varieties is 49.7%.

Country	**Posterior mean**	**2.5%**	**97.5%**	$P(\beta > 0)$
Australia	0.02	-0.06	0.1	72%
Ghana	0.04	-0.04	0.12	83.7%
Great Britain	-0.05	-0.13	0.03	12.6%
Hong Kong	0.03	-0.05	0.1	77.7%
India	-0.01	-0.08	0.06	42.1%
Jamaica	-0.02	-0.09	0.06	34.6%
USA	0.07	-0.04	0.17	90.6%

Table 2: Bayesian regression results for the information content of the infinitive given HELP (main effects).

In contrast, Table 3 presents very strong evidence in support of the effect of predictability, as determined by the information content of HELP given the infinitive. In all varieties, the result is the same: the greater the information content of HELP, or, in other words, the more surprising it is that the verb is used as a complement of *help*, and not in another function, the higher the chances of the marked infinitive. The average overlap between the posterior distributions is high and constitutes 62%.

Country	**Posterior mean**	**2.5%**	**97.5%**	$P(\beta > 0)$

Australia	0.12	0.05	0.19	99.9%
Ghana	0.1	0.03	0.16	99.7%
Great Britain	0.07	0	0.14	96.5%
Hong Kong	0.16	0.09	0.23	100%
India	0.09	0.03	0.15	99.95%
Jamaica	0.14	0.08	0.2	100%
USA	0.13	0.04	0.22	99.6%

Table 3: Bayesian regression results for the information content of HELP given the infinitive (main effect).

Figure 3 displays the differences between the percentages of *to* in all examples and in those where HELP is highly informative given the infinitive (top 5% of all scores in each variety). One can see that the proportion of *to*-infinitives is higher in the highly informative contexts than on average across all varieties.

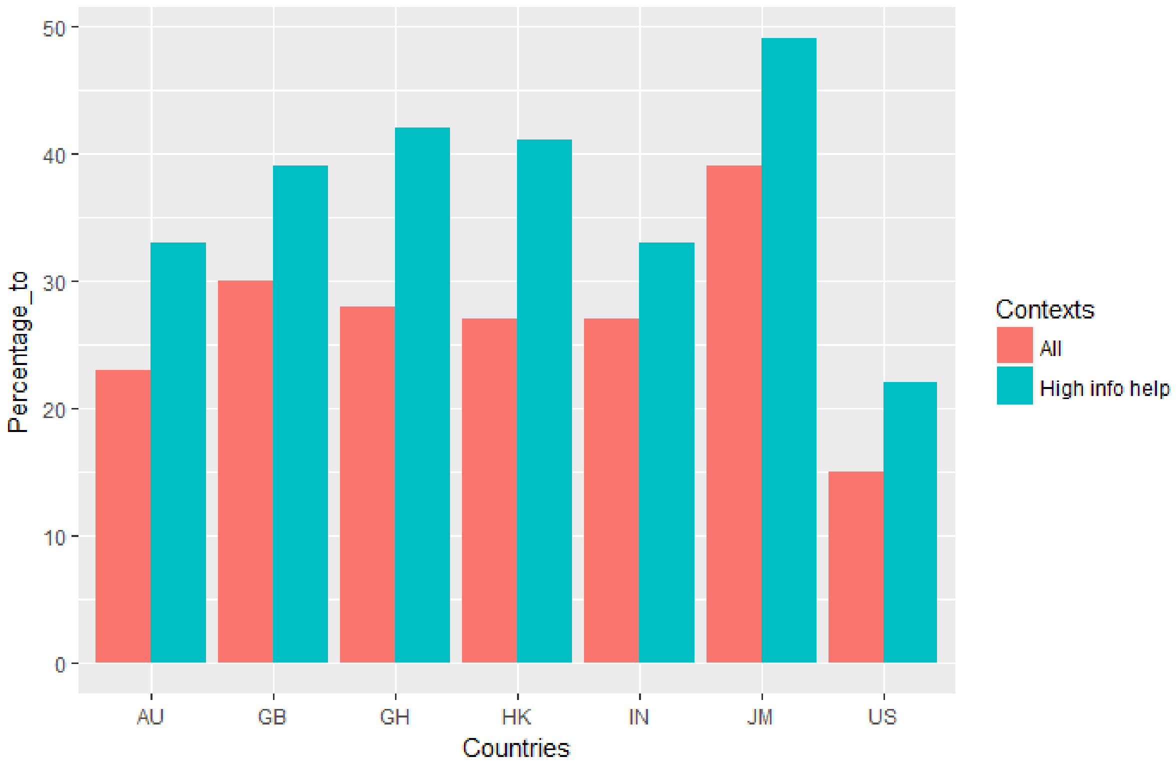

Figure 3: Percentages of *to*-infinitives in all contexts and in those where *help* given the infinitive is highly informative (top 5% of the scores).

Finally, Table 4 displays the results for the interaction term between the two information-theoretic variables. The results suggest that the variables interact in five out of seven varieties. There is no evidence of interaction in Great Britain and India. The interaction is of the same kind: when both measures are high, the chances of the *to*-infinitive are smaller than what one would expect if the effects were additive. As an illustration, the interaction in the Jamaican model is presented in Figure 4. The plot was created with the help of a Generalized Additive Model (Wood 2006), with the predicted probabilities of the *to*-infinitive as the response (based on the corresponding Bayesian model) and the two information-theoretic measures as the predictors. The warmer the colour, the higher the chances of the *to*-infinitive. The results for the other varieties are similar. The average overlap between the posterior distributions is 52%.

Country	**Posterior mean**	**2.5%**	**97.5%**	$P(\beta > 0)$
Australia	-0.03	-0.07	0.01	6%
Ghana	-0.03	-0.06	0.01	5.7%
Great Britain	-0.01	-0.05	0.03	35.2%
Hong Kong	-0.05	-0.09	-0.02	0.2%
India	0.001	-0.03	0.04	51%
Jamaica	-0.06	-0.1	-0.02	0.1%
USA	-0.03	-0.08	0.02	9.3%

Table 4: Bayesian regression results for the interaction term between two information-theoretic measures.

As can be seen in Figure 4, the increase in the proportion of *to*-infinitives is obvious only in the most informative contexts. Additional analyses with the help of Generalized Additive Mixed Models support this observation. In all varieties except for the Indian one, the effect of the variable representing the information content of *help* given the infinitive is somewhat non-linear, without a clear tendency for the contexts with low and middle informativity, and an obvious positive effect in favour of the *to*-infinitive only in highly informative contexts. In the Indian data, there is no evidence of non-linear relationships. The results obtained for the other predictors are corroborated.

This non-linearity can be explained in two ways. First, one can imagine that language users are more sensitive to informativity when it has higher values, and less sensitive to the differences between less informative contexts. Second, it may well be that the corpus-based estimates of informativity are less reliable for the verbs associated with less informative contexts (cf. Jaeger 2006: Section 5.1). More research is needed in order to answer this question.

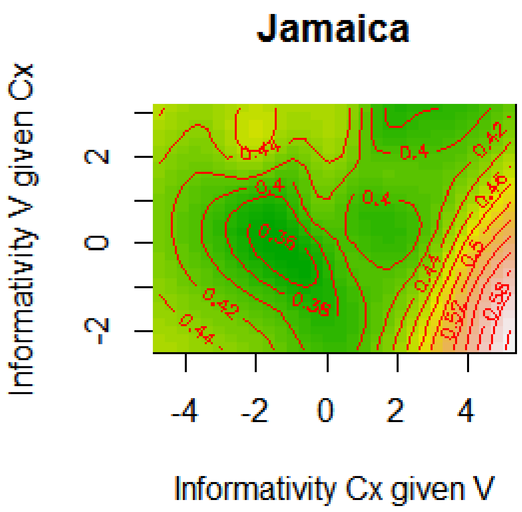

Figure 4: Interaction between two information-theoretic measures in the Jamaican variety.

6.2.2. Cognitive complexity, *horror aequi* and their interaction

As one can see from Tables 5–7, the results support all theory-driven predictions in all countries. Table 5 displays the main effects of linguistic distance. With each word between *help* and the infinitive, the odds of the *to*-infinitive credibly increase. There is some variation in the strength of this effect, however, with the American variety displaying the weakest effect, and the Indian one the strongest. The average overlap between the posterior distributions is 35.2%.

Country	Posterior mean	2.5%	97.5%	$P(\beta > 0)$
Australia	0.37	0.24	0.49	100%
Ghana	0.5	0.41	0.6	100%
Great Britain	0.45	0.34	0.57	100%
Hong Kong	0.53	0.43	0.63	100%
India	0.6	0.48	0.72	100%
Jamaica	0.36	0.25	0.47	100%
USA	0.3	0.18	0.42	100%

Table 5: Bayesian regression results for linguistic distance (main effect).

Table 6 shows the main effects of the presence of *to* before *help*. The chances of the *to*-infinitive decrease if there is *to* before *help*. There is some variation, again: the Hong Kong data display the weakest effect, and the USA the strongest effect. The average overlap between the posterior distributions is 47.5%.

Country	Posterior mean	2.5%	97.5%	$P(\beta > 0)$
Australia	-1.07	-1.21	-0.93	0%
Ghana	-0.95	-1.06	-0.85	0%
Great Britain	-1	-1.12	-0.89	0%
Hong Kong	-0.86	-0.96	-0.76	0%
India	-1.01	-1.14	-0.88	0%
Jamaica	-1.06	-1.17	-0.96	0%
USA	-1.11	-1.3	-0.94	0%

Table 6: Bayesian regression results for the presence of *to* before *help* (main effect).

The positive interaction terms (see Table 7) indicate that the odds of the *to*-infinitive become higher, as the linguistic distance between *help* and the infinitive increases. The estimates of the interaction term display very little geographic variation. Not surprisingly, the average overlap between the posterior distributions is very high: 67.1%.

Country	Posterior mean	2.5%	97.5%	$P(\beta > 0)$
Australia	0.28	0.18	0.38	100%
Ghana	0.27	0.19	0.34	100%
Great Britain	0.31	0.22	0.4	100%
Hong Kong	0.31	0.23	0.39	100%
India	0.32	0.21	0.42	100%
Jamaica	0.34	0.26	0.43	100%
USA	0.24	0.13	0.34	100%

Table 7: Bayesian regression results for the interaction term between *linguistic distance* and the presence of *to* before *help*.

6.2.3. Other variables

First, let us consider the degree of formality represented by the average word length in a website text. The posteriors in Table 8 show the effect of adding one letter on the log-odds of the *to*-infinitive vs. the bare infinitive. In some countries (Great Britain, Australia, Hong Kong and Jamaica) the average word length has a highly probable positive effect on the chances of the *to*-infinitive, as predicted. The strongest effect is observed in Great Britain. The Indian data show, surprisingly, the opposite tendency. The average overlap is rather modest, only 39.9%.

Country	Posterior mean	2.5%	97.5%	$P(\beta > 0)$
Australia	0.19	0.02	0.37	98.4%
Ghana	0.06	-0.1	0.22	75.5%
Great Britain	0.39	0.21	0.56	100%
Hong Kong	0.16	0	0.31	97.7%
India	-0.27	-0.45	-0.1	0.2%

Jamaica	0.12	-0.06	0.29	90.7%
USA	0.04	-0.18	0.27	64.2%

Table 8: Bayesian regression results for the average word length.

Next, the morphological form *helping* increases the chances of *to* in all varieties, as shown in Table 9. The Indian variety has the smallest posterior mean, whereas the USA data display the strongest effect. The average overlap between the posterior distributions is 47.7%.

Country	**Posterior mean**	**2.5%**	**97.5%**	$P(\beta > 0)$
Australia	0.56	0.42	0.7	100%
Ghana	0.57	0.43	0.71	100%
Great Britain	0.55	0.42	0.69	100%
Hong Kong	0.47	0.32	0.61	100%
India	0.2	0.04	0.35	99.5%
Jamaica	0.52	0.38	0.67	100%
USA	0.74	0.58	0.9	100%

Table 9: Bayesian regression results for the effect of the morphological form *helping*.

As for the other morphological forms, one finds the following hierarchy in most countries:

(10) *helping* > *helps* > *helped* > *help*

This hierarchy is based on the posterior means, which are the greatest for the form *helping*, followed by *helps* and *helped*, and the smallest for *help*. The latter was computed on the basis of the posterior means of the three other forms. There are two exceptions. In the Indian model, *helps* is more strongly associated with the *to*-infinitive than *helping*. In the Ghanaian one, the estimates of *helps* and *helped* are very close.

The next contrast is between zero and overt Helpees. If the Helpee is absent, there is a very high probability that the chances of the *to*-infinitive being used increase in all countries (see Table 10). This is where the usefulness of the Bayesian approach becomes obvious: if one only made decisions based on the *p*-values, it would be more difficult to notice that the Ghanaian variety behaves very similarly to the other varieties. The strongest effect is found in the Jamaican data. The average overlap between the posterior distributions is 52.4%.

Country	**Posterior mean**	**2.5%**	**97.5%**	$P(\beta > 0)$
Australia	0.29	0.14	0.45	100%
Ghana	0.14	-0.01	0.28	96.5%
Great Britain	0.35	0.2	0.5	100%
Hong Kong	0.32	0.2	0.44	100%
India	0.29	0.17	0.41	100%
Jamaica	0.43	0.28	0.59	100%
USA	0.39	0.22	0.56	100%

Table 10: Bayesian regression results for the expression of the Helpee (no explicit Helpee).

Table 11 displays the results for inanimate Helpees contrasted with animate ones. There is some evidence that this distinction plays a role only in the Jamaican variety, where the chances of the posteriors being positive are 9%. This means that inanimate Helpees occur more frequently with the bare infinitive in that subcorpus than animate ones. The average overlap between the posterior distributions is 54%.

Country	**Posterior mean**	**2.5%**	**97.5%**	$P(\beta > 0)$
Australia	0.11	-0.1	0.3	84.7%
Ghana	0.02	-0.19	0.24	58%
Great Britain	-0.11	-0.32	0.11	16.3%

Hong Kong	-0.04	-0.21	0.12	29.8%
India	-0.05	-0.20	0.11	26.2%
Jamaica	-0.15	-0.38	0.07	9%
USA	0.003	-0.25	0.25	51.4%

Table 11: Bayesian regression results for the semantics of the Helpee (inanimate).

Finally, Table 12 shows the numbers that represent the effect of transitivity of the infinitive on the presence of *to*. One can see that high probabilities (close to 100%) are observed only in Hong Kong and India. In the other countries, there is no strong bias in either direction. The average overlap between all pairs of posterior distributions is 59.2%, which is relatively high. A separate check (not shown here) reveals that the presence of clause complements had no clear effects on the use of the infinitive in any of the varieties.

Country	Posterior mean	2.5%	97.5%	$P(\beta > 0)$
Australia	-0.003	-0.13	0.17	48.2%
Ghana	0.05	-0.09	0.19	74.9%
Great Britain	-0.02	-0.15	0.11	38.4%
Hong Kong	0.13	0	0.25	97.8%
India	0.15	0.02	0.27	99.5%
Jamaica	0.08	-0.05	0.22	89.7%
USA	0.03	-0.13	0.19	64%

Table 12: Bayesian regression results for valency of the infinitive (transitive).

6.2.4. Interactions between information-theoretic variables and other factors

In addition, all pairwise interactions between the information-theoretic variables and the other variables were tested in each variety. An interaction was taken into account only when the HDI of the interaction term did not include zero. The resulting interactions differ substantially across the varieties. This is why they were not included in the models presented above. An analysis of these interactions allows one to make several generalizations, however. First, nearly all of the interactions involve the predictability of HELP given the infinitive, and not the other information-theoretic measure. Second, they are weak and do not change the direction of the relationships, with the exception of the Jamaican data, where one sees a reversed effect of the Helpee's animacy in highly informative contexts, so that animate Helpees are more often used with the *to*-infinitives than the inanimate ones. However, the tendency was also not very convincing in the model without the interaction, in the first place. Third, in almost all of them, the effect of the non-information-theoretic variables slightly decreases in highly informative contexts. The only exception is found in the Indian variety, where the strength of the formality effect increases with information content. These exceptions, as well as the reasons for the decreasing effect of the other variables in highly informative contexts, require further investigation.

7. Discussion

In general, the bare infinitive is the preferred variant in all varieties discussed here. The highest proportion of the bare infinitive is observed in the USA data, whereas the lowest proportion is found in the Jamaican subcorpus, although the difference is not very large.

The results of the previous studies are largely corroborated. The variables related to *horror aequi* and the principle of cognitive complexity behave in accordance with the expectations in all varieties. They interact, such that the effect of *to* before *help* weakens with linguistic distance between *help* and the infinitive. The varieties also behave similarly with regard to the form *helping*, which substantially increases the chances of the *to*-infinitive. It is followed by *helps* and *helped* in most varieties. These findings also support the idea that inflectional forms of words have their own semantic, pragmatic, stylistic and collocational properties, which speakers are sensitive to (Newman & Rice 2006). As for register variation, the expected effect of formality is observed in Australia, Great Britain, Hong Kong and Jamaica: the more formal the communication, as measured by the average word length in a text, the greater the chances of the *to*-infinitive being chosen. However, one finds an opposite effect in India. This surprising finding requires further investigation. We also find high geographic uniformity with regard to the absence or presence of the Helpee. As for the other variables (valency of the infinitive and semantics of the Helpee), credible effects are found only in the data from Hong Kong, India and to some extent in Jamaica.

Let us now turn to the central question of the present study: does information content help predict the use of the bare or *to*-infinitive, other factors being controlled for? The answer to this question is positive: in highly informative contexts, the chances of the marked infinitive increase. Interestingly, it is the information content of HELP before the infinitive that matters, rather than the information content of the infinitive itself. The infinitives that are associated with high information content of HELP are, as a rule, highly frequent verbs, such as *be, have, do, say, ask, believe, tell, use* and *go*.[5] A few examples are shown in (11). These verbs appear in many diverse constructions, which explains the high information content of HELP. Note that in some varieties the effect is weaker if the infinitive itself is highly surprising given the construction, as well, as can be concluded from the interactions between the information-theoretic measures.

(11) a. *Growing plants will **help** you **to be** patient*. (Hong Kong, blog, 3585980)

b. *It will **help** your partner **to have** clear insight regarding your travelling habits.* (India, general, 623003)

c. *…if I try to **help** him **to do** it better, he gets an attitude and yells "I don't care about baseball"* (USA, general, 44601).

It is worth mentioning that the type of predictability effects observed in the present study is rather unusual. Most information-theoretic studies of grammatical alternations show that speakers tend to provide extra marking on unexpected units, such as the more frequent use of *that* before relative clauses (e.g. *I like the way (that) she moves*) that are less predictable from the lexical properties of the noun phrase (Wasow et al. 2011) or the more frequent case marking of semantically untypical direct objects in Japanese (Kurumada & Jaeger 2015). In the case of *help*, however, the marking is on the infinitive, but the infinitive is not surprising itself. What is unexpected, is the construction in which it appears. Still, the effect found in

[5] An investigation of the precise role of verb frequency is left for future research due to its high correlation with the information content measures and the danger of multicollinearity.

this study can be considered a manifestation of the general tendency to maximize communicative efficiency because the additional coding is provided in ‘tricky’ contexts. If a verb is ‘unfaithful’ to HELP, it requires additional marking in order to be recognized as part of the construction.

Notably, the average overlap of the posterior distributions for the main effect of this information-theoretic variable is the second highest (following only the interaction term between complexity and *horror aequi*). This means that the varieties display high similarity with regard to this factor. In contrast, linguistic distance between *help* and the infinitive (as the main term) and text formality display the smallest average overlap between the posterior distributions. In both cases, the Indian variety behaves in a way that is different from the other varieties.

To summarize, the online web-based corpora of English varieties provide us with evidence that the expectedness of a given verb to serve as a complement of HELP or to perform a different function can be added to the list of factors that predict the user’s choice between the constructional variants. These effects are very similar across the varieties. This finding, in addition to the existing evidence from different languages and various linguistic phenomena, allows one to conclude that the tendency to minimize formal coding in predictable contexts and to maximize it in surprising contexts is a universal and independent communicative constraint.

References

Aylett, Matthew & Alice Turk. 2004. The smooth signal redundancy hypothesis: A functional explanation for relationships between redundancy, prosodic prominence, and duration in spontaneous speech. *Language and Speech*, 47(1). 31-56.

Bell, Alan, Jason Brenier, Michelle Gregory, Cynthia Girand and Dan Jurafsky. 2009. Predictability Effects on Durations of Content and Function Words in Conversational English. *Journal of Memory and Language*, 60(1). 92-111.

Biber, Douglas, Stig Johansson, Geoffrey Leech, Susan Conrad & Edward Finegan. (1999). Longman Grammar of Spoken and Written English. Harlow: Longman.

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

Bürkner, Paul-Christian. In press. brms: An R Package for Bayesian Multilevel Models using Stan. *Journal of Statistical Software*.

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

Davies, Mark. 2013. *Corpus of Global Web-Based English: 1.9 billion words from speakers in 20 countries*. Available online at http://corpus.byu.edu/glowbe/.

Dixon, R.M.W. 1991. *A New Approach to English Grammar, on Semantic Principles*. Oxford: Clarendon Press.

Duffley, Patrick J. 1992. *The English infinitive*. London &New York: Longman.

Givón, Talmy. 1990. *Syntax: A Functional-Typological Introduction. Vol. 2.* Amsterdam: John Benjamins.

Haiman, John. 1983. Iconic and economic motivation. *Language* 59(4). 781–819.

Huddleston, Rodney and Geoffrey K. Pullum. 2002. The Cambridge Grammar of the English Language. Cambridge: Cambridge University Press.

Jaeger, T. Florian. 2006. *Redundancy and Syntactic Reduction in Spontaneous Speech*. PhD diss., Stanford University.

Jaeger, T. Florian. 2010. Redundancy and reduction: speakers manage syntactic information density. *Cognitive Psychology* 61 (1). 23-62. doi: 10.1016/j.cogpsych.2010.02.002

Jaeger, T. Florian 2011. Corpus-based research on language production: Information density and reducible subject relatives. In Emily M. Bender & Jennifer E. Arnold (eds.), *Language from a cognitive perspective: Grammar, usage, and processing. studies in honor of Tom Wasow*, 161–197. Stanford, CA: CSLI Publications.

Jaeger, T. Florian & Esteban Buz. In press. Signal reduction and linguistic encoding. To appear in Eva M. Fernández & Helen Smith Cairns (eds.), *Handbook of Psycholinguistics*.

Jurafsky, Daniel, Alan Bell, Michelle L. Gregory & William D. Raymond. 2001. Probabilistic relations between words: Evidence from reduction in lexical production. Bybee, Joan L. & Paul Hopper (eds.), *Frequency and the Emergence of Linguistic Structure,* 229–254. Amsterdam: John Benjamins.

Klein, Dan & Christopher D. Manning. 2003. Accurate Unlexicalized Parsing. In *Proceedings of the 41th Annual Meeting of the Association for Computational Linguistics*, 423-430. Accessed from http://nlp.stanford.edu/manning/papers/unlexicalized-parsing.pdf (last access 26.10.2016).

Kruschke, John. 2011. *Doing Bayesian Data Analysis. A Tutorial with R and BUGS.* Amsterdam etc.: Elsevier.

Kurumada, Chigusa & T. Florian Jaeger. 2015. Communicative efficiency in language production: Optional case-marking in Japanese. *Journal of Memory and Language* 83. 152-178. doi: 10.1016/j.jml.2015.03.003

Leben, William. 1973. *Suprasegmental Phonology*. PhD dissertation, MIT.

Levy, Roger & T. Florian Jaeger. 2007. Speakers optimize information density through syntactic reduction. In Bernhard Schlökopf, John Platt & Thomas Hoffman (eds.), *Advances in neural information processing systems (NIPS)* Vol. 19, 849–856. Cambridge, MA: MIT Press.

Lind, Age. 1983. The variant forms of *help to*/*help Ø*. *English Studies* 64. 263–275.

Lohmann, Arne 2011. Help vs. help to - a multifactorial, mixed-effects account of infinitive marker omission. *English Language and Linguistics* 15(3). 499-521.

Mahowald, Kyle, Evelina Fedorenko, Steven T. Piantadosi & Edward Gibson. 2013. Info/information theory: Speakers choose shorter words in predictive contexts. *Cognition* 126. 313 – 318.

Mair, Christian. 2002. Three changing patterns of verb complementation in Late Modern English: a real-time study based on matching text corpora. *English Language and Linguistics* 6(1). 105–31.

Mair, Christian. 2015. Responses to Davies and Fuchs. *English World-Wide* 36(1), 29–33.

Newman, John & Sally Rice. 2006. Transitivity schemas of English EAT and DRINK in the BNC. In Stefan Th. Gries & Anatol Stefanowitsch (eds.), *Corpora in Cognitive Linguistics: Corpus-based Approaches to Syntax and Lexis*, 225–260. Berlin/New York: Mouton de Gruyter.

Pastore, Massimiliano. 2017. *overlapping*: Estimation of Overlapping in Empirical Distributions. R package version 1.4.0. https://CRAN.R-project.org/package=overlapping

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.

Rohdenburg, Günther. 2009. Grammatical divergence between British and American English in the nineteenth and early twentieth centuries. In Ingrid Tieken-Boon van Ostade & Wim van der Wurff (eds)., *Current Issues in Late Modern English* (Linguistic Insights 77), 301–330. Bern: Peter Lang.

Schmid, Hans-Jörg. 2000. *English Abstract Nouns as Conceptual Shells. From corpus to cognition*. Berlin & New York: Mouton de Gruyter.

Shannon, Claude E. 1948. A Mathematical Theory of Communication, *Bell System Technical Journal* 27. 379–423 & 623–656.

Stan Development Team. 2015. Stan: A C++ Library for Probability and Sampling, Version 2.8.0. URL http://mc-stan.org/.

Szmrecsanyi, Benedikt. 2004. On operationalizing syntactic complexity. In Gérard Purnelle, Cédrick Fairon & Anne Dister (eds.), *Le poids des mots. Proceedings of the 7th International Conference on Textual Data Statistical Analysis*. Louvain-la-Neuve, March 10-12, 2004, Vol. 2, 1032–1039. Louvain-la-Neuve, Presses universitaires de Louvain.

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. Available at http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0077006 (accessed 21 November 2015)

Walter, Mary Ann, & T. Florian Jaeger. 2008. Constraints on optional *that*: A Strong word form OCP effect. In R. L. In Edwards, P. J. Midtlyng, K. G. Stensrud, & C. L. Sprague (Eds.), *Proceedings from the Annual Meeting of the Chicago Linguistic Society*, 505–519. Chicago, IL: CLS.

Wasow, Thomas, T. Florian Jaeger & David M. Orr, 2011. Lexical variation in relativizer frequency. In Horst J. Simon & Heike Wiese (eds.), *Expecting the unexpected: Exceptions in grammar*, 175–195. Berlin: De Gruyter Mouton.

Wood, Simon N. 2006. *Generalized Additive Models: An Introduction with R*. Chapman and Hall/CRC.