Cognitive indigenization effects in the English dative alternation

Röthlisberger, Melanie; Grafmiller, Jason; Szmrecsanyi, Benedikt

Link to publication on Research at Birmingham portal

Cognitive indigenization effects in the English dative alternation

Melanie Röthlisberger (KU Leuven), Jason Grafmiller (KU Leuven) and Benedikt Szmrecsanyi (KU Leuven)

Corresponding author: Melanie Röthlisberger
                       Department of Linguistics
                       Blijde-Inkomststraat 21 PO Box 03308
                       3000 Leuven
                       Belgium
                       melanie.rothlisberger@kuleuven.be

Co-authors: Jason Grafmiller, jason.grafmiller@kuleuven.be

                Benedikt Szmrecsanyi, benszm@kuleuven.be

Cognitive indigenization effects in the English dative alternation

Melanie Röthlisberger

Jason Grafmiller

Benedikt Szmrecsanyi

Abstract

We advance theory formation in Cognitive Sociolinguistics by exploring the extent to which language users' probabilistic grammar varies regionally. For this purpose, we investigate the effects of constraints that influence the choice between the two syntactic variants in the well-known dative alternation (*I give Mary a book* vs. *I give a book to Mary*) across nine post-colonial varieties of English. Using mixed-effects logistic regression and adopting a large-scale comparative perspective, we illustrate that on the one hand, stability in probabilistic grammars prevails across speakers of diverse regional and cultural backgrounds. On the other hand, traces of indigenization are found in those contexts where shifting usage frequencies in language-internal variation seem to have led to regional differences between users' probabilistic grammar(s). Within a psycholinguistically grounded model of probabilistic grammar, we interpret these results from various explanatory perspectives, including language contact phenomena, second language acquisition, and semantic variation and change.

1 Introduction

Recent years have seen the emergence of studies exploring the regional variability of constraints that influence syntactic variation in postcolonial varieties of English (e.g. Bresnan and Hay 2008; Bresnan and Ford 2010; Hundt and Szmrecsanyi 2012). The present study contributes to this conversation through an investigation of variability in the well-known English dative alternation, exemplified in (1), across nine international varieties of English.

(1) a. the ditransitive dative variant

I'd given [Heidi] recipient [my T-shirt] theme . <ICE-GB:S1B-066>

b. the prepositional dative variant

And I'd given [the key] theme to [Helen] recipient. <ICE-CAN:S1A-058>

The methods and theoretical scope of the present study are firmly grounded in the functional-cognitivist and corpus-based tradition of dative alternation research. Of central concern are the language-internal and external factors that influence the variable ordering between the theme and recipient arguments of a given dative verb.

In the spirit of previous work, we restrict our attention therefore to contexts where either variant is available, following standard practice in variationist sociolinguistics (e.g. Kendall et al. 2011). This variationist perspective regards the two variants as very close paraphrases, or in other words, as "alternative ways of saying 'the same' thing" (Labov 1972: 188). While this assumption is not uncontroversial, see e.g. Goldberg (2002), recent experimental work has supported the variationist perspective with regard to syntactic alternations. Perek (2012) shows that if speakers are given the task of sorting constructions based on their conceptual representation, they rely more on the semantic overlap between the variants than on the constructions' syntactic form to guide their judgment; in other words, they tend to pit ditransitive and prepositional dative constructions against other locative prepositional constructions (Perek 2012).

Abundant research has shown that the factors governing the dative alternation are multifaceted and non-deterministic: no single feature (or set of features) categorically determines the choice of a given variant (see Bernaisch et al. 2014; Bresnan et al. 2007; Theijssen et al. 2013; among many others). Instead, numerous factors probabilistically influence the use of one or the other variant. These include pronominality, definiteness, and length of the respective constituents (e.g. pronominal recipients favor the ditransitive, pronominal themes the prepositional dative), along with the semantics of the token in question: abstract uses of *give* – *give the offer a second thought* – favor the ditransitive, while uses representing physical transfer – *give my card to them* – favor the prepositional dative variant. From a comparative perspective, there is some evidence that these factors may vary in subtle ways across different speech communities (Bresnan and Ford 2010; Tagliamonte 2014), however the extent of conditioning factors' potential cross-lectal variability is still not well understood (see Bernaisch et al. 2014).

The present study builds upon this research tradition in two ways. First, while recent studies have focused on the prototypical ditransitive verb *give* (e.g. Bernaisch et al. 2014; Bresnan and Hay 2008), we analyze fully 83 alternating dative verbs (see Section 3.2). In doing so, we find that some subtle patterns of variability among individual factors do not necessarily generalize from the prototypical case, i.e. *give*, to the grammar as a whole. Second, we include data from a geographically wider and more diverse set of English varieties than previous comparative studies (e.g. Bresnan and Ford 2010; Tagliamonte 2014), including both native (L1) varieties such as British, Canadian, and New Zealand English, as well as varieties whose speakers are primarily second language (L2) speakers of English, such as Indian, Jamaican, and Hong Kong English. Given the considerable body of work on the dative alternation, there is still much we do not know about the cross-varietal plasticity of the (probabilistic) factors shaping the dative grammar(s) of different speakers and communities. By drawing on production data from nine national varieties of English and including an extensive set of dative verbs, this large-scale comparative study contributes to patching this hole in our current understanding of the English dative alternation.

Our primary interest thus lies in delimiting the scope of syntactic variation within and among different varieties of English around the world. We seek to answer two key questions: i) to what extent do we find a stable (probabilistic) grammar across nine international varieties of English; and ii) are some grammatical constraints more amenable to regional differences than others? In essence, we are interested in the

extent to which speakers of different varieties of the same language rely on the same processes and/or cues when choosing between dative variants.

Theoretically, we assume a model of grammar that is at its core dynamic, probabilistic, and usage-based (e.g. Bybee and Hopper 2001), and extend this model to our investigation of cross-varietal syntactic variation à la Bresnan and Hay (2008) and Bresnan and Ford (2010). Our study also ties in with recent research in Cognitive Sociolinguistics – a relatively novel linguistic subdiscipline that merges the main viewpoint of Cognitive Linguistics, namely that language is entrenched within one's general cognitive abilities, with a sociovariationist view, that is, an interest in the social and cultural forces that drive variation in human interaction (e.g. Geeraerts et al. 2010; Harder 2010; Kristiansen and Geeraerts 2013). Cognitive sociolinguistics confronts the inherent heterogeneity of language as a social construct and is concerned with the effect that cognitive and sociocultural forces exert on the formation of distinct lects. From a cognitive (socio)linguistic perspective, variationist studies such as the present one can be seen as investigations of the forces shaping the interaction between 'formal onomasiological variation' and 'speaker and situation related variation' (Geeraerts et al. 2010: 7–8) – specifically variation among different regional lects. Recent work in that subfield demonstrates that syntactic choices within and across varieties of a given language are governed by language-internal forces that can exhibit subtle degrees of variability across regions (e.g. Bresnan and Hay 2008; Mukherjee and Hoffmann 2006), time (e.g. Wolk et al. 2013), and register (e.g. Grafmiller 2014; Gries 2013). The majority of these studies employ logistic regression analysis to gauge the variable effects of language-internal constraints on a binary syntactic choice on the premise that these statistical models reflect speakers' grammatical knowledge. The small number of studies that further test the models' predictions on behavioral data with supplementary rating task experiments (e.g. Bresnan and Ford 2010) largely validate those corpus-based results.

Probabilistic grammar models also fit squarely with functional approaches to grammatical structure and variation, e.g. MacDonald (2013), which assume that users are subject to the same psychological processes shaping production and comprehension, and are thus likely to make similar syntactic choices, all else being equal. For instance, Bresnan et al. (2007) found that in conversational American English, speakers opt for the dative variant that places the 'easier' or more accessible constituent before the less accessible one. This tendency illustrates the general bias for language users to place 'easy' elements first. 'Easy' in this sense refers to those elements that are more quickly retrieved from (long-term) memory (MacDonald 2013: 4), and an element may be easier to retrieve by virtue of it being more frequent, shorter, less syntactically complex, more conceptually salient, or having been recently mentioned. Uttering the easier elements first gives the speaker enough time to plan and produce the more difficult constituents. This general pattern of ordering in the dative alternation has been observed in numerous native and non-native varieties of English (e.g. Bresnan and Ford 2010; De Cuypere and Verbeke 2013; Tagliamonte 2014).

Importantly, we argue that social meaning and socially conditioned variation is entirely compatible with probabilistic grammar models. Community-specific social forces, e.g. language attitudes or stylistic preferences, undoubtedly shape biases in individual speakers' production and comprehension, while at the same time, ad hoc meaning formation that arises during individuals' interactions can lead to innovation and greater variability among syntactic forms and their semantic cues. The resulting

patterns are in turn reflected in specific forms' distributions across different social groups/contexts, as all usage-based theories assume. This variation in the use of specific constructions may be driven by social forces, e.g. stylistic preferences among registers, by cognitive forces, e.g. biases in L2 acquisition/usage, by situational forces, e.g. language contact scenarios, or by normal dialectal drift.

The development of subtle, region-specific grammatical variation in postcolonial Englishes' datives represents a process we call *cognitive indigenization*. *Indigenization* or *nativization* is used in research in New Englishes to refer to "the emergence of locally characteristic linguistic patterns" (Schneider 2007:6). Recent studies have shown that indigenization does not only take place at the lexis-grammar interface but also on more fine-grained levels of linguistic knowledge, namely in the underlying stochastic patterns that make up speakers' probabilistic grammar (see Szmrecsanyi et al. 2016). Shifting usage frequencies in language internal variation can thereby lead to a gradient localized acculturation of probabilistic constraints (such as end-weight) in the grammar of speakers from different communities, an effect that has been termed *probabilistic indigenization* (Szmrecsanyi et al. 2016). By adding a *cognitive* dimension, we would like to stress the outcome of probabilistic indigenization, namely the "lectalization" or creation of distinct lects (that speakers are aware of) at the level of very subtle gradience.

The remainder of the paper is structured as follows: Section 2 sketches the data and methodology. Section 3 presents the results, and is followed by the discussion in Section 4. Conclusions and final remarks are offered in Section 5.

2 Data and methods

2.1 The data

For our investigation of the dative alternation, we tap into the International Corpus of English (ICE) series, which samples naturalistic language data from 12 spoken and written registers across numerous national varieties of English (see Greenbaum 1996). In order to facilitate comparative studies between varieties, each ICE component follows the same corpus design: each variety comprises a collection of 500 texts of approximately 2000 words each – ~1 million words in total (60% spoken; 40% written).

The varieties sampled here include:

- British English (BrE)
- Canadian English (CanE)
- Irish English (IrE)
- New Zealand English (NZE)
- Jamaican English (JamE)
- Singapore English (SinE)
- Indian English (IndE)
- Hong Kong English (HKE)

- Philippine English (PhiE)

2.2 Retrieving dative tokens

All dative tokens were extracted from the corpus using a list of dative verbs adapted from previous literature (Bresnan et al. 2007; De Cuypere and Verbeke 2013; Levin 1993; Mukherjee and Hoffmann 2006; Wolk et al. 2013 and others) (see example (2)]). At first, this list contained any verb known to occur in either the ditransitive or prepositional dative variant in Standard English (Levin 1993). Supposedly non-interchangeable verbs in Standard English (e.g. *donate the money to charity ~ *donate charity the money*) were included in the list since some such verbs may in fact vary in non-Standard varieties of English. At present, there exists no exhaustive list(s) of all interchangeable dative verbs in all the varieties studied here. Therefore, a given verb was considered interchangeable if it occurred in both ditransitive and prepositional variants in the ICE corpora, or in independent datasets, e.g. the full GloWbE Corpus (Davies 2013). If we found at least 5 instances of the verb in each variant, the verb was considered interchangeable.

(2) *accord, advise, allocate, allot, allow, answer, appoint, ask, assign, assure, award, bequeath, bid, bring, call, carry, cause, cede, charge, concede, convey, cost, deal, deliver, demonstrate, deny, develop, drop, entrust, explain, extend, feed, flick, flip, forward, get, gift, give, grant, guarantee, hand, impart, inform, issue, keep, lease, leave, lend, loan, lose, mail, name, offer, owe, pass, pay, permit, play, pose, post, prescribe, present, promise, propose, provide, quote, read, recommend, refuse, render, sell, send, serve, set, show, sing, slip, submit, suggest, supply, take, teach, tell, throw, toss, vote, wish, write, yield*

Through intensive manual coding, the variable context was then restricted to those dative tokens where the alternating variant was grammatically acceptable with a near-identical meaning (following previous approaches, see Bresnan et al. [2007], and others). Hence, we weeded out observations involving particle verbs (e.g. *I gave her back her book*), passivized verbs (e.g. *the book was given to him*), elliptical structures (e.g. *I gave a rose to Mary and to Joe*), coordinated verbs, and clausal (any non-noun phrase) or non-overt constituents (e.g. *Tell them that we are coming*), as well as any case that was not variable with either a prepositional or ditransitive dative, namely:

- Beneficiary constructions (e.g. *We get them typed photocopies*)

- Constructions involving a spatial goal (e.g. *I bring my presents to school*)

- Idioms or fixed expressions (e.g. *Bring it directly to the boil*)

Pronominal constituents were retained since they are not exclusively restricted to one or the other variant. When it was unclear whether a particular dative token could be paraphrased by the other variant involving the same lexical constituents, we conducted a region-specific search in either Google or the Corpus of Global Web-based English (Davies 2013) to determine whether the paraphrase was attested. Region-specific searches ensured that expressions that might be idiomatic and hence non-alternating in variety A, would still be included in variety B if their alternating variant could be found in that variety. The dataset was then restricted to dismiss prepositional datives with extremely long recipients (>18 words) and

ditransitive datives with extremely long themes (>23 words) in order to eliminate those cases where the likelihood was relatively slim (if not zero) of the other variant occurring.

The dative dataset to be analyzed here spans 8549 interchangeable datives (see Table A in appendix for proportional distributions across varieties and Table B for the distribution across verbs).

2.3 Explanatory factors

In order to assess speakers' probabilistic grammar and to measure the combined impact of the constraints, the data was annotated for the factors commonly shown to influence dative choice. Apart from animacy and verb sense, all factors listed below were coded fully automatically. Automatic coding of factors across such a vast number of different varieties is not unproblematic – especially for the factor definiteness – and we discuss the issues arising from such an approach in the relevant section(s) below. Overall, the factors fall into two categories: language-internal and language-external constraints.

2.3.1 Language-internal factors

Length

End-weight – often measured in terms of constituent length – is one of the most influential factors when choosing a dative variant (Bresnan et al. 2007; Gerwin 2014: 48), and is used to refer to the general tendency in English to place short constituents before long ones (Behaghel 1909; Hawkins 1994).

To reduce multicollinearity in the data, we make use of a log transformed WEIGHTRATIO (number of characters in recipient divided by the number of characters in theme) instead of separate length measurements (Bresnan and Ford 2010: 174). Taking (3) as an example, the lengths of the theme and recipient are 20 and 23 characters respectively. We calculate the natural log of the weight ratio by *ln*(# of characters in recipient / # of characters in theme) = *ln*(23/20) = 0.140.

> (3) *Under the law, LTO should not issue [professional licenses] to [drug addicts or dependents].* <ICE-PHI:W2D-007>

Based on previous literature, we would assume the first constituent in either variant to be shorter than the second, that is, the smaller the weight ratio (<0) the more likely the ditransitive dative becomes, while the larger the weight ratio (>0) the more likely the prepositional dative.

Syntactic complexity

Wasow and Arnold (2003) illustrate the importance of syntactic complexity independent of length measurements. In a similar vein, Berlage (2014) shows in her study on noun phrase complexity that the presence or absence of postmodifying elements functions as an influential determinant of variation in that complex noun phrases tend to follow simple ones. We coded for a binary distinction between constituents heads with postmodification – coded as 'complex' – and those without

7

any postmodification – coded as 'simple' (exemplified in 4) (labels in the model: RecComplexity, ThemeComplexity).

(4) *[…] they promised [the non-Russian peoples of the vast tsarist empire]$_{complex}$ [self-determination]$_{simple}$* <ICE-SIN:W2E-004>

Given the literature (MacDonald 2013), we expect simple constituents to precede more complex ones in both the ditransitive and prepositional dative. In other words, a simple recipient increases the likelihood of a ditransitive dative while a simple theme increases the likelihood of a prepositional dative.

Pronominality

The pronominality of a constituent has been shown to influence the ordering of constituents in a number of syntactic alternation phenomena (see Bresnan and Ford 2010: 175). To avoid data sparseness, we made use of a binary distinction between pronouns (personal and impersonal) and non-pronouns (all other NP types) (exemplified in 5) (RecPron, ThemePron). More than half of the dative tokens in our dataset contain pronominal recipients and non-pronominal themes; roughly 21% contain two non-pronominal constituents.

(5) *One of the lighthouse keepers, David Lyall, sent [them]$_{pron}$ to [the dealer Henry Travers]$_{non-pron}$, […]* <ICE-NZ:W2B-021>

Following previous literature (Bernaisch et al. 2014; Bresnan et al. 2007; Wolk et al. 2013), we expect pronominal recipients to increase the likelihood of a ditransitive dative and pronominal themes to increase the likelihood of a prepositional dative.

Discourse givenness

Previous research (Arnold et al. 2000; Collins 1995: 43) has demonstrated the importance of information status with regard to the ordering of the constituents. In our study, discourse givenness was coded as a binary variable ('given' vs. 'new'). If the lemma of the head noun occurred in the 100 preceding words of discourse or was a personal pronoun, it was coded as 'given'. All other constituents were coded as 'new' (labels: RecGivenness, ThemeGivenness).

(6) *There is so much that can be got out of story-telling. It is not just to entertain **the child** but also to feed **him** with information of **his** cultural background, to teach **him** moral values and to enhance family cohesiveness. There are different types of stories and different ways of presenting them. To simplify things, stories could be categorized into family stories and classical stories. Family stories – these stories give [**the child**]$_{given}$ [an idea of himself and the family he belongs to]$_{new}$.* <ICE-SIN:W2D-020>

Findings from previous literature suggest given constituents precede new ones (Arnold et al. 2000; Bresnan et al. 2007). Hence, we would expect given recipients to increase the likelihood of a ditransitive dative and given themes to increase the likelihood of a prepositional dative.

Note that the automatic coding of givenness is not foolproof. If, for instance, the discourse participant(s) had mentioned a different child in the preceding 100-word window in example (6), automatic coding procedure would fail to notice that the child

in the dative variant has not been previously mentioned and is hence not discourse given. A similar caveat applies to pronominal constituents (generally coded as 'given') where the actual referent of the pronoun would need to be manually verified. Such a task is not feasible for the current study but clearly a desirable asset for future work.

Definiteness

In their experiment with American and Australian subjects, Bresnan and Ford (2010) report definiteness and length to be the main factors in their model of the dative alternation. In our study, themes and recipients were coded for definiteness (labels: RECDEFINITENESS, THEMEDEFINITENESS) following the procedure outlined in Garretson et al. (2004): Any constituent that allowed an existential reading in the context of *There is/are__* (as opposed to a deictic interpretation) was coded as 'indefinite' (for instance, bare nouns, indefinite pronouns, etc.). Constituents that contained a proper noun or pronoun as head or started with a definite article, demonstrative or any element tagged as definite in Garretson et al. (2004)[1], were coded as 'definite'.

> (7) *Jim Molyneaux is set to give [the Prime Minister]$_{def}$ [a piece of his mind]$_{indef}$ when the pair meet this week.*<ICE-IRE:W2E-002>

In accordance with the patterns found in Bresnan and Ford (2010) and others, we expect definite elements to precede indefinite ones. In other words, a definite recipient should increase the likelihood of a ditransitive dative while the likelihood of a prepositional dative is increased with a definite theme.

The automatic coding of definiteness is also problematic. For instance, Sand (2004) shows that speakers of L2 varieties tend to use the definite article in contexts where Standard English doesn't allow it, e.g. with generic nouns. However, the overuse of the definite article is not a phenomenon restricted to L2 varieties – it has also been observed in English spoken in Scotland, Northern England, South Wales, Ireland, and Southwest England (Filppula 1999: 69) as well as in Ireland, Newfoundland, Singapore, Jamaica, Orkney, and Shetland (Siemund 2013: 97). To verify the reliability of our automatic coding procedure for definiteness, we randomly selected 100 tokens – 50 tokens with a definite and 50 tokens with an indefinite recipient – from IndE, IrE, JamE, and SinE English, i.e. those varieties shown to have diverging usage patterns of definiteness markers as indicated above. We focused on the definiteness coding of the recipients because Sand (2004) remarks that it is animate NPs, such as recipients, that tend to be additionally marked with a definite article. After manually verifying the coding for false positives (NP marked as definite when indefinite) and false negatives (NP marked as indefinite when definite), we found five miscoded tokens in IndE (mostly with 'people'), two miscoded tokens in IrE, one miscoded token in JamE, and one miscoded token in SinE. While we are thus aware of the complications arising from our automatic coding procedure, the small number of miscoded noun phrases, the unfeasibility of manually verifying over 8,000 tokens, and the fact that the use of definite articles with generic nouns is not unique to L2

---

[1] Garretson et al. (2004) tag the following lexical items as definite: *the, this, that, those, these, her, his, its, my, our, their, your, all, both, each, either, every, most, neither, last,* and *next.*

Englishes all seem to legitimize usage of the automatic coding procedure of definiteness adopted in this paper. We are aware that any conclusion we draw based on the factor 'definiteness' will have to be tentative.

Person of recipient

Bresnan and Nikitina (2009) find a weak effect of recipient person (RECPERSON) on dative choice: speakers prefer the ditransitive significantly more if the recipient is 'local' (1st or 2nd person pronouns) than 'non-local'. Following their approach, we coded first person in- and exclusive and second person (non-)specific as 'local' (Cueni 2004: 11) since they are imminent participants in the speech act. All third person pronouns and non-pronominal NPs were coded as 'non-local'.

(8) *Because she brought the pictures and showed [them]$_{non-local}$ to [me]$_{local}$ and I said okay.* <ICE-PHI:S1A-036>

Animacy of recipient

Animacy (RECANIMACY) seems to have only subtle effects on word order in the dative alternation. Still, regional differences in the influence of recipient animacy have been shown to exist (Bernaisch et al. 2014: 19; Bresnan and Hay 2008). Adopting methods of earlier work (Wolk et al. 2013) and using a simplified version of the guidelines in Zaenen et al. (2004), we coded each constituent for a binary distinction between 'animate' (human, animal) and 'inanimate'. We excluded theme animacy as a factor because themes are overwhelmingly inanimate in our data (98.7%).

(9) *But before she could continue, the bell at the entrance tinkled and she tottered off to give [the new customer]$_{animate}$ [some service]$_{inanimate}$.* <ICE-SIN:W2F-006>

Assuming that animate referents are more accessible, that is 'easier', than inanimate ones (Branigan et al. 2008), we anticipate effects of animacy to align with the influence of previous constraints: animate recipients increase the likelihood of a ditransitive dative.

Concreteness of theme

Bresnan et al. (2007) include THEMECONCRETENESS in their model as a way to compensate for the simplified binary distinction of animacy. In our study, themes were hence coded for whether they referred to a concrete object and thus perceivable by one of the five senses following procedures in Garretson et al. (2004). Concreteness of the theme was coded using the annotation of verb semantics: if the whole verb phrase expressed 'transfer' or 'future transfer' (see Section *Verb Sense*), the theme was coded as 'concrete'. All other instances were coded as 'non-concrete'.

(10) *President Wee Kim Wee, who offered [cakes and drinks]$_{concrete}$ to [the 10 scouts who spent the morning sprucing up the Istana], said: […].* <ICE-SIN:W2C-010>

Based on the findings in Bresnan et al. (2007), we expect concrete themes to increase the likelihood of a prepositional dative and non-concrete themes to increase the likelihood of a ditransitive dative.

Verb sense

VERBSENSE was coded according to a five-level distinction (Bresnan et al. 2007; see also Levin 1993: 45–48), namely expressing 'transfer of possession of a concrete object' as in (11a), 'future transfer/possession of objects' as in (11b), 'prevention of transfer/possession', (11c), 'communication' of information (11d), or 'abstract', that is, all senses that cannot be categorized among the first four (11e).

(11)   a. *They give everybody a piece of paper.* (*give*.t)

           b. *Carl had promised her this car.* (*promise*.f)

           c. *My grandfather had denied my mother her chance of an education.* (*deny*.p)

           d. *She told me the whole story.* (*tell*.c)

           e. *You are paying me attention.* (*pay*.a)

2.3.2 Language-external factors

Variety

Each dative variant was annotated for VARIETY according to the ICE-corpus it was extracted from (see Section 2.1).

Register

In addition to variety, we also annotated for REGISTER using four coarse-grained register categories. This division is based on Koch and Oesterreicher (1985) who distinguish between graphic and phonemic code on the one hand, and informal versus formal on the other. This leaves us with four categories: SpokInf (spoken informal), SpokForm (spoken formal), WritInf (written informal), and WritForm (written formal) (see Table C in the appendix for a detailed list).

Corpus metadata

Each ICE-component provides specialized register classifications for each text, including mode (spoken/written), file identification (FILEID), text category, and subregister. These were used in the random structure of the regression analysis. We will elaborate more on this below.

3 Analysis and results

In order to analyze the contribution of each of the constraints on the choice of dative construction, we make use of mixed-effects logistic regression modeling. Logistic regression estimates the simultaneous effect of a set of factors on a binary outcome and gives an indication of the probability of observing one of the variants – in our case the prepositional dative (Gelman and Hill 2007; Hosmer and Lemeshow 2000; Pinheiro and Bates 2000). Mixed-effects modeling takes not only the combined set of factors into account, but also allows for so-called random effects – by-group idiosyncractic variation that is specific to the dataset. Using mixed-effects models enables us to better generalize beyond the particular data sample to, for instance, all verbs or speakers of a particular variety. The statistical analysis was conducted with the lme4 package in R (R Core Team 2014; Bates et al. 2015).

Our initial model included all the factors listed in Section 3.2 as fixed effects (apart from VERBSENSE), as well as the three-way interaction of each with the higher order interaction VARIETY * REGISTER (no other higher order interactions were considered). The numeric factor WEIGHTRATIO was standardized by subtracting the median and dividing the value by two standard deviations. Traditionally, numeric factors are centered around the mean (Gelman 2008). We opted for the median in this case because it captured the center of the distribution of the factor better than the mean. The purpose of this process was to reduce potential covariation among WEIGHTRATIO and other predictors, and to create a predictor whose estimated effect size is on a scale comparable to that of a binary categorical predictor (see Gelman 2008). Both VARIETY and REGISTER were coded using deviation contrasts where the proportion of responses for each level is compared against the grand mean across all levels (see Menard 2010: 97). Random factors included random intercepts for nested factors of corpus structure (e.g. file, text category, mode, subregister, etc.), verb, verb sense, recipient head, theme head, and verb-theme pairs, along with by-verb and by-VERBSENSE slopes for variety. Including the multiple levels of corpus structure as well as verb-, theme- and recipient-specific effects in the random component of the model is essential to ensure that the basic assumption of the non-independence of data points is not violated (Gries 2015: 99).

Model selection followed the backwards elimination procedure outlined by Zuur et al. (2009: 120–122). Starting from the maximal model, we first identified the optimal random structure, removing those random components that did not significantly improve model fit according to likelihood ratio tests. Next, we determined the optimal fixed effects structure in similar top-down fashion, by first removing any non-significant interaction terms, followed by non-significant main effects. The predicted outcome of the model was the log odds of the prepositional dative variant.

3.1 Model summary

The final model (shown in 12) includes a by-verb random intercept, an interaction of VERBLEMMA and VERBSENSE, heads of the theme, and a nested random effect of FILEID within text category within subregister within mode ('CorpusStructure'), as well as an interaction of VARIETY and REGISTER, RECPRONOMINALITY, THEMECONCRETENESS, and WEIGHTRATIO. None of the other initial interactions with REGISTER or VARIETY (in the beyond optimal model) came out as significant.

(12)   Dative model 1; Response = {ditransitive, prepositional }

$$\text{Response} \sim (1|\textsc{VerbLemma/VerbSense}) + (1|\text{ThemeHead}) +$$
$$(1|\text{CorpusStructure}) + \textsc{RecComplexity} + \textsc{RecGivenness} +$$
$$\textsc{ThemeComplexity} + \textsc{RecPerson} + \textsc{RecDefiniteness} +$$
$$\textsc{ThemePron} + \textsc{RecAnimacy} + \textsc{ThemeGivenness} +$$
$$\textsc{ThemeDefiniteness} + \textsc{Variety} * (\textsc{Register} +$$
$$\textsc{RecPron} + \textsc{ThemeConcreteness} + \textsc{WeightRatio})$$

Summary statistics for the model give a very good index of concordance *C* of 0.982, which indicates that the model is able to discriminate well between ditransitive and prepositional dative (values higher than .8 are indicative of a good fit to the data [Baayen 2008]). The classification accuracy of the model is 93.6%, which is significantly better than the baseline of 69% when always choosing the most frequent (ditransitive) dative ($p_{binom} < 0.001$).

For model validation, we randomly divided our dataset 100 times into a training set (consisting of roughly 75% of the data) and a test set (which contained the remaining 25%). We then fitted the model to each training set and calculated its predictions on the corresponding test set, measuring the accuracy of each of these 100 models in the probability of correctly predicted outcomes. Mean accuracy was 90.5%, which indicates a good model fit; the accuracy measures ranged from 89.5% for the poorest to 91.6% for the best model fit.

The condition index κ =11.8 points to existent but not overly harmful collinearity (Baayen 2008: 182). The variance inflation factor (VIF) for each of the factors indicates that much of the estimated variance of all higher order interactions with Variety is associated with the corresponding main effect. Hierarchical cluster analysis also reveals that multicollinearity mainly exists between the factor Variety and its interaction terms. Thus, extra caution will be exercised when interpreting the results.

3.2 Random effects

The random effects of verb lemma and head of the theme explain the most variance in the random structure of the model (see Table 1). The importance of the lexical units of theme and verb for dative choice has been demonstrated previously (Bresnan and Ford 2010: 202).

Table 1. Estimated variances and standard deviations of random effects in the model.

Groups	Variance	Std. Dev
Mode:GenreCoarse:GenreFine:FileID	0.6304	0.7940
Mode:GenreCoarse:GenreFine	0.1019	0.3193
ThemeHead	1.6968	1.3026
VerbLemma	6.4453	2.5388
VerbLemma: VerbSense	0.4238	0.6510

In our model, the themes *fee, lead, choice, problems*, and *one* show a very strong preference for the ditransitive dative (13), while the opposite is the case for *it*, *birth*, *attention*, *evidence*, and *effect* (14). These preferences, especially in the case of prepositional datives, are possibly due to tokens that were considered idiomatic in some varieties but not all (e.g. *give birth to someone* was considered interchangeable in BrE, see Section 2.2).

(13)   *pay the government a fee* <ICE-SIN:S2B-006>
(14)   *the Prosecution had given evidence of his statement to the Investigating Officer* <ICE-SIN:S1B-069>

Regarding VERBLEMMA, we find that *wish, promise, permit, allow, tell*, and *teach* are attracted to the ditransitive dative (15), while *explain, submit, get, recommend, bequeath*, and *pose* display the opposite preference (16).

(15)   *wish him luck* <ICE-PHI:S1B-030>
(16)   *explained the situation to her* <ICE-IND:W2F-002>

Zooming in on VERBSENSE, the strongest preferences for the ditransitive dative have *give*.a, *show*.c, *pass*.t, *hand*.a, and *grant*.a (17), while *show*.a, *hand*.t, *leave*.c, *give*.t, and *offer*.c favor the prepositional dative (18).

(17)   *gave us all a big hug* (abstract) <ICE-CAN:S1A-001>
(18)   *gave my card to Aunt Ellen* (transfer) <ICE-HK:W1B-012>

Table 2 gives the most extreme adjustments to the intercept by VERBSENSE: positive values indicate a preference for the prepositional dative, negative values a preference for the ditransitive dative (cf. also Levin 1993; Gerwin 2014; Wasow and Arnold 2003). The diverging preferences of *give* meaning 'transfer of concrete objects' and *give* in the abstract sense (see Table 2) indicates that the abstract sense of this verb lemma drives *give*'s overall preference for the ditransitive.

Table 2. Most extreme adjustments to the intercept for VERBSENSE. Positive values indicate a preference for the prepositional dative, negative values indicate a preference for the ditransitive dative.

VerbSense	adjustments to the intercept
show.a	0.533
hand.t	0.502
leave.c	0.397
give.t	0.391
offer.c	0.388
grant.a	- 0.415
hand.a	- 0.551

pass.t	- 0.561
show.c	- 0.620
give.a	- 0.805

Table 2 indicates that abstract senses of both *give* (19) and *hand* (20) show a preference for the ditransitive dative while the transfer of concrete objects as in (21) and (22) is preferably expressed with a prepositional dative.

(19)   *give the new customer some service* <SIN:W2F-006>
(20)   *But handing the newly-privatised steelmaker a UK monopoly*
                  <GB:W2C-007>
(21)   *give the flowers to the First Lady* <IRE:S2A-017>
(22)   *handed a letter to the Principal* <JA:S1B-077>

Krifka (2003) notes that the prepositional dative is lexically constrained: the theme undergoes movement, that is, the theme is transferred from the subject into the possession of the recipient. Such movement is only possible with concrete objects (and not abstract entities such as *ideas*), hence we can expect *give* expressing 'transfer of possession' to occur more frequently in the prepositional dative. At the same time, the adjustments of other verb senses in Table 2 (e.g. *show*.a as in *show our appreciation to them*) indicate that the preferences of the verb per se for either variant might overrule semantic considerations. To disentangle the various effects of both verb and verb sense, a more in-depth analysis is required, which is beyond the scope of the current study.

3.3 Main effects

Table 3 summarizes the coefficients of the main factors in the model. The column labelled $\widehat{\beta}$ indicates the estimates of the coefficients on a logit-scale. Positive values indicate a preference for the prepositional dative (the predicted outcome), negative values indicate a preference for the ditransitive dative. SE specifies standard errors. The results of the statistical analysis can be summarized as follows.

First, the constraints in the model have the expected effect on the choice of dative variant given the literature. For instance, the longer the recipient is in relation to the theme, the greater the odds for a prepositional dative. Pronominality also has the expected effect in that the ditransitive dative increases in likelihood if the recipient is pronominal, and the prepositional dative becomes more likely if the theme is pronominal. Similarly, the effects of animacy (of the recipient), accessibility, person (of the recipient), definiteness, and complexity are congruent with the findings of previous research: Whenever a constituent is given, local, animate, definite, or simple, the model indicates that language users tend to place it first in the ordering of the constituents. In other words, if the recipient is given, local, animate, definite, or simple, the ditransitive is the preferred option across the board. If the theme is given, animate, definite, or simple, the prepositional dative increases in likelihood.

Second, the overall likelihood of a prepositional or ditransitive dative does not vary cross-regionally apart from IndE and CanE where the probability of either dative

variant deviates from the overall mean significantly. In CanE, the ditransitive dative is more likely, while in IndE, the odds of a prepositional dative are much higher compared to all other varieties. A similarly high rate of prepositional datives in IndE has also been observed by Mukherjee and Hoffmann (2006) and De Cuypere and Verbeke (2013). Taking into account the contextual constraints in connection with these constructional preferences, we might conclude that CanE has the highest preference for a prototypical ditransitive dative (simple, local, animate, pronominal, and definite recipient with a simple, non-pronominal, and indefinite theme) while IndE prefers the prepositional dative in the same context, more so than any other variety.

Table 3. Main effects of individual factors in the model. Model predictions are for the prepositional dative (only significant factors shown).

Factor	$\hat{\beta}$	SE	p
(INTERCEPT)	2.525	0.405	<0.001
RECIPIENT COMPLEXITY			
simple ⇨ complex	0.898	0.204	<0.001
THEME COMPLEXITY			
simple ⇨ complex	-0.692	0.164	<0.001
RECIPIENT PERSON			
local ⇨ non-local	0.882	0.175	<0.001
RECIPIENT ACCESSIBILITY			
given ⇨ new	0.388	0.130	<0.01
RECIPIENT ANIMACY			
animate ⇨ inanimate	0.994	0.140	<0.001
THEME PRONOMINALITY			
non-pronoun ⇨ pronoun	1.552	0.468	<0.001
RECIPIENT PRONOMINALITY			
pronoun ⇨ non-pronoun	1.945	0.191	<0.001
RECIPIENT DEFINITENESS			
definite ⇨ indefinite	0.556	0.144	<0.001
THEME DEFINITENESS			
indefinite ⇨ definite	0.696	0.126	<0.001
WEIGHT RATIO (rec/theme)	2.950	0.230	<0.001
VARIETY			
all ⇨ CanE	-1.586	0.365	<0.001
all ⇨ IndE	0.919	0.256	<0.001

Next, the importance of each factor in the model was calculated by measuring the decrease in goodness-of-fit when leaving the factor out of the model. This was done with the Anova() function in the car package in R. As Figure 1 indicates, WEIGHTRATIO and RECPRON are the two most important factors in the model, followed by VARIETY and RECANIMACY. THEMECONCRETENESS, THEMEGIVENNESS, and REGISTER are not significantly contributing to the model.
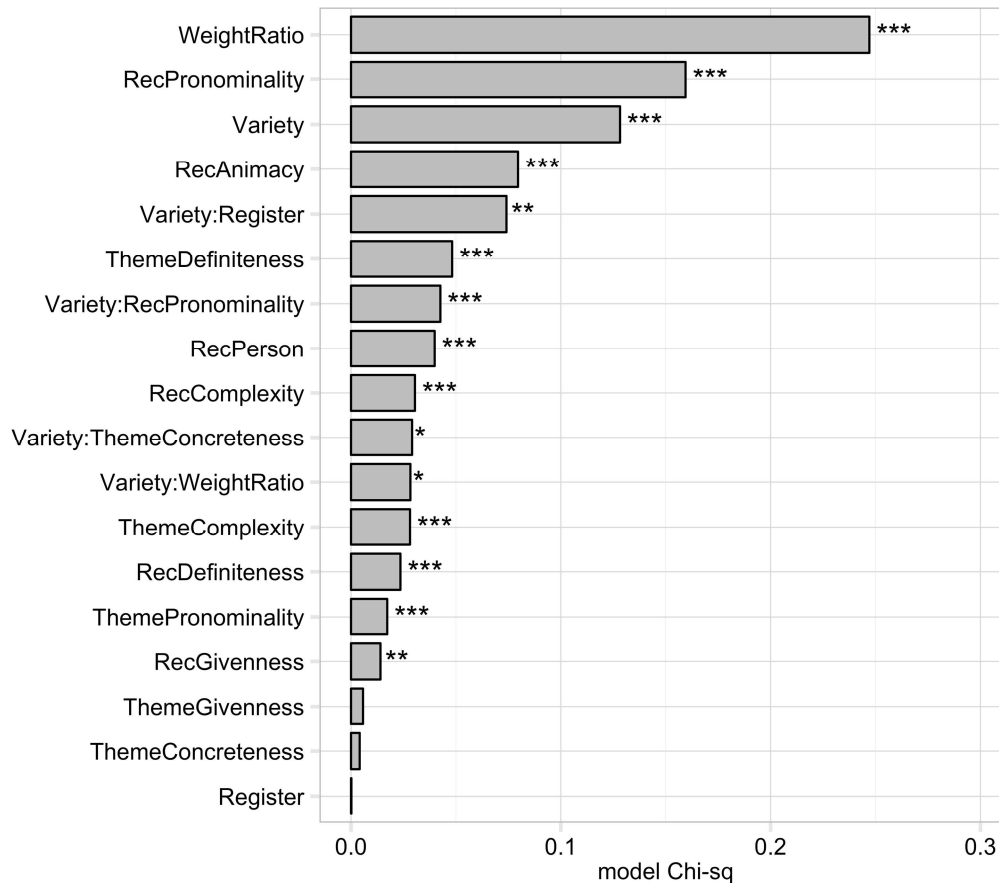
Figure 1. Variable importance (decrease in model goodness-of-fit if factor removed) of all factors in the model. Significance is indicated by $p<0.05$ = *; $p<0.01$ = **; $p<0.001$ = ***.

3.4 Interactions

Table 4 reports all significant model coefficients of the interaction between VARIETY and language-internal conditioning factors. The coefficients indicate that higher-order terms have a similar influence on dative choice as the respective main effects. However, the magnitude of the influence varies across some regions. If the coefficient estimates of a main factor (for instance, RECPRONOMINALITY) and its interaction term (VARIETY * RECPRONOMINALITY) have the same signs, the effect is stronger in that specific variety (compared to all other varieties). If the coefficient estimate of a main factor and its interaction term have opposite signs, the effect of that factor is weaker in that specific variety.

Zooming in, the interaction terms indicate that

- recipient pronominality has a greater effect in Indian and Canadian English and a weaker effect in Jamaican English,

- the effect of end-weight (short before long) is weaker in IndE and stronger in JamE (compared to all other varieties),

- the effect of THEMECONCRETENESS in CanE follows the expected pattern given the literature, namely that concrete themes are placed first in the ordering, and

18

- the varying effect of REGISTER across varieties points towards stylistic differences in New Zealand, Irish, Jamaican, and Hong Kong English that do not follow the overall trend.

Table 4. Interaction effects in the model between VARIETY and REGISTER and language-internal factors. Model predictions are for the prepositional dative (only significant factors shown).

Factor	β	SE	p
VARIETY : RECIPIENT PRONOMINALITY			
CanE + non-pronoun	0.902	0.402	0.025
IndE + non-pronoun	1.108	0.353	0.002
JamE + non-pronoun	-1.253	0.402	0.002
VARIETY : WEIGHT			
IndE	-1.080	0.452	0.017
JamE	1.960	0.606	0.001
VARIETY : THEME CONCRETENESS			
CanE + concrete	1.250	0.397	0.002
VARIETY : REGISTER			
IrE + SpokForm	0.692	0.278	0.013
IrE + SpokInf	-0.604	0.287	0.035
HKE + SpokInf	0.679	0.244	0.005
HKE + WrittenForm	-0.912	0.293	0.002
HKE + WrittenInf	0.566	0.220	0.010
JamE + SpokInf	-0.703	0.312	0.024
JamE + WrittenForm	0.873	0.433	0.044
NZE + WrittenForm	0.673	0.295	0.023

Setting aside stylistic differences across varieties of English for the present paper, the three interaction terms THEMECONCRETENESS, WEIGHTRATIO, and RECPRONOMINALITY merit further exploration.

Out of these three higher order terms, THEMECONCRETENESS is the only factor whose main effect is not significant in our model (β = -0.473, SE > 0.31, p>0.128) (see Figure 1). Only in CanE does the effect of THEMECONCRETENESS significantly influence the choice of dative variant: the likelihood of a prepositional dative increases in case of concrete themes. The preference thus follows MacDonald's (2013) processing principles in that 'easy' elements (in this case concrete themes) are posited first.

Generally, WEIGHTRATIO has a similar effect across all varieties: an increase in WEIGHTRATIO increases the likelihood of a prepositional dative since the recipient

increases in length in comparison to the theme. However, the increase in likelihood, that is, the effect size of the constraint, differs in its strength across varieties. In IndE, the influence of WEIGHTRATIO is weaker and the probability of using a prepositional dative is not increased by the effect of end-weight as much as it is in the other varieties. This could be attributed to an overall greater likelihood of the prepositional dative in IndE to start with. In contrast, length differences between recipient and theme have a stronger effect for speakers of JamE; they show a stronger preference for the ditransitive when WEIGHTRATIO decreases (i.e. when the theme increases in length compared to the recipient) and a stronger preference for the prepositional dative when WEIGHTRATIO increases (i.e. when the recipient increases in length compared to the theme).

Similarly to other factors, the effect of recipient pronominality does not differ across varieties with regard to the choice of variant: pronominal recipients cue a ditransitive dative, non-pronominal recipients cue a prepositional dative. However, the strength of this constraint varies across regions. The effect is stronger in Indian and Canadian English and weaker in Jamaican English than in the average speech community. In other words, the probability of a prepositional dative is higher in IndE and CanE when the recipient is non-pronominal than when it is pronominal compared to all other varieties. In JamE, however, the likelihood of a prepositional dative is smaller when the recipient is non-pronominal than when it is pronominal compared to all other varieties.

4 Discussion

Through a multivariate analysis of the dative alternation across nine geographically diverse varieties of English, we have uncovered two patterns of note. On the one hand, we find that the "core" probabilistic grammar underlying the dative alternation is largely stable across these different varieties. On the other hand, we observe that small differences nonetheless exist in the degree of sensitivity that speakers of different varieties demonstrate towards some of the factors that constrain dative choice. The aim of this study was to explore the extent to which international varieties of English share a probabilistic grammar, and to investigate which linguistic factors, if any, are amenable to regional differences. Our findings suggest that a shared underlying pattern of usage in the dative alternation prevails across varieties of English: language users tend to opt for the syntactic variant where a shorter constituent is followed by a longer one, an animate constituent is followed by an inanimate one, and so on. While such stability at the level of syntax may not seem surprising at first, it is rather striking given the considerable degree of variability we find among these varieties at other levels of linguistic structure, e.g. phonology and lexis (see Schneider 2007 for review). Notwithstanding this overall stability in speakers' probabilistic grammar, the effect size of three constraints, namely theme concreteness, length, and recipient pronominality, differs across varieties. By assuming a usage-based model of language, we have thus highlighted that general processes of language production and comprehension (e.g. short before long) vary regionally. Let us therefore examine these processes in more detail.

According to MacDonald (2013), incremental language production can be explained by the interplay between three factors: *Easy First*, *Plan Reuse*, and *Reduce Interference*. An *Easy First* bias in speech production and planning leads a speaker

to select early those linguistic units (words, phrases, etc.) that are easier to retrieve from long-term memory. "Easier" in this sense is typically characterized as frequent, shorter, less syntactically complex, conceptually entrenched, and given in the discourse (MacDonald 2013: 3). At the same time, speakers tend to reuse previously heard syntactic plans and closely related structures that they retrieve from long-term memory in a process that MacDonald (2013: 4) calls *Plan Reuse*. The third process, *Reduce Interference*, refers to the minimization of interference from a semantically closely related lexeme during the utterance of a word by increasing the number of linguistic units between the two words. MacDonald argues that these three principles of language production and planning jointly govern utterance form. For instance, animate nouns have been shown to be easier to retrieve from memory than inanimate nouns, hence the tendency for animate agents to be realized in subject position, e.g. *The boy smashed the window* (Bock 1982). At the same time, passive sentences, e.g. *The window was smashed by the boy*, often involve inanimate subjects, as patient arguments tend to be inanimate. As the forces of *Easy First* (animate first) and *Plan Reuse* (priming for passive voice with a passive biased verb) might conflict in the choice of passive vs. active, we expect utterance planning time to increase for passive voice – a prediction supported by evidence (Ferreira 1994). These principles also generate the link from "individual-level behaviors" to "population-level linguistic phenomena" (Scott-Phillips and Kirby 2010: 411). By summing over millions of utterances and language producers, the consistent interplay between the three principles creates statistical regularities in language usage (MacDonald 2013: 5).

The outcome of this interplay is reflected in our model. On the one hand, speakers tend to choose that dative variant where the first constituent fulfills all aforementioned requirements of being "easy". *Easy First* is thus a principle that combines the various influences of the factors in our model, such as length, frequency, givenness, and concreteness. Since the combination of these factors constitutes language users' probabilistic grammar, we can assume that the prevailing stability in effect direction that we observe across regional varieties of English can be attributed to the principle of *Easy First*. On the other hand, while *Easy First* seems to strengthen stability of speakers' probabilistic grammar since, irrespective of the linguistic material, easy comes first, *Plan Reuse* constantly reinforces the regularization of linguistic input and can ultimately strengthen diverging statistical patterns of use. Since *Plan Reuse* (as well as the other two principles) relies on the linguistic material at hand, changes in this material can result in differences in the statistical regularities that speakers make and eventually in diverging probabilistic grammars. As a consequence of these diverging statistical regularities, the strength of the effects of the individual predictors that modulate these regularities change as well. Hence, which (syntactic) variant is cued and thus easier for speakers to produce, or entrenched enough to be reused in language planning may not necessarily be the same for individual speakers but will depend on their linguistic experience (see Ellis 2002: 145). At this point, we cannot profess to be able to provide exhaustive explanations for the regional variation in the strength of some predictors that we observe. Rather, we would like to suggest three somewhat speculative but plausible explanations how such variation might arise as the result of (random but expected) modifications in the cue strength of predictors through the constant reinforcement of structural patterns by *Plan Reuse*.

First, linguistic experience and input varies due to the general conditions of language or dialect contact, which naturally vary from region to region as speakers of different dialects and/or native languages interact in their new environment. Such contact leads to the emergence of localized linguistic forms on the level of syntax and morphology in the formative stages of New Englishes – a process that Schneider (2007: 44) calls "structural nativization". Structural nativization generally results in new combinations of syntactic constructions with lexical items. In cases where new lexical items occur frequently enough in these syntactic constructions, the abstractions of regularities that speakers make (in order to be able to generalize beyond the linguistic input) leads to changes in the constraints governing language structure. These constraints are, in turn, learned during processes of language acquisition (Ellis 2002: 144) and become part of speakers' grammatical knowledge (Gahl and Garnsey 2004). In short, changes in lexical choices in syntactic variants can influence the impact that the underlying cues have on syntactic variation.

Second, processes involved in second language acquisition and substrate influences may also shape users' choices in a given context. We note that some of the largest deviations in individual factor effects in our model occur in the L2 varieties; it is in IndE and JamE where the effect of weight ratio and recipient pronominality deviates significantly from the global average. Effects of second language acquisition impact not only structural nativization processes but also lead to an increased usage of the more transparent syntactic variant – in our case the prepositional dative (Leufkens 2013: 345, 346; see also Siegel et al. 2014). This in turn can lead to changes in the strength of specific cues as variants are used by L2 speakers in contexts where L1 speakers would not use them. For instance, Mukherjee and Hoffmann (2006) explain the large proportion of prepositional datives in IndE by drawing attention to the fact that *give* frequently occurs as a light verb in that variety, as in (23). Also more generally, the kind of verb-complementation profiles that *give* is used with in IndE differs from British English (Mukherjee and Hoffmann 2006: 154–155). De Cuypere and Verbeke (2013: 180–181) further suggest that the popularity of light verbs in IndE is due to their high frequency in the substrate languages. In addition, the necessity of an explicit dative case marker in the Indian vernacular languages (e.g. *ko* in Hindi as in (24)) might have increased the use of the prepositional dative in IndE in contexts diverging from L1 usage (see also Haspelmath 2013).

(23)   *give a satisfactory and convincing explanation to any one of them*
       <ICE-IND:W1B-016>

(24)   Hindi
       maiṃ   apnī   bahan=ko   yah   kitāb   deti hūṃ.
       I   my   sister=to.RECIPIENT the   book.THEME   give.
       'I give my sister the book.'
       (De Cuypere & Verbeke 2013:181)

A similar substrate effect can be observed in the contact situation between Jamaican Creole and Jamaican English. According to the Atlas of Pidgin and Creole Language Structure (APiCS) Online (Michaelis et al. 2013), speakers of Jamaican Creole use ditransitive constructions as in (25) with verbs of physical transfer of possession

followed by recipient and theme without any additional grammatical marking on the recipient (contrary to what one would expect in Standard English).

(25)   Jamaican Creole
         Di   uman gi   di   bwai   di   fuud.
         DET   woman give DET   boy.RECIPIENT DET   food.THEME.
         'The woman gave the boy the food.'
         (Farquharson 2013)

Bruyn et al. (1999: 330) provide examples from several other creoles that highlight that the double-object construction with an unmarked recipient constitutes the most frequent if not only option, irrespective of whether the recipient occurs before or after the theme. The high frequency of double-object constructions seems to be inherent to creoles despite the fact that not all lexifier languages had those double-object constructions to begin with.

Since in both India and Jamaica, most speakers acquire the substrate language as their first language (see Meade [2001: 175–176] for the Jamaican context), transfer effects would result in ditransitive and prepositional datives being used in different contexts in both IndE and JamE. In addition to the evidence provided here, transfer of cue strength (that is, the effect size of constraints) from the first language can lead to gradient shifts in linguistic preferences and changes in speakers' probabilistic grammar (MacWhinney 1997: 129).

Third, the variation we observe might not only be due to changes in contact-induced lexical variation or substrate effects but also result from constructional and/or semantic changes that arise in the course of everyday language usage. As speakers use the ditransitive or prepositional datives in different ways in different contexts, the range of meanings associated with either variant – their semasiological profiles – will likely change, and these changes are reflected in the lexical items that fill their syntactic slots. This entails that the range of different lexical items might be more diverse in one variant as compared to the other, and that this difference in diversity (i.e. the degree of semasiological heterogeneity) might differ from variety to variety. The latter hypothesis is supported by studies that show that universal processes of language acquisition can influence the type frequency in syntactic variants. For instance, research in first language acquisition has shown that up to a certain age, children associate the use of the ditransitive dative with specific lexical items and do not abstract syntactic constructions beyond the input they receive (Dodson and Tomasello 1998: 606). Similarly, second language learners tend to associate the use of the ditransitive dative with specific lexical items (for instance, pronouns) or certain discourse contexts while the use of the prepositional dative is not as semantically restricted (McDonough 2006: 193, 194).

In order to find supporting evidence of the lexical specificity of dative variants, we performed a distinctive collexeme analysis on the dative data to measure the strength of association on the paradigmatic and syntagmatic level among lexical items and either of the two dative variants (see Stefanowitsch and Gries 2003).
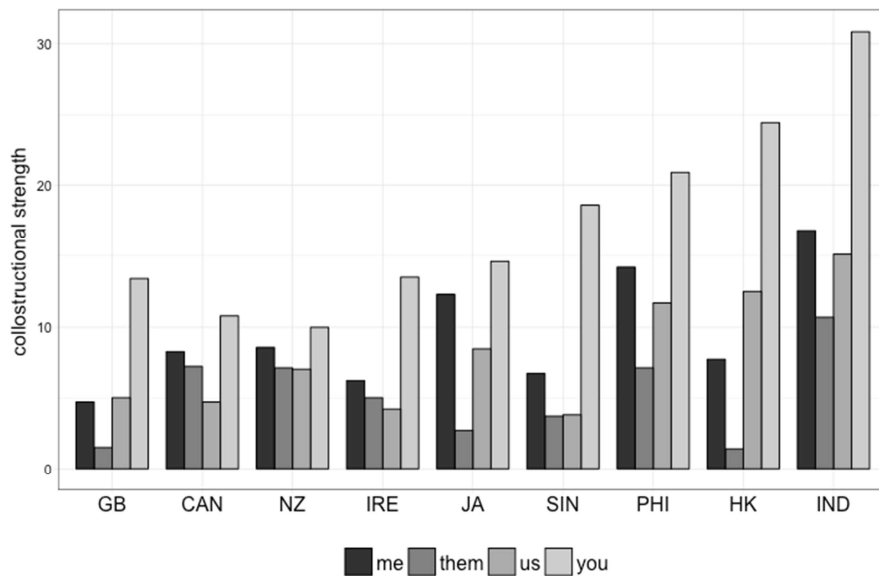
Figure 2. Collostructional strength between the four most frequent recipients and the ditransitive variant across nine national varieties of English.

Figure 2 shows that second language learners in postcolonial varieties of English (JamE, SinE, PhilE, HKE, and IndE) associate pronominal recipients more strongly with the ditransitive construction than other varieties of English, that is, the ditransitive dative is lexically more specified in those varieties.

In the end, the constant reinforcement of such diverging usage patterns through the principle of *Plan Reuse* can result in diverging statistical regularities. For instance, in IndE, we observed that recipient pronominality is a very strong cue for the choice of dative variant, which is reflected in the fact that speakers of IndE are exposed to a large number of ditransitives with a pronominal recipient. Cross-varietal differences with regard to the variants' lexical profile can thus lead to deviations in the underlying factors that constrain linguistic variation.

That the operation of linguistic constraints is limited by lexical considerations is nothing new (see Bybee and Hopper 2001: 2). We have shown, however, that the strength of these linguistic constraints varies subtly between different varieties of the same language. What is more, recipient pronominality and length are not only the two factors that differ significantly across varieties, they are also the most influential constraints in dative choice (see Figure 1). Our findings thus suggest that the factors that emerge as the most amenable to probabilistic indigenization are also the most prominent cues. Hence, even though we might never be able to fully predict which factors in linguistic variation might deviate across different dialects or varieties, we can assume that the most reliable cues are the ones most probably prone to change in strength (see also Grafmiller [2014] for evidence of this tendency across registers). Why is it that CanE, JamE and IndE exhibit the greatest difference? While the forces of structural change suggested here might point us into the direction of the reasons for different degrees of probabilistic indigenization across varieties, we cannot conclusively answer that question with the data currently at hand.. Furthermore, cross-constructional comparison reveals that the set of varieties that diverge the most from the global mean is not consistent (see Grafmiller et al. 2016). If – as suggested above – the semasiological profile of the variants influences the

statistical abstractions that speakers make, and assuming that the semasiological profiles differ from variant to variant and from alternation to alternation, we can expect construction-specific statistical deviations in the influence of different predictors across varieties. Research is under way right now to find a rationale behind these patterns of cross-constructional deviations but no final conclusion can be offered yet.

In contrast to previous research (Bresnan and Hay 2008; Bresnan and Ford 2010), recipient animacy did not turn out to be regionally variable in our study. One reason might be the greater lexical diversity of our dataset. Unlike prior studies, for instance Bresnan and Hay (2008), our dataset contains a large number of different verbs and not just the prototypical *give*. To explore this further, we performed a supplementary analysis on a dataset restricted to tokens with the verb *give* and included only those factors that were statistically significant in Bresnan and Hay's original study.[2] The GIVE model of our reduced dataset shows a significant change in the effect of RECANIMACY in IndE and HKE compared to all other varieties. As Figure 3a illustrates, the effect of RECANIMACY on the choice of dative variant disappears almost completely in HKE. In IndE by contrast, an inanimate recipient raises the likelihood of a prepositional dative far more than in any other variety. For comparison, Figure 3b shows the effect of RECANIMACY by variety reported in our full model containing all dative verbs. Apart from the fact that an inanimate recipient generally increases the probability of a prepositional dative, no cross-varietal differences can be discerned. These supplementary findings indicate that pooling over a large number of verbs seems to obscure potential meaningful differences in that some effects seem to be sensitive to the lexical items that are used as syntactic constituents. Since those verb-specific sensitivities vary across varieties, these results strengthen our argument that frequent co-occurrence of, in this case, *give* with certain lexical (animate) recipients in a subset of the varieties lead to deviations in the underlying stochastic patterns that abstract from regularities of linguistic input.

---

[2] These are, in detail, a random slope for SPEAKER, the main factors RECGIVENNESS, separate length measurements, the interaction of THEMEGIVENNESS with VERBSEMANTICS, the interaction of RECANIMACY and VARIETY, and recipient and theme pronominality as coded in Bresnan and Hay (2008: 18).
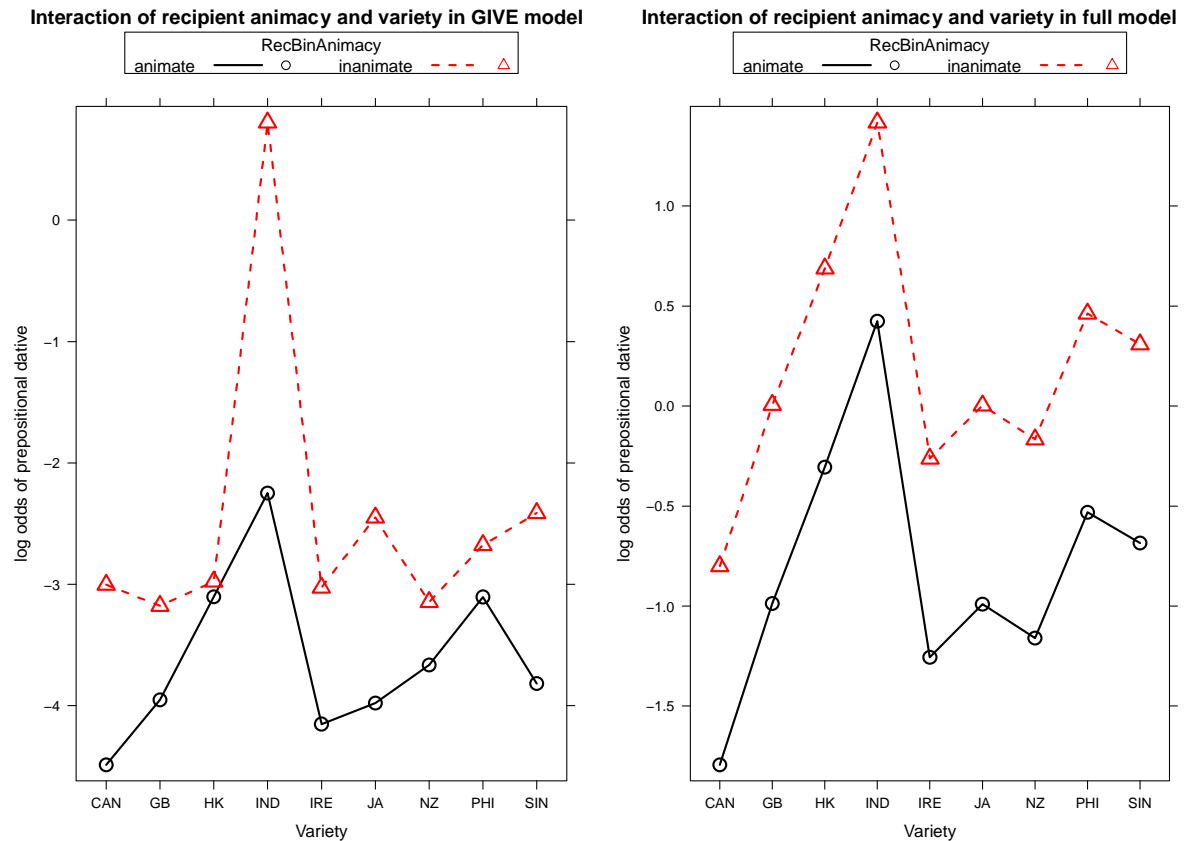
Figure 3a (left). Variability of effects of recipient animacy across VARIETY for the verb *give.* Predicted variant is the prepositional dative. Summary statistics of the GIVE model indicate a very good index of concordance C = 0.974 and multicollinearity is nearly absent (k = 7.4, max VIF = 4.3).

Figure 3b (right). Variability of effects of recipient animacy across VARIETY for all verbs. Predicted variant is the prepositional dative.

The results of our study tie in with recent research in Cognitive Sociolinguistics in that we view variation in language from a cognitive as well as socio-cultural perspective. Both the processes of *Easy First* and *Plan Reuse* are well situated in the cognitive domain and their interplay results in the statistical abstractions that we observe in our model. These cognitive processes are constantly reinforcing the underlying structural patterns that have been reshaped by contact-induced lexical variation, substrate effects and constructional changes as a result of the socially interactive nature of language. The emergence of cognitive indigenization can thus only be adequately recognized if we take both the social as well as cognitive nature of language into account.

Finally, a caveat is in order here. To test the cognitive plausibility of statistical models such as the current one, corpus-based analyses have been comparing the models' performance with the prediction accuracy of native speakers obtained in experimental settings (see Klavan and Divjak 2016: 357). Even though such studies show that language users' implicit knowledge of variation patterns reflects on the whole the usage probabilities attained from statistical models much closer than expected (e.g. Bresnan 2007; Bresnan and Ford 2010), this is not always the case. Comparisons often reveal marginal but existent differences between observational

aggregate data and behavioral individual data. We thus have to be circumspect when drawing conclusions about speakers' linguistic knowledge based on the results from regression models. While regression techniques might not necessarily mirror the cognitive reality in speakers' mind with 100% accuracy, they can still be used to assess the relative weighting of simultaneously interacting constraints on language performance and are thus valid and cognitively realistic approximations (Klavan and Divjak 2016: 379). Truly cognitive models (e.g. memory-based learning, naïve discriminative learning) are currently being developed (see Milin et al. 2016) which will certainly enhance our grasp of speakers' grammatical knowledge in future work.

5 Conclusion

The present study set out to explore the extent to which language users' probabilistic grammar varies regionally. Using mixed-effects logistic regression and adopting a large-scale comparative perspective, we investigated the effects of constraints that influence the choice between the two syntactic variants in the well-known dative alternation (*I give Mary a book* vs. *I give a book to Mary*) across nine (post-colonial) varieties of English. Our results illustrate that on the one hand, stability in probabilistic grammars prevails across English language users of diverse regional and cultural backgrounds (see Heller et al. in press). On the other hand, traces of indigenization are found in those contexts where shifting usage frequencies of language-internal variation seem to have led to regional differences between users' probabilistic grammar(s). These differences reflect the dynamics of exposure to (more or less subtly) different linguistic input since variability in frequencies emerges when successive generations of speakers are exposed to different sets of dative exemplars. This variability in frequencies is a result of natural variation in the frequencies of specific lexical items, features and/or syntactic structures, and can be attributed to the influence of second language acquisition, first language substrate effects, and universals of language and dialect contact. Constant reinforcement of diverging frequency patterns eventually results in differences on a more fine-grained level of language users' grammatical knowledge, namely in the underlying stochastic patterns that modulate linguistic choices. Our study thus emphasizes the importance of synthetizing the cognitive with the social dimension in explaining linguistic variation on the very subtle level of stochastic regularities.

We admit that the limitations of the current study to only one syntactic alternation might restrict further generalizations that we could make. It is thus desirable of future work to advance our understanding of the effect of cognitive indigenization by widening the perspective to other (syntactic) alternations and to add to the methodological toolbox by extending the analysis to include memory-based learning techniques and rating task experiments. Such experiments are increasingly being employed as a means of externally validating results from corpus-based analyses and assessing the cognitive reality of different theoretical models (see Bresnan 2007; Divjak et al. 2016). We would thus expect the results of future rating task experiments to correspond with the predictions of our corpus model.

Besides extending the analysis to other (syntactic alternations), future studies might also need to overcome possible limitations of the datasets that researchers investigate. Our study has shown that potential meaningful differences between

varieties could be obscured if the analysis employs a (in various ways) restricted dataset. Future research thus needs to consider linguistic variation from multiple angles, that is, modeling over large-scale datasets as well as closer syntactic and semantic analyses of particular lexical items and constructions. With such an expanded perspective, we are in a position to enhance our knowledge of the granularity of syntactic structure, that is, the extent to which "grammar" is tied to individual (micro-) constructions and lexical items.

Acknowledgment

References

Arnold, Jennifer E., Anthony Losongco, Thomas Wasow & Ryan Ginstrom. 2000. Heaviness vs. newness: The effects of structural complexity and discourse status on constituent ordering. *Language* 76(1). 28–55.

Baayen, R. Harald. 2008. *Analyzing linguistic data: A practical introduction to statistics using R*. Cambridge, New York: Cambridge University Press.

Bates, Douglas, Martin Mächler, Benjamin M. Bolker & Steven C. Walker. 2015. Fitting linear mixed effect models using lme4. *Journal of Statistical Software* 67(1). 1–48.

Behaghel, Otto. 1909. Beziehungen zwischen Umfang und Reihenfolge von Satzgliedern. *Indogermanische Forschungen* 25. 110–142.

Berlage, Eva. 2014. *Noun Phrase Complexity in English*. Cambridge: Cambridge University Press.

Bernaisch, Tobias, Stefan Th. Gries & Joybrato Mukherjee. 2014. The dative alternation in South Asian English(es): Modelling predictors and predicting prototypes. *English World-Wide* 35(1). 7–31.

Bock, Kathryn. 1982. Toward a Cognitive Psychology of Syntax: Information Processing Contributions to Sentence Formulation. *Psychological Review* 89(1). 1–47.

Branigan, Holly P., Martin J. Pickering & Mikihiro Tanaka. 2008. Contributions of Animacy to Grammatical Function Assignment and Word Order during Production. *Lingua* 118(2). 172–189.

Bresnan, Joan. 2007. Is syntactic knowledge probabilistic? Experiments with the English dative alternation. In Sam Featherston & Wolfgang Sternefeld (eds.), *Roots: Linguistics in Search of Its Evidential Base*, 75–96. Berlin: Mouton de Gruyter.

Bresnan, Joan, Anna Cueni, Tatiana Nikitina & Harald Baayen. 2007. Predicting the Dative Alternation. In Gerlof Boume, Irene Krämer & Joost Zwarts (eds.), *Cognitive Foundations of Interpretation*, 69–94. Amsterdam: Royal Netherlands Academy of Science.

Bresnan, Joan & Marilyn Ford. 2010. Predicting syntax: Processing dative constructions in American and Australian varieties of English. *Language* 86(1). 168–213.

Bresnan, Joan & Jennifer Hay. 2008. Gradient grammar: An effect of animacy on the syntax of give in New Zealand and American English. *Lingua* 118(2). 245–259.

Bresnan, Joan & Tatiana Nikitina. 2009. The Gradience of the Dative Alternation. In Linda Uyechi & Lian Hee Wee (eds.), *Reality Exploration and Discovery: Pattern Interaction in Language and Life*, 161–184. Stanford: CSLI Publications.

Bruyn, Adrienne, Pieter Muysken & Maaike Verrips. 1999. Double-object constructions in the creole languages: Development and acquisition. In Michel DeGraff (ed.), *Language creation and language change: Creolization, diachrony and development*, 329–373. Cambridge, MA: MIT Press.

Bybee, Joan & Paul Hopper. 2001. *Frequency and the Emergence of Linguistic Structure*. Amsterdam: John Benjamins.

Collins, Peter. 1995. The indirect object construction in English: an informational approach. *Linguistics* 33. 35–49.

Cueni, Anna. 2004. Predicting the outcome of the choice between the dative constructions of English. Stanford, ms.

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/.

De Cuypere, Ludovic & Saartje Verbeke. 2013. Dative alternation in Indian English: A corpus-based analysis. *World Englishes* 32(2). 169–184.

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

Dodson, Kelly & Michael Tomasello. 1998. Aquiring the transitive construction in English: the role of animacy and pronouns. *Journal of Child Language* 25(3). 605–622.

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

Farquharson, Joseph T. 2013. Jamaican structure dataset. In Susanne Maria Michaelis, Philippe Maurer, Martin Haspelmath & Magnus Huber (eds.), *Atlas of Pidgin and Creole Language Structures Online*. Leipzig: Max Planck Institute for Evolutionary Anthropology. (Available online at http://apics-online.info/contributions/8, Accessed on 2017-01-20.).

Ferreira, Fernanda. 1994. Choice of passive voice is affected by verb type and animacy. *Journal of Memory and Language* 33. 715–736.

Filppula, Markku. 1999. *The Grammar of Irish English: Language in Hibernian style*. London and New York: Routledge.

Gahl, Susanne & Susan Garnsey. 2004. Knowledge of Grammar, Knowledge of Usage: Syntactic Probabilities Affect Pronunciation Variation. *Language* 80. 748–775.

Garretson, Gregory, M. Catherine O'Connor, Barbora Skarabela & Marjorie Hogan. 2004. Coding practices used in the project Optimality Typology of Determiner Phrases. corpus.bu.edu/documentation/BUNPCorpus_coding_practices.pdf.

Geeraerts, Dirk, Gitte Kristiansen, and Yves Peirsman, eds. 2010. *Advances in Cognitive Sociolinguistics*. Berlin/New York: De Gruyter Mouton.

Gelman, Andrew. 2008. Scaling regression inputs by dividing by two standard deviations. *Statistics in Medicine* 27(15). 2865–2873.

Gelman, Andrew & Jennifer Hill. 2007. *Data analysis using regression and multilevel/hierarchical models*. Cambridge / New York: Cambridge University Press.

Gerwin, Johanna. 2014. *Ditransitives in British English Dialects*. Berlin: Mouton de Gruyter.

Goldberg, Adele E. 2002. Surface generalizations: An alternative to alternations. *Cognitive Linguistics* 13–14. 327–356.

Grafmiller, Jason. 2014. Variation in English genitives across modality and genres. *English Language and Linguistics* 18(03). 471–496.

Grafmiller, Jason, Benedikt Heller, Melanie Röthlisberger & Benedikt Szmrecsanyi. 2016. Syntactic variation and probabilistic indigenization in World Englishes. Paper presented at the New Ways of Analyzing Syntactic Variation 2, Ghent, Belgium. (18 October, 2017).

Greenbaum, Sidney. 1996. *Comparing English worldwide: the International Corpus of English*. Oxford, New York: Clarendon.

Gries, Stefan Th. 2013. Sources of variability relevant to the cognitive sociolinguist, and corpus- as well as psycholinguistic methods and notions to handle them. *Journal of Pragmatics* 52. 5–16.

Gries, Stefan Th. 2015. The most under-used method in corpus linguistics: multi-level (and mixed-effects) models. *Corpora* 10(1). 95–125.

Harder, Peter. 2010. *Meaning in Mind and Society: A Functional Contribution to the Social Turn in Cognitive Linguistics*. Berlin / New York: Mouton de Gruyter.

Haspelmath, Martin. 2013. Ditransitive Constructions: The Verb "Give." In Matthew S. Dryer & Martin Haspelmath (eds.), *The World Atlas of Language Structures Online*. Leipzig: Max Planck Institute for Evolutionary Anthropology. http://wals.info/chapter/105.

Hawkins, John A. 1994. *A performance theory of order and constituency*. Cambridge; New York: Cambridge University Press.

Heller, Benedikt, Benedikt Szmrecsanyi & Jason Grafmiller. in press. Stability and fluidity in syntactic variation world-wide: The genitive alternation across varieties of English. *Journal of English Linguistics*.

Hosmer, David W. & Stanley Lemeshow. 2000. *Applied Logistic Regression*. 2nd. ed. (Wiley Series in Probability and Statistics). New York: Wiley.

Hundt, Marianne & Benedikt Szmrecsanyi. 2012. Animacy in early New Zealand English. *English World-Wide* 33. 241–263.

Kendall, Tyler, Joan Bresnan & Gerard Van Herk. 2011. The dative alternation in African American English: Researching syntactic variation and change across sociolinguistic datasets. *Corpus Linguistics and Linguistic Theory* 11. 361–389.

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

Koch, Peter & Wulf Oesterreicher. 1985. Sprache der Nähe - Sprache der Distanz: Mündlichkeit und Schriftlichkeit im Spannungsfeld von Sprachtheorie und Sprachgeschichte. *Romanistisches Jahrbuch 36*. 15–43.

Krifka, Manfred. 2003. Semantic and Pragmatic Conditions for the Dative Alternation. *International Conference on English Language and Linguistics*, 1–14. Seoul, Korea.

Kristiansen, Gitte & Dirk Geeraerts (eds.). 2013. *Contexts of use in Cognitive Sociolinguistics*. (Thematic Issue of Journal of Pragmatics 52).

Labov, William. 1972. *Sociolinguistic patterns*. Philadelphia: University of Philadelphia Press.

Leufkens, Sterre. 2013. The transparency of creoles. *Journal of Pidgin and Creole Languages* 28(2). 323–362.

Levin, Beth. 1993. *English verb classes and alternations: A preliminary investigation*. Chicago: University of Chicago Press.

MacDonald, Maryellen C. 2013. How language production shapes language form and comprehension. *Frontiers in Psychology* 4. 1–16.

MacWhinney, Brian. 1997. Second Language Acquisition and the Competition Model. In A.M.B. De Groot & Judith F. Kroll (eds.), *Tutorials in bilingualism: Psycholinguistic perspectives*, 113–142. Mahawa, NJ: Lawrence Erlbaum Associates.

McDonough, Kim. 2006. Interaction and syntactic priming: English L2 speakers' production of dative constructions. *Studies in Second Language Acquisition* 28(2). 179–207.

Meade, Rocky R. 2001. *Acquisition of Jamaican phonology.* Delft: De Systeem Drukkers.

Menard, Scott W. 2010. *Logistic regression: from introductory to advanced concepts and applications.* Los Angeles: SAGE.

Michaelis, Susanne Maria, Philippe Maurer, Martin Haspelmath & Magnus Huber (eds.). 2013. *APiCS Online.* Leipzig: Max Planck Institute for Evolutionary Anthropology. http://apics-online.info/.

Milin, Petar, Dagmar Divjak, Strahinja Dimitrijevic & R. Harald Baayen. 2016. Towards cognitively plausible data science in language research. *Cognitive Linguistics* 27(4). 507–526.

Mukherjee, Joybrato & Sebastian Hoffmann. 2006. Describing verb-complementational profiles of New Englishes: A pilot study of Indian English. *English World-Wide* 27. 147–173.

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

Pinheiro, José C & Douglas M. Bates. 2000. *Mixed-Effects Models in S and S-PLUS.* New York: Springer.

R Core Team. 2014. *R: A Language and Environment for Statistical Computing.* Vienna: R Foundation for Statistical Computing. http://www.R-project.org/.

Sand, Andrea. 2004. Shared morpho-syntactic features in contact varieties of English: article use. *World Englishes* 23(2). 281–298.

Schneider, Edgar. 2007. *Postcolonial English: Varieties Around the World.* Cambridge: Cambridge University Press.

Scott-Phillips, Thomas C. & Simon Kirby. 2010. Language evolution in the laboratory. *Trends in Cognitive Sciences* 14(9). 411–417.

Siegel, Jeff, Benedikt Szmrecsanyi & Bernd Kortmann. 2014. Measuring analyticity and syntheticity in creoles. *Journal of Pidgin and Creole Languages* 29(1). 49–85.

Siemund, Peter. 2013. *Varieties of English: A typological approach.* Cambridge: Cambridge University Press.

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

Szmrecsanyi, Benedikt, Jason Grafmiller, Benedikt Heller & Melanie Röthlisberger. 2016. Around the world in three alternations: modeling syntactic variation in varieties of English. *English World-Wide* 37(2).

Tagliamonte, Sali A. 2014. A comparative sociolinguistic analysis of the dative alternation. In Rena Torres-Cacoullos, Nathalie Dion & André Lapierre (eds.), *Linguistic variation: Confronting fact and theory*, 297–318. London, New York: Routledge.

Theijssen, Daphne, Louis ten Bosch, Lou Boves, Bert Cranen & Hans van Halteren. 2013. Choosing alternatives: Using Bayesian Networks and memory-based learning to study the dative alternation. *Corpus Linguistics and Linguistic Theory* 9(2). 227–262.

Wasow, Thomas & Jennifer Arnold. 2003. Post-verbal constituent ordering in English. In Günter Rohdenburg & Britta Mondorf (eds.), *Determinants Of Grammatical Variation in English*, 119–154. Amsterdam: Mouton de Gruyter.

Wolk, Christoph, Joan Bresnan, Anette Rosenbach & Benedikt Szmrecsanyi. 2013. Dative and genitive variability in Late Modern English: Exploring cross-constructional variation and change. *Diachronica* 30(3). 382–419.

Zaenen, Annie, Jean Carletta, Gregory Garretson, Joan Bresnan, Andrew Koontz-Garboden, Tatiana Nikitina, Mary Catherine O'Connor & Thomas Wasow. 2004. Animacy Encoding in English : why and how. In Donna Byron & Bonnie Webber (eds.), *Proceedings of the 2004 ACL Workshop on Discourse Annotation, Barcelona, July 2004*, 118–125. East Strouddsburg, PA: Association for Computational Linguistics.

Zuur, Alain F., Elena N. Ieno, Neil J. Walker, Anatoly A. Saveliev & Graham M. Smith. 2009. *Mixed Effects Models and Extensions in Ecology with R*. New York: Springer.

Appendix

Table A. Overall proportion of dative tokens by variety and variant.

Variety	ditransitive dative	prepositional dative	Total
CanE	630 (73.3%)	230 (26.7%)	860
BrE	609 (73.9%)	215 (26.1%)	824
HKE	794 (66.1%)	407 (33.9%)	1201
IndE	563 (56.1%)	440 (43.9%)	1003
IrE	611 (74.2%)	212 (25.8%)	823
JamE	665 (73.2%)	243 (26.8%)	908
NZE	699 (71.3%)	282 (28.7%)	981
PhiE	619 (65.9%)	321 (34.1%)	940
SinE	735 (72.8%)	274 (27.2%)	1009
TOTAL	5925	2624	8549

Table B. Overall distribution of dative tokens by verb and variant.

Verb	ditransitive	prepositional	Total
accord	4	3	7
advise	3	0	3
(re-)allocate	0	12	12
allot	1	3	4
allow	66	7	73
answer	1	0	1
appoint	1	1	2
assign	8	30	38
assure	3	0	3
award	4	12	16
bequeath	0	3	3
bid	9	6	15
bring	82	122	204
carry	0	1	1
cause	44	53	99
cede	0	2	2
charge	21	6	27
concede	0	6	6

convey	1	34	35
deal	3	5	8
deliver	3	23	26
demonstrate	0	4	4
deny	30	5	35
drop	30	2	32
entrust	0	5	5
explain	2	48	53
extend	3	32	35
feed	5	9	14
flick	0	1	1
forward	3	8	11
get	2	7	9
give	3,686	926	4,612
grant	41	29	70
guarantee	10	1	11
hand	39	28	67
impart	0	9	9
inform	5	1	6
issue	1	27	28
keep	6	0	6
leave	31	33	64
lend	26	38	64
loan	1	2	3
lose	0	1	1
(e-)mail	6	7	13
name	4	0	4
offer	161	83	244
owe	36	29	65
pass	15	46	61
pay	91	210	301
permit	5	0	5
play	8	12	20
pose	0	16	16
post	0	4	4
prescribe	1	1	2
present	3	62	65
promise	18	2	20
propose	0	5	5
quote	3	6	9
read	10	16	26

recommend	1	12	13
refuse	2	2	4
render	2	7	9
return	1	0	1
sell	21	83	104
send	326	218	544
serve	13	11	24
set	7	2	9
show	316	71	387
sing	3	2	5
slip	4	1	5
submit	0	63	63
suggest	1	6	7
take	0	2	2
teach	118	16	134
tell	422	23	445
throw	6	6	12
toss	1	1	2
vote	0	1	1
wish	99	0	99
write	46	48	94
yield	0	5	5

Table C. Division of the subregisters in the ICE corpora into spoken formal (SpokForm), spoken informal (SpokInf), written formal (WritForm) and written informal (WritInf).

Register	subregisters in ICE
SpokForm	public dialogues, scripted monologues
SpokInf	private dialogues, unscripted monologues
WritForm	academic writing, popular writing, instructional writing
WritInf	student writing, letters, reportage, persuasive writing, creative writing