The be- versus get-passive alternation in world Englishes

Hundt, Marianne ; Dallas, Bethany ; Nakanishi, Shimon

Abstract: Multifactorial studies of the be:get-passive alternation are still rare. On the basis of the International Corpus of English, this is the first investigation to use mixed modelling for the passive alternation in world Englishes. Overall, our findings reveal that regional differences are far less important than language-internal constraints, with Inner and Outer Circle varieties largely sharing a core grammar. Additionally, while there is qualitative evidence confirming the interchangeability of the two passive allostructions, our generalised-mixed model reveals that the choice is still heavily influenced by the different semantic origins of the variants, evident not only in the adversative semantics of theget-passive but also in a tendency to prefer human/animate subjects. The strong prescriptive reaction to the progressive passive in the Late Modern period, however, has not resulted in a marked preference forgetoverbein progressives in our data.

DOI: https://doi.org/10.1111/weng.12633

Received: 13 February 2023 | Revised: 26 May 2023 | Accepted: 2 July 2023

DOI: 10.1111/weng.12633

ORIGINAL ARTICLE

WORLD ENGLISHES WILEY

The *be*- versus *get*-passive alternation in world Englishes

Marianne Hundt[1] | **Bethany Dallas**[1] | **Shimon Nakanishi**[2]

[1]University of Zürich

[2]Kyoto University

Correspondence
Prof. Dr. Marianne Hundt, English Department, University of Zurich, Plattenstrasse 47, CH-8032 Zürich, Switzerland.
Email: mhundt@es.uzh.ch

Abstract

Multifactorial studies of the BE:GET-passive alternation are still rare. On the basis of the *International Corpus of English*, this is the first investigation to use mixed modelling for the passive alternation in world Englishes. Overall, our findings reveal that regional differences are far less important than language-internal constraints, with Inner and Outer Circle varieties largely sharing a core grammar. Additionally, while there is qualitative evidence confirming the interchangeability of the two passive allostructions, our generalised-mixed model reveals that the choice is still heavily influenced by the different semantic origins of the variants, evident not only in the adversative semantics of the GET-passive but also in a tendency to prefer human/animate subjects. The strong prescriptive reaction to the progressive passive in the Late Modern period, however, has not resulted in a marked preference for GET over BE in progressives in our data.

1 | INTRODUCTION

Passive constructions have attracted the attention of linguists in world Englishes research for various reasons. On the one hand, there is evidence of structural borrowing, as is the case with the *kena*-passive in Singaporean English (see Fong, 2004: 97–98 or Kim & Sato, 2013). On the other hand, various studies have observed differences in the frequency of BE- and GET-passives, not only across different varieties of English world-wide but also across the written vs. spoken mode. Several explanations have been offered for the observed variability, including the different historical

origins of the two constructions and/or prescriptive influence which, in turn, may affect different varieties to different degrees. Both passives combine a catenative verb (BE or GET) with a past participle. As GET does not possess the NICE properties – it needs *do*-support for negation, inversion, contrast and emphasis – we refer only to BE as an auxiliary (see Quirk et al., 1985: 160).

One of the challenges for research on the passive allostructions is their gradient nature, with prototypical or 'central' passives (Svartvik, 1966; Quirk et al., 1985: 167–68) at one end and those with an adjectival participle at the other end of the cline (cf. *They were/got hit* vs. *They were/got interested*). Another obstacle to corpus-based research into passives was that, before the advent of Part-of-Speech (PoS)-tagged corpora, studies had to rely on lexical retrieval, which proves particularly problematic for BE-passives.

With a very few exceptions, existing studies focus on either the BE-passive or the GET-passive, and few of these studies take a multifactorial approach at modelling the BE-passive (in opposition to active transitives, as in Hundt et al., 2016 or Hundt et al., 2021). Biewer (2009), who investigates both GET- and BE-passives looks at them separately rather than as a choice context. Only two previous studies have modelled the BE:GET-passive alternation so far, namely Fehringer's (2022) multifactorial analysis of the alternation in a regional variety of British English (BrE) and Bohmann et al.'s (2023) study on diachronic American English (AmE), which takes a bottom-up, vector-based approach at modelling semantic aspects of the choice between the two variants.

The aim of the present paper therefore is to fill a gap in existing research. On the basis of data from a broad range of regional varieties of English as well as a range of variety types, we model the factors that influence the choice between a BE- and a GET-passive. The predictor variables in our multifactorial approach are mostly derived from previous research into the two passive constructions, which we review in Section 2 of the paper. On account of the gradient nature of passives,[1] a clear definition of the dependent variable is an important starting point for data retrieval and analysis. We provide background on these aspects in Section 3, along with details on the approach to statistical modelling that we take as well as the operationalisation of the predictor variables. The results of our analyses are presented in Section 4 and their wider relevance for the theoretical modelling of constructional variation in world Englishes is discussed in the concluding section. Specifically, the results on the passive add to a growing body of research showing that world Englishes share a core grammar, including important aspects of variability in the use of alternating forms.

2 | PREVIOUS RESEARCH

2.1 | Passives in world Englishes

As pointed out above, the majority of previous studies looks at either the BE-passive or the GET-passive in world Englishes. In their study on the active:passive alternation, Hundt et al. (2016) find that there are very few regional differences in academic writing, with no obvious influence of different substrates in the second-language varieties. They attribute the avoidance of the BE-passive in AmE academic writing to prescriptive influence. Hundt et al.'s (2021) multifactorial study reveals more fine-grained differences in the effect size that predictor variables like animacy, length of the subject or givenness play in the choice between active transitives and BE-passives.

Earlier studies have looked at GET-constructions in world Englishes generally (Bruckmaier, 2017; Coto-Villalibre, 2014) or focused more specifically on passives with GET (most others). GET-passives have been studied against the background of different theoretical frameworks, notably generative (Alexiadou, 2006, 2012; Wanner, 2009) and functional (Givón, 1993; Quirk et al., 1985) ones, and using different methodological approaches, such as corpus data (Rühlemann, 2007; Schwarz, 2017) or experimental evidence (Thompson et al., 2013, 2018). The range of Englishes that have been investigated in previous studies is also quite broad, including varieties of English as a first language (InnerCircle) from different regions (specifically northern and southern hemisphere varieties, see Leech et al., 2009 or Hundt et al., 2008), Outer Circle (OC) varieties with AmE or BrE as their matrilect (namely Philippine English vs. Singaporean, Indian, Hong Kong or Fiji English, see Hundt, 2009; Coto-Villalibre, 2015, 2016), as well as English as a

second dialect (ESD) in Jamaica (Bruckmaier, 2016). Fehringer (2022) adds a regional variety of BrE (Tyneside English) to these.

Previous research is somewhat inconclusive with respect to regional differences in the distribution of GET-passives, which are typically presented as normalised frequency (either by corpus size or relative to other GET-constructions). Collins (1996) finds the GET-passive to be more frequent in Australian English (AusE) than in BrE and AmE while Indian English (IndE) uses GET-passives even more frequently than any of the Inner Circle (IC) varieties. Some studies (Leech et al., 2009) provide evidence for the GET-passive being more frequent in AmE than in BrE, while the results of other studies indicate that settler varieties such as Australian (AusE) and New Zealand English (NZE) fall between the two matrilects in their use of GET-passives (Sussex, 1982; Hundt, 1998: 78). There is preliminary evidence that OC varieties align with BrE (namely IndE) or AmE (notably PhilE and SingE), and not necessarily their historical matrilects (Hundt, 2009). One of the possible reasons for these different findings could be that they are based on different corpora which were sampled at different times using different sampling frames. Collins' (1996) investigation uses the American Brown and the English Lancaster-Oslo/Bergen (LOB) corpus, which are collections made from texts published in 1961, and compares them to the Australian Corpus of English and the Kolhapur Corpus of IndE. The latter used the same sampling frame as the Brown and LOB corpora, albeit on material sampled in the second half of the 1980s, about 25 years after the publication of the texts in the reference corpora. With evidence of a growing use of the GET-passive (ongoing change), the diachronic dimension is likely to have had a skewing effect with respect to regional differences across the Brown-type corpora (see Hundt, 1998). Studies based on the *International Corpus of English* (ICE) include material from spoken as well as written usage, and within written texts, both published and unpublished material. This is important because previous research indicates that GET-passives are used more frequently in speech than in writing (Hundt, 2009), while they are particularly rare in formal writing (Mair, 2006: 111; Leech et al., 2009: 158). Moreover, Bohmann et al.'s (2023) multifactorial analysis of the passive alternation in the *Corpus of Historical American English* has shown that the association of the GET-variant with informal contexts persists across time, while other factors weaken. In other words, variation according to mode (speech vs. writing) and register (formal vs. informal) is likely to have an influence on the BE:GET-passive alternation, and it remains to be seen how these factors interact with regional variation.

2.2 | Constraints on the BE: GET passive alternation

The two passive constructions have quite different diachronic trajectories, which in turn have led to different contextual preferences. While the BE-passive is the older of the two variants, with incipient grammaticalisation of this passive variant dating back to the Old English (OE) period, the GET-passive is a much more recent development.

Like other Germanic languages, OE made use of two verbs, namely *beon/wesan* 'be' and *weorðan* 'become'. English lost the latter, more dynamic variant of the emergent passive in the Middle English period (see Fischer, 1992: 250 or Petré, 2010). The variant with GET developed in the Late Modern period, with the earliest unambiguous passives attested from the end of the seventeenth century (see Hundt, 2001: 71). Brinton (1994: 159) points out that the two passives may eventually specialise to stative and dynamic variants of the passive construction:

> Though the *get* passive is not yet fully standardized in English, its existence may ultimately lead to the restriction of the *be* form to stative or statal passive meaning and the *get*form to dynamic or inchoative passive meaning.

A precursor to these more dynamic passives are causative transitive constructions with animate subjects such as *get sth done*, which also had reflexive variants, as in *he got himself fired* (Hundt, 2001). Unlike prototypical BE-passives, which typically have non-human or inanimate subjects, the origins of the GET-passive thus foster a preponderance for animate subjects. These origins also explain why the subject in GET-passives is still felt to carry some of the

responsibility for the action it undergoes, unlike the prototypical patient in a BE-passive (*He got/was fired*; see Hatcher, 1949; Vanrespaille, 1991; Carter & McCarthy, 1999). Because of this semantic difference in the involvement of the grammatical subject, GET-passives have been said to combine with an agent *by*-phrase less easily and often than BE-passives (see Matthews, 1993: 22 or Carter & McCarthy, 1999: 51). Our data will allow us to verify how strong these semantic aspects are in determining the choice between BE and GET to form the passive.

While there are historical reasons that the semantics of the subject might have a limiting effect on the availability of GET to form passives, aspectual restrictions in the earlier history of the BE-passive might favour the choice of GET over BE in progressives. The earliest progressive BE-passive dates back to the late 1760s (Bergen, 2013). Moreover, the combination of the progressive with passive BE as auxiliary was one of the most hated grammatical innovation at the time and attracted massive reaction from prescriptive grammarians (see Anderwald, 2016). Smith and Rayson (2007) have therefore argued that the GET-passive might be favoured in combination with the progressive even in Present-Day English, as it avoids the combination of two forms of BE.

Since Hatcher's (1949) paper, the special semantics of the GET-passive has been repeatedly addressed in the literature (Chappell, 1980; Downing, 1996). According to this, GET-passives are prone to combine with verbs such as *fire*, *arrest* or *sack* that express actions which have a negative effect on the subject. The studies that have looked at the alleged 'adversative' semantics of the GET-passive have mostly focused on the proportion of adversative vs. non-adversative meanings of the construction itself. Fehringer (2022) finds that in her older speakers, there is an effect of semantics in that BE shows a clear preference for benefactive contexts, but in younger speakers, semantics of the construction no longer has an effect. Similarly, Bohmann et al. (2023) find that the semantic constraint weakens across time in their AmE data. It will be interesting to see whether semantics has an effect on the passive alternation in our world Englishes data.

2.3 | Research questions

With the exception of Fehringer (2022) and Bohmann et al. (2023), this is the only study that takes a multifactorial approach to the BE:GET passive alternation and the first to do so for a broad range of world Englishes. With this approach, we aim to answer the following research questions:

RQ1: Which of the factors in our analysis impact on the choice between the BE- and the GET-passive and how do they interact?

RQ2: Does regional variety emerge as a significant/important factor, and if so, do variety type (IC vs. OC), area (northern vs. southern hemisphere, Asian vs. non-Asian varieties) or diachronic connection (British-based vs. American-based varieties) play a role?

The second RQ is closely related to the question whether the emergence of new Englishes goes hand in hand with a process of nativisation or indigenisation, which does not typically result in categorical differences but rather regional preferences for usage patterns (Schneider, 2007: 22) or what Szmrecsanyi et al. (2016: 133) refer to as 'probabilistic indigenization', namely:

> the process whereby stochastic patterns of internal linguistic variation are reshaped by shifting usage frequencies in speakers of post-colonial varieties. To the extent that patterns of variation in a new variety A, for example, the probability of item x in context y, can be shown to differ from those of the mother variety, we can say that the new pattern represents a novel, if gradient, development in the grammar of A. These patterns need not be consistent or stable … but they nonetheless reflect the emergence of a unique, region-specific grammar.

TABLE 1 Overview of world Englishes included.

Variety	Acronym	Region	Variety type
British English	BrE	Europe	Inner Circle
Irish English	IrE	Europe	Inner Circle
Canadian English	CanE	North America	Inner Circle
Australian English	AusE	Pacific	Inner Circle
New Zealand English	NZE	Pacific	Inner Circle
Hong Kong English	HKE	Asia	Outer Circle
Singapore English	SingE	Asia	Outer Circle
Indian English	IndE	Asia	Outer Circle
Philippine English	PhilE	Pacific	Outer Circle
Jamaican English	JamE	Caribbean	Outer Circle (ESD)

On the basis of previous research, we predict to find differences in the use of GET-passives across different types of variety, with IC varieties being, overall, more attuned to the colloquial origin of the GET-passive and hence more likely to avoid it in the more formal, written register (see Hundt, 2009).

We also expect GET-passives to more readily combine with progressive aspect, particularly in IC varieties, due to the late emergence of the progressive BE-passive and the tendency to avoid the combination of two forms of auxiliary BE (passive and progressive) in succession. Varieties with a substrate that requires marking for aspect more regularly (like IndE, see Sharma, 2009) are expected to be overall more open to combinations of both progressive and perfect aspect with both passive constructions.

Seeing that the GET-passive, due to its path of constructionalisation, is inherently more agentive, we expect the GET-passive to show an overall greater propensity to be used with human subjects, while the BE-passive is likely to be used more freely with inanimate subjects and a *by*-phrase openly assigning responsibility to an external agent.

As in previous research on grammatical alternations (Szmrecsanyi et al., 2016), we expect language-internal factors like subject animacy or aspect to play a greater role in the choice of passive construction than variety type (IC vs. OC) or regional variety.

3 | METHODOLOGY

Our data come from the ICE, which provides a set of standard, one-million-word corpora compiled along the same sampling principles (Greenbaum, 1996). Seeing that the GET-passive is still increasing and diachronic change might also affect the predictors for the choice in the passive alternation (see Fehringer, 2022), we limit our analysis to the first-generation ICE components and exclude any of the ICE-Age 2 components such as ICE-Nigeria or ICE-Sri Lanka (see Hundt, 2020). Moreover, not all currently available ICE components have been syntactically annotated, which would make retrieval of BE-passives problematic. As a result, the regional spread of our data is somewhat limited: our OC varieties come from Asia as African varieties were underrepresented in first-generation ICE components, and those that were included have not been syntactically annotated, yet. We were able to include a total of 10 world Englishes in our study, among them five varieties from countries where English is the first language for the majority of speakers (IC), four where it is an institutionalised second language (OC) and one where standard English is spoken as a second dialect of English alongside a creole variety, namely Jamaica. Table 1 gives an overview of the regional varieties we include in our study. We did not include US English in our data because the spoken part of that ICE component has

not been completed, and our aim is to compare the alternation across both spoken and written contexts of use. North American varieties are represented by Canadian English, a close relative of US English.

Since first-generation ICE components represent language use from the late twentieth-century, our study can serve as a starting point for research on more recent (albeit less carefully compiled) corpora such as the *Global Web-based English* (GloWbE) (www.english-corpora.org/glowbe/) or the *News on the Web* (NOW) (https://www.english-corpora.org/now/) corpora.

3.1 | Retrieving passives from the tagged components of ICE corpora

We accessed the ICE corpora online[2] via the Zürich Dependency bank (see Lehmann & Schneider, 2012 for details), which allowed us to make use of the PoS annotation of the corpora. We retrieved all forms of the catenatives followed by a past participle, allowing for a total of five/four words between the catenative and the participle.[3] The algorithms retrieve affirmative and negative variants of the passives, instances in declaratives and interrogatives as well as variants with intervening adverbs or hesitation markers like *uh* or *ehm* common in the spoken part of the corpora. Interestingly, the algorithm also retrieved the occasional (potential) passive with a bare participle; the following example from ICE Hong Kong illustrates this (even though BE and the bare participle are only adjacent and not part of the same verb phrase in (1), which was not included in our set of passives):

(1) They *are Welcome* Back by Pricilla Chan, Just wake up from dream by Cass Pang, and All time favourite by Shirley Kwan. (ICE-HK, W1B-009)[4]

The number of potential passives retrieved allows us to analyse all constructions with GET (between 161 and 460 per variety, see Table 3, subsection 4.1). The number of potential BE-passives turns out to be too large for exhaustive analysis. Since our main interest is in gauging the relative importance that a common set of predictor variables has on the choice of one construction in the BE:GET-passive alternation, analysing a random subset from the BE-concordances is sufficient.[5] We decided to select a total of 150 BE-passives for each variety for further analysis.

3.2 | Definition of the variable context and manual post-editing

We treat the BE:GET-passive as an alternation, as two different ways of saying the same thing. The following examples illustrate that they co-occur, with example (2) illustrating nicely that both variants are, in fact, occasionally used interchangeably with the same lexical verb.

(2) Anything that has to *get gone* into can *be gone* into in the morning, Rose said. (ICE-IRE, W2F-018)
(3) See that the calculated interest is correct and discrepancies, if any should *be noted* and *get rectified*. (ICE-IND, W2A-013)

In construction grammar, constructions are 'conventional, learned form-function pairings at varying levels of complexity and abstraction' (Goldberg, 2013: 17) and the variants with the two catenatives can thus be conceived of as allostructions (see Cappelle, 2006) of an overarching, more abstract passive construction. Passive constructions combine a patient subject NP with a catenative and the past participle of a lexical verb; the agent of the related active construction is optionally expressed in a prepositional phrase, typically introduced with *by*. The relationship between the more abstract passive and its allostructions is shown in Figure 1. At the constructeme level, the catenative (C) is underspecified whereas in the two allostructions, the catenative is lexically specified as BE vs. GET, respectively.

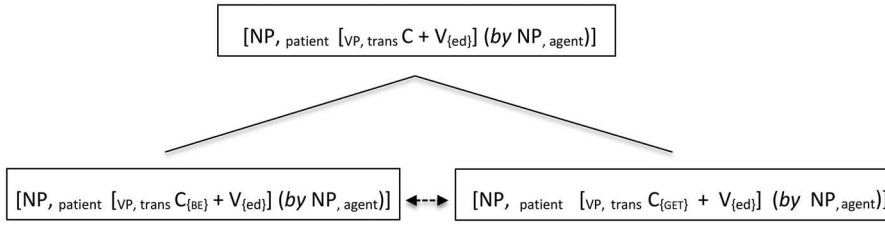

FIGURE 1 Constructeme-allostruction relation of the English passive alternation.

Previous research has shown that the passive allostructions are characterised by different contextual preferences: they have an affinity to different registers, verb semantics, animacy of the subject and different propensities to combine with tense, aspect and modality (TAM). Nevertheless, we expect these to be tendencies rather than categorical rules. Importantly, they assign the same semantic role to the NP in subject position, which undergoes the action expressed by the lexical verb.

Since combinations of BE and GET with a past participle are gradient phenomena that are also used to express non-passive semantics, it is important to clearly define the variable context. We only included dynamic, central BE-passives in our dataset. These typically denote events rather than states, and while passive *They were arrested* is ungrammatical with gradation adverbs such as *very*, stative copular *They were frightened* combines with gradation adverbs, indicating that *frightened* functions like an adjective rather than a verb. As central BE-passives are also non-reciprocal, we included constructions like *They were loved* in our dataset but not *They were engaged/married* as only the latter combines with *to each other*. With respect to the GET-passive, similar criteria apply: instances that allow for a gradation adverb such as *He got (very) worried* were excluded as the participle is an adjective. We also excluded constructions with reflexive or reciprocal semantics, such as *He got shaved* and *They got married*, and unclear instances in which a reflexive interpretation was possible. In the following example, either an external agent is sorting them out or the subjects are actively doing the sorting themselves:

(4) … and they'*re getting sorted out*. (ICE-GB, S2A-006)

English passives can also take the indirect object of a related ditransitive construction as their subject. This holds also for GET-passives, as example (5) shows.

(5) … my sister *got given* a hamster once for her birthday by a family that was moving away. (ICE-CAN, S1A-053)

Tests for passive-hood would lead us to expect that the lexical use of GET (in the sense of 'obtain') and the passive use cannot occur in a coordinated verb phrase. However, ICE-NZ yields the following example, which we included in our set of GET-passives:

(6) We went to a lovo and he *got food poisoning, and badly bitten* by mosquitoes. (ICE-NZ, W2F-011)

In manually post-editing the data, we excluded a number of false positives, both erroneously retrieved instances as well as those that did not fit our definition of the dependent variable. Some false positives from the spoken part of the corpora are due to repetition or planning errors, in which case the retrieval algorithm misidentified the repetition of the catenative or the second verb as the lexical verb potentially forming a passive, as in (7) and (8). Online spoken production occasionally leads to a change in catenative before the lexical verb is selected, as in (9) and (10); such instances were not included in our dataset:

(7) Well it's *been* it's *been* functioning as a hotel also uh apparently. (ICE-PHI, S1B-038)
(8) I suppose if you're in a country that's leading the world or starting to *get become* really prosperous … (ICE-AUS, S1B-014)
(9) … *we get we are separated* from the English speaking people … (ICE-IND, S1A-010)
(10) … a lot of these uh brands *will get will be stolen* or cheated … (ICE-JAM, S1B-051)

The algorithm retrieved instances where BE/GET did not form a complex verb phrase with the participle. This kind of false positive is illustrated in (1) as well as (11)–(15):

(11) Mr Moore said land-holders could still *be* liable for court costs *involved* in defending their leases against claims. (ICE-AUS, W2C-008)
(12) In thematic terms it *would be* how *trapped* the characters become in their lives and the roles that are determined for them by others, usually parents. (ICE-NZ, W1A-004)
(13) It *gets* red when *activated*. (ICE-SIN, S1A-008)
(14) The whole process is thus streamlined and clients *get processed* results much more quickly. (ICE-HK, W2A-033)
(15) you've *got to* actually *have built* i think. (ICE-NZ, S1A-068)

Our definition of the BE:GET-passive led us to exclude further false positives. In example (16), the participle functions as an adjective as the sentence could easily be rephrased as *We were then unprepared* … . On the same grounds we excluded example (17), which does not qualify as a passive despite the *by*-phrase that could potentially be taken to encode the external agent.[6] The participles in (18) and (19) can combine with an adverb like *very* or *completely*, so function as adjectives. This also holds for the participle in example (20)a. where the *by*-phrase, moreover, is the cause of the action rather than a constituent that could function as a subject in an active transitive construction:

(16) We *were* not then *prepared* to live within our means as a nation … . (ICE-IRE, S1B-060)
(17) a. Shipley *is undaunted* by the prospect. (ICE-NZ, W2B-013)
b. *The prospect ***un***daunted* Shipley.
(18) Yeah stuff that everyone knows really well so they can go along and *get pissed* … (ICE-AUS, S1A-011)
(19) it's not in the area that i'd like to *get involved* in. (ICE-NZ, S1A-055)
(20) a. You really *get hooked* by these astronomical events. (ICE-CAN, W2B-028)
b. *These astronomical events really hooked you.
c. These astronomical events get you really hooked.

The following examples were excluded because BE and GET here function as catenatives in stative/resultative constructions rather than dynamic passives:

(21) Both of them *were covered* in spatters of a clear substance that looked like Clearasil gel. (ICE-GB, W2F-002)
(22) … he *was* totally *changed* actually … (ICE-IND, S1A-010)
(23) … we've got to *get finished* by Christmas. (ICE-GB, S1B-029)

A complication for the delineation of the variable is added by the fact that, particularly in contact varieties of English, BE is also used to form perfects (Werner, 2016), which makes (24) potentially ambiguous between a passive and a perfect reading. Such potentially ambiguous instances were excluded from our dataset:

(24) When the solution of gelatinized starch *is cooled*. (ICE-HK, W1A-019)

Finally, BE/GET + participle are also used in a broad number of idiomatic contexts, such as BE/GET *used to*, *get rid of*, *get stuffed* in the sense of 'go away', *get carried away*, *get hooked* or *get fed up*.

(25) This was not a situation the East Germans *were used to*; … (ICE-GB, W2F-012)
(26) … Rosie sort of said oh well *get stuffed*. (ICE-AUS, S1A-045)
(27) we're going to *get rid of* that wood anyway. (ICE-CAN, S1A-005)
(28) I don't *get carried away with* huge transit mixer business … (ICE-AUS, W1B-030)
(29) I just *got hooked on* it. (ICE-JAM, S1A-062)
(30) … I *got fed up* with his brother … (ICE-PHI, S1A-057)

The BE-passives were annotated by two authors, each. Inter-annotator agreement (based on the Kappa coefficient, Carletta, 1996) was high at around 80.8%. Additionally, we were conservative in our decision to include BE-passives and excluded instances where one coder had raised doubts. The GET-passives were all coded by the first author (two rounds of coding for consistency) and double-checked by the second author. We provide the complete dataset at https://osf.io/dx5bz/.

3.3 | Statistical modelling

We fit a Generalised Mixed Model Tree (glmertree) to our data. This approach combines the linear mixed-modelling approach with a decision tree and can be easily implemented in the glmertree package available in R (Fokkema & Zeileis, 2019). The advantage of this approach is that decision trees provide a straightforward way of interpreting the interactions of predictor variables in a model (see Fokkema et al., 2018) while the mixed-modelling approach allows us to include the verb lemma as a random effect.[7] Since fitting the glmertree with all predictor variables failed to work, we first fit a Random Forest (RF) using the party package in R (Strobl et al., 2009) to determine which of the predictor variables were significant. For the glmertree, we then started out with the first four significant predictors from the RF (namely VARIETY, REGISTER, ANIMACY, SEMANTICS), which did return a valid model and proceeded to add the other predictors in the order they were returned in the RF. We test for model fit with Somers2. The glmertree returned a good model fit with VARIETY, REGISTER, ANIMACY, SEMANTICS and TENSE (see subsection 4.2), but failed to work with any of the other predictors added into the model. We provide the R codes we used for fitting the RF and glmertree at https://osf.io/dx5bz/.

Table 2 gives an overview of the variables we include in the RF as well as the random effect included in the glmertree. Level codes in brackets were used to improve the legibility of the glmertree. We comment on the operationalisation of those factors that are not straight forward.

For the verb lemma, we chose to normalise spelling variation to BrE, that is words ending in *-ize* or *-or* were coded as *-ise* and *-our* across all varieties. In cases where there were two verbs in a conjoined verb phrase, we annotated the first one, as in the following instance where V_LEMMA is *review*:

(31) All these decks *have been reviewed* recently or <u>were borrowed</u> to aid the exploration of differences in sound quality between transports. (ICE-GB, W2B-040)

With respect to register, we use the macro categories from ICE, which are essentially a combination of the mode of production (speech vs. writing) and formality (a cline from the most informal, that is, dialogues, to the most formal, printed texts). A more fine-grained analysis is not really warranted given the small sub-samples for some categories (for example, telephone conversations comprise of 10 samples against the 90 face-to-face conversations).

As pointed out in Section 2, ASPECT (simple vs. perfect vs. progressive) is likely to have an effect on the choice of the GET- over the BE-passive, both for diachronic reasons (recency of the progressive passive) and potential influence

TABLE 2 Summary of variables for the BE:GET passive alternation in ICE.

Variable	Label	Levels
Catenative	CATENATIVE	BE, GET
Regional variety	VARIETY	BrE (GB), IrE (IR), CanE (CA), NZE (NZ), (AusE) AU, HKE (HK), PhilE (PH), IndE (IN), SingE (SG), JamE (JA)
Verb lemma	V_LEMMA	value of the base form of the verb
Register/mode	REGISTER	spoken dialogue (sd), spoken monologue (sm), written unpublished (wu), written published (wp)
Tense	TENSE	past (p), non-past (np), non-finite (nf)
Aspect	ASPECT	simple (s), perfect (perf), progressive (prog)
Modality	MODAL	core modal (m), semi-modal (sm), no (semi-)modal (n)
Animacy of the subject	ANIMACY	human (h), animate (a), inanimate (i)
Semantics	SEMANTICS	adversative (adv), non-adversative/neutral (neut)
Agent	BY_PHRASE	(agent) *by*-phrase (y), no agent (n)

of language contact (importance of aspect marking in the main substrate for IndE, Hindi). Note that only modal (32) and finite (33) progressives were coded as progressive because it is in these contexts that two instances of auxiliary BE would occur and the GET-passive offers a strategy to avoid this co-occurrence.[8] In other words, instances like (34) or (35) were coded as 'nf' (non-finite) and 's' (simple). We further coded for function rather than form in cases where contact varieties had a bare present participle (as in (36)) or used an allograph of *been* (as in (37)).[9]

(32) Wouldn't she *be getting paid* for her holidays. (ICE-AUS, S1A-049)
(33) Saudi general said a camp *is being set* up to house eight thousand refugees. (ICE-IND, S2B-008)
(34) … I saw men *being led* to their execution. (ICE-AUS, W1B-006)
(35) But buyers don't have to worry about *getting caught*. (ICE-HK, W2B-036)
(36) Concept of time has been changed and thus stream of consciousness method *has be employed*. (ICE-IND, W1A-018)
(37) A new full-screen editor to create and modify text files *has* now *being included* with DOS 5.0. (ICE-SIN, W2B-031)

With respect to predictor variable MODAL, we coded both core modals (m) and semi-modals (sm), with the latter subsuming Future Time Expressions (FTE) *going to/gonna*:

(38) … lawyers for both sides would argue over what evidence *could be presented* in the trial. (ICE-CAN, S2B-003)
(39) I don't want to get hurt. (ICE-PHI, S1A-018)
(40) I wasn't *gonna get paid*. (ICE-AUS, S1A-043)

ANIMACY of the subject is a gradient phenomenon. Among the different approaches to the question how best to operationalise this predictor variable, we opt (with slight modification) for the one outlined in Zaenen et al. (2004), a distinction into three categories (human, animate, inanimate), where 'animate' subsumes organisations (collectives of humans, see (38)) and animals, with the latter also applying to viruses and bacteria (Zaenen et al., 2004: 121). As the effect of ANIMACY in linguistics is different from the way it is defined in biology, Zaenen et al. (2004: 121) also included intelligent machines and vehicles in the category 'animate', whereas we decided to code them as 'inanimate'. Instances in which the name of a place (*Downing Street Number 1*) is used metonymically to refer to the people in the house, on the other hand, are labelled as 'animate' (that is, collectives of humans).

(41) The team who played last week that was *the team* that apparently got hit twelve times. (ICE-CAN, S1A-031)

While Zaenen et al. (2004) do not explicitly comment on how plants and mushrooms were classified, it makes sense to subsume them under the label 'animate', particularly seeing that viruses and bacteria are included within this category by Zaenen et al. (2004) and in the light of examples such as (42).

(42) I wondered how the owner of the home felt about the apple tree *getting poisoned*. (ICE-CAN, S2B-039)

Occasionally, the intended subject of a passive has to be inferred, as in the following instance, where it has to be human, as humans typically ride animals, an action that usually involves a certain amount of control.

(43) "There are 100 ways to *get hurt* in riding steers, and riding bulls," states Johansen … (ICE-CAN, W2D-020)

With the help of contextual cues, it was possible to assign one of the three category labels ('h', 'a' and 'I') to the few instances with omitted subjects in our data.

Another important aspect of the passive alternation is the semantics, the question of whether the GET-passive is more likely to combine with verbs expressing adversative meaning. We encode the semantics as a binary distinction between adversative and neutral/non-adversative meaning. However, we code for semantics on the level of the construct rather than the verb only, taking in the context into account. An example where context is key to interpretation of the semantics of the construct comes from *downgrade* in (44), which is used in a context where the subject actively tries to bring this change about and thus cannot be read as adversative. However, we decided to code all negated instances of adversative verbs, as in (45), as 'a' despite the fact that the negative particle counteracts the semantics of the verb.

(44) I am trying desperately to *get downgraded*. (ICE-SIN, W1B-005)
(45) How a lot of people did not *get killed* is beyond me. (ICE-AUS, W2C-018)

For a prepositional phrase with *by* to be analysed as a *by*-agent phrase, it had to be a PP and be conceivable as the subject in a related active rendering of the passive construction. In example (46), the element headed by the preposition *by* thus does not qualify as a *by*-agent. Conversely, not all agents are introduced by *by*, as example (47) illustrates, which we did include among the instances with an overt agent phrase.

(46) … we *get punished* by having our the interest on our savings taxed. (ICE-AUS, S1B-030)
(47) … you do *get charged* from the firms who run these systems … (ICE-AUS, S1B-043)

4 | RESULTS

4.1 | Summary statistics

In addition to the 1500 BE-passives (150 per variety), we retrieved and annotated a total of 909 GET-passives. Table 3 shows the total number of potential passives (first row) and the number of true positives (second row) across the different ICE components.

As the figures in the second row of Table 3 show, there is no immediate distinction between IC and OC varieties with respect to the overall number of GET-passives: AusE, NZE, CanE as well as IndE yield more than 100 GET-passives, whereas BrE has even fewer than PhilE and SingE.

TABLE 3 Distribution of get-passives across world Englishes.

	AusE	BrE	CanE	HKE	IndE	IrE	JamE	NZE	PhilE	SingE
GET + Ved	312	229	342	161	340	333	241	460	229	218
GET-passives	116	53	122	39	130	82	82	146	70	69

As far as the distribution of the two passive allostructions across the registers/modes is concerned, we see a complementary distribution, with the BE-passive favouring the written mode at 62% (930/1500) and the GET-passive the spoken mode at 78.98% (718/909). Within writing, it is interesting to observe that prescriptive influence might have affected the distribution as only 11.1% of all BE-passives are found in published written texts, whereas 50.9% occur in unpublished writing. GGET-passives, on the other hand, clearly lean towards the most informal register: 53.9% occur in spoken dialogues, compared to 25.1% in spoken monologues. This is in line with previous research.

The overall tendency with respect to tense, aspect and modTAM is one where BE-passives more easily combine with other elements in the VP, which is evidenced by the higher proportion of non-finite catenatives (48.7% vs. 36.6% for GET-passives), perfect passives (13.7% vs. 1.3%) and combinations with a modal/semi-modal (25.3% vs. 13.2%). The proportion of progressive passives is slightly higher for GET (4.6%) than for BE (3.4%), but the difference is not significant ($p = 0.13177$ in a chi-squared test).

With respect to semantic aspects, we also see the expected distribution of the BE-passive favouring inanimate subjects (77.4%) and the GET-passive favouring human subjects (63.6%), that is, there is a clear division of labour that reflects the constructional origins of the GET-passive with constructions that have a human, agentive subject. This has the expected repercussions on the co-construction with a *by*-agent phrase, which is slightly more common with BE- than with GET-passives (16.7% vs. 7.7%). The GET-passive also shows the expected preponderance for adversative contexts (45.3%), but 12.1% of BE-passives also co-occur with adversative verbs, illustrated in (48)–(50).

(48) Two years later his world fell apart when Mary, 26, and son David, 4, *were killed* in a car accident. (ICE-AUS, W2B-001)

(49) … I *was* once *drowned* in a river, in the arms of a priest. (ICE-HK, W2F-020)

(50) A nine-year-old boy *was raped* and *killed* there. (ICE-JAM, W2C-017)

The example from ICE-HK additionally shows substrate influence from Chinese: in both Mandarin and Cantonese (according to our informants), 'to drown' does not automatically imply death by drowning.

4.2 | Modelling constraints on the passive alternation

Figure 2 shows the results of the RF (ntree = 500, mtry = 8) with the ranking of the predictor variables;[10] Somers2 returned a very good model fit with C = 0.937435. Figure 2 shows that REGISTER, VARIETY, ANIMACY and SEMANTICS are the top four predictor variables.

Including the first five predictors from the RF in the GLMM returned a valid model with excellent model fit (C = 0.9452). Adding ASPECT and/or MODAL results in error messages because of the high number of Eigenvalues. In other words, two language-external predictors (REGISTER and VARIETY), two semantic factors (ANIMACY of the subject and (non-)adversative SEMANTICS) and one VP-related predictor (TENSE) turn out to be relevant predictors for the choice between the two passive variants across our ten world Englishes.

The detailed results of the glmertree analysis are provided in the Appendix. Interestingly, it is ANIMACY rather than REGISTER that provides the first split of the data (Node 1), with animate and human subjects on the left of the tree and inanimate subjects on the right of the tree, showing an overall tendency to take GET- and BE-passives, respectively.

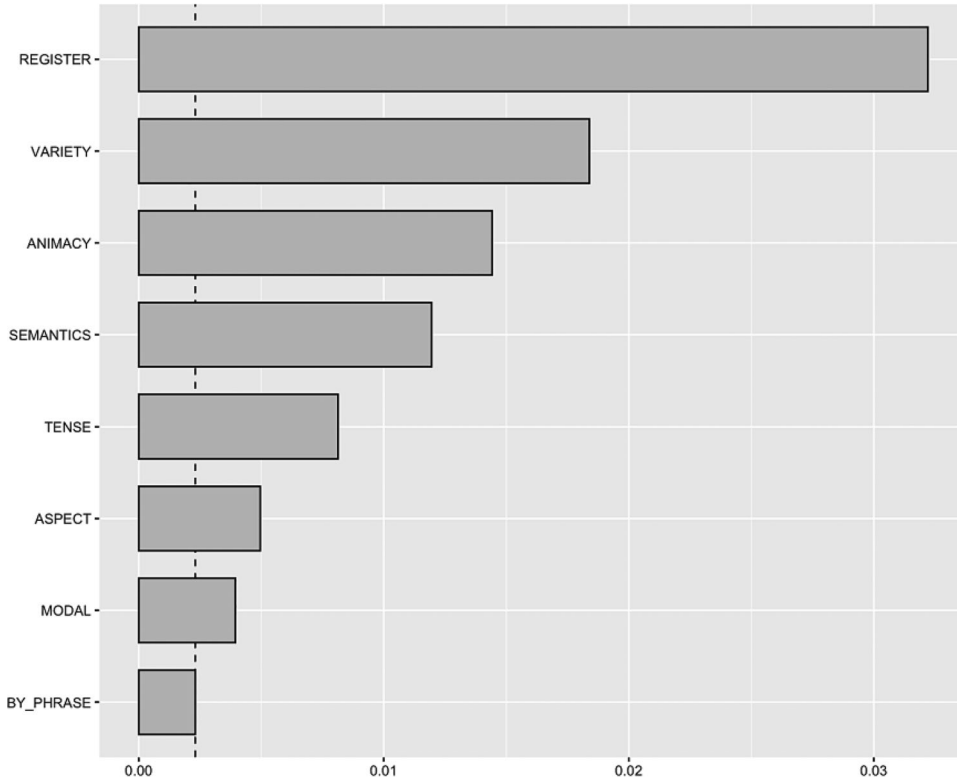

FIGURE 2 RF analysis of passive alternation across ICE components – variable importance.

VARIETY, which ranks second in the RF, occurs much lower down in the glmertree. Because of the large number of nodes and world Englishes included in the model, it is necessary to zoom in on the respective sections of the tree plot. At Node 2, the data are split by REGISTER into spontaneous spoken (Node 3) vs. the other registers (Node 8).

Within Node 3, SEMANTICS provides the next split, with 'adversative' contexts showing interaction with VARIETY (see Figure 3): in spontaneous speech, and for most varieties (with the exception of BrE and HKE), adversative semantics strongly predict the GET-passive (Node 6). In no-adversative contexts across all varieties, GET-passives are alpreferred over BE-passives (Node 7). Notably, the effect of VARIETY at Node 4 does not occur along groupings according to variety type or region as GB and HK are grouped together vis à vis the remaining varieties.

At Node 8, passives with human and animate subjects in spoken monologues and written texts split according to the predictor SEMANTICS again. This time, adversative semantics do not show any interaction with VARIETY: at Node 9, the mode of interaction provides the split in the data, with written texts showing a preference for BE-passives when compared with spoken monologues (cf. Nodes 10 and 11) in Figure 4a. Turning to Figure 4b, we see that in neutral contexts and printed written communication, BE-passives are preferred (Node 13). In monologues and unprinted texts, we observe interaction with VARIETY (Node 14), with only HKE, PhilE and SingE still dispreferring the variant with GET even in more informal communication. Note that again, the effect of VARIETY does not coincide with any regional groupings or type of World Englishes as both IndE and JamE pattern along the same lines as the IC varieties.

If we now turn to the part of the tree that indicates the distributional profile for passives with inanimate subjects, Node 17 shows an early split by REGISTER into spoken vs. written production. The tendency for adversative contexts to strongly favour GET-passives also holds for inanimate subjects in speech (Node 18 in Figure 5).

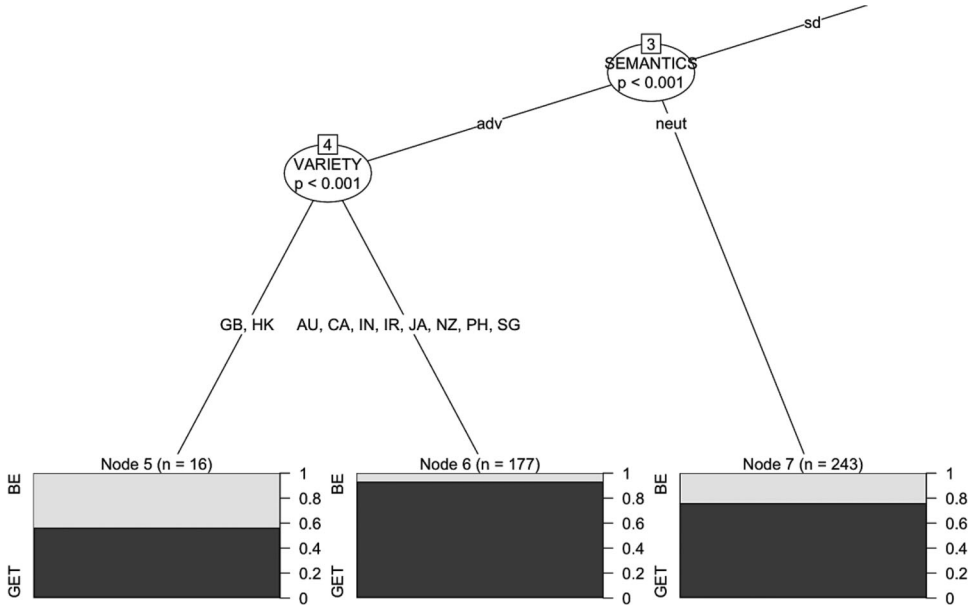

FIGURE 3 Partial glmertree for passives in dialogues with human and animate subjects.

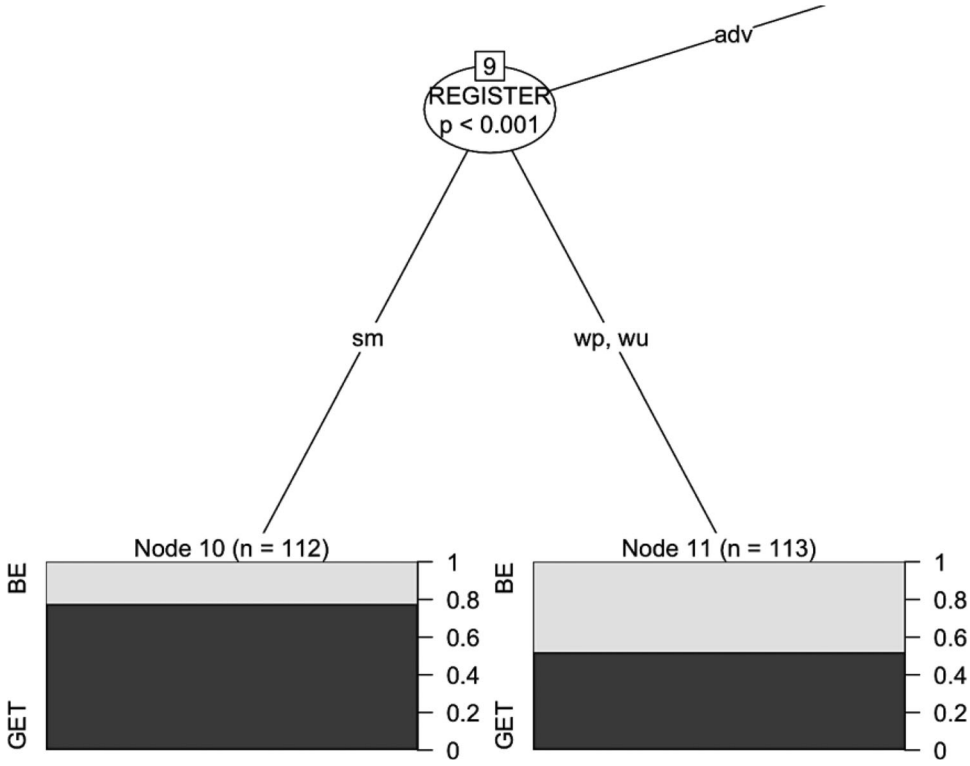

FIGURE 4a Partial glmertree for passives with human in sm, wp and wu.

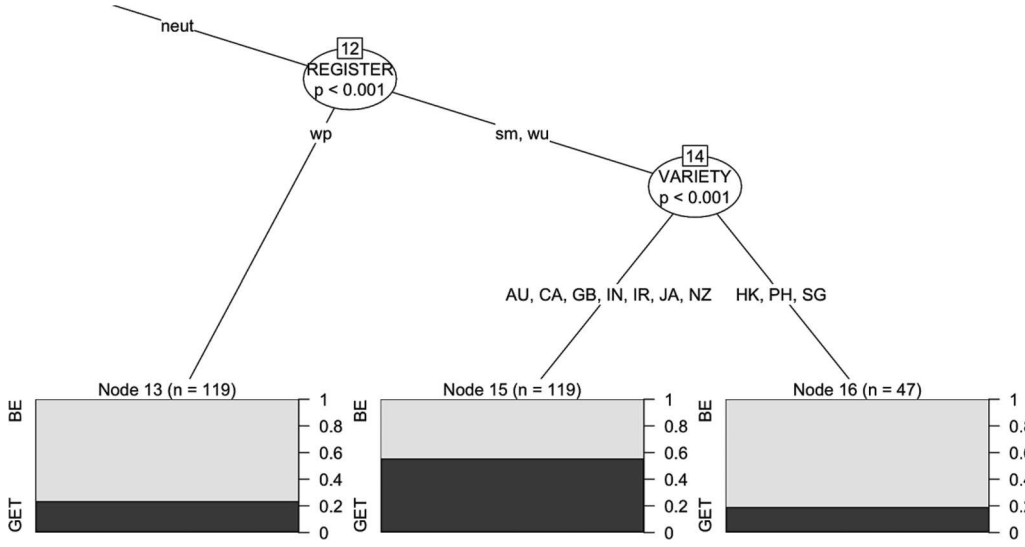

FIGURE 4b Partial glmertree for passives animate subjects in sm, wp and wu.

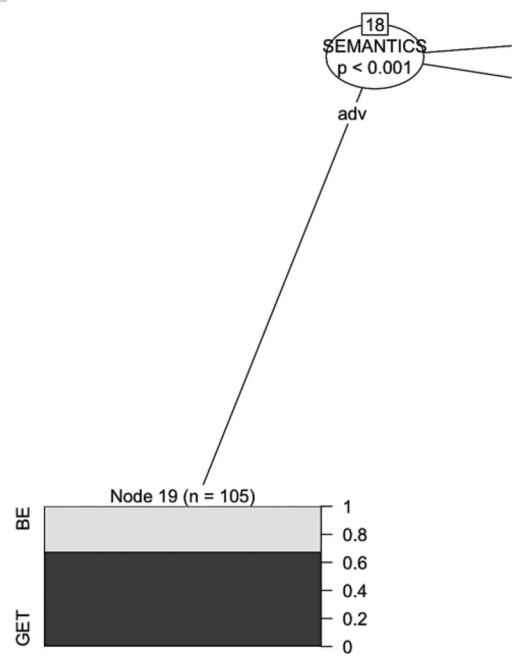

FIGURE 5 Partial glmertree for passives with inanimate subjects in spoken, adversative contexts.

In non-adversative contexts (Figure 6), we see that spoken monologues disprefer GET-passives more than spoken dialogue (Node 20). For both contexts, we see a split by VARIETY (Nodes 21 and 24), but again without any grouping by region, variety-type or matrilect.

Finally, if we turn to the right-most part of the glmertree, we see that for inanimate subjects in written texts, the model shows an early split by VARIETY, with most varieties (Figure 7a) showing additional interaction with SEMANTICS (Node 35), with adversative contexts providing more of an opening for GET-passives in written texts (Node 36),

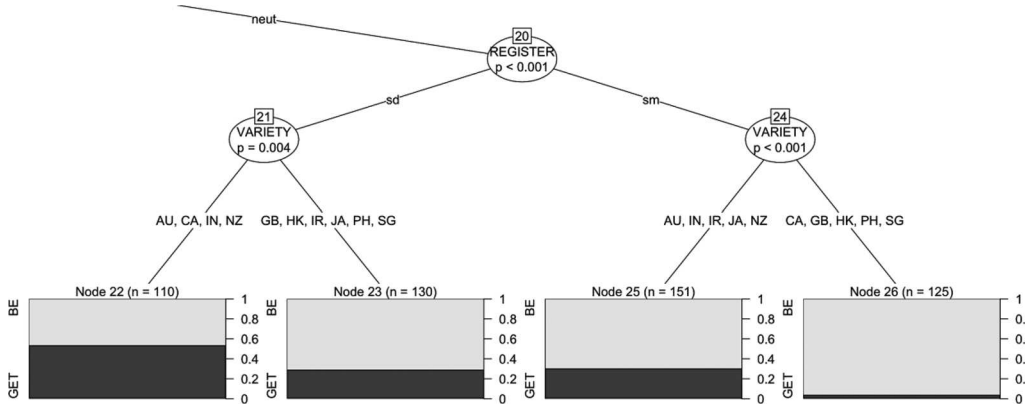

FIGURE 6 Partial glmertree for neutral passives with inanimate subjects in spoken interaction.

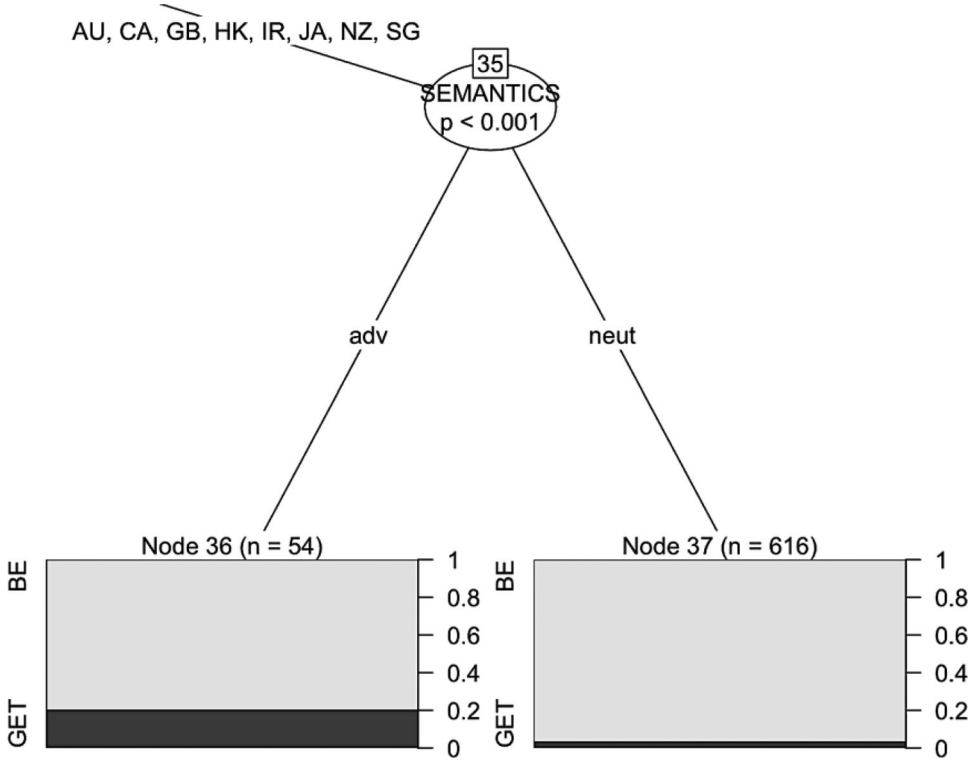

FIGURE 7a Partial glmertree for passives with inanimate subjects in written texts.

which are otherwise dominated (both proportionally and with respect to raw frequencies) by BE-passives (Node 37), as expected.

In written IndE and PhilE, the preference for GET-passives with inanimate subjects and in adversative contexts is more pronounced than in the other varieties (Node 29 in Figure 8). For non-adversative, printed contexts, IndE is somewhat more open to using GET in passives than PhilE (Nodes 32 and 33). In unprinted writing, both varieties are somewhat more open to using GET-passives with inanimate subjects (Node 34). This shows that OC varieties like PhilE and IndE are atuned to the stylistic differences between the two passive allostructions.

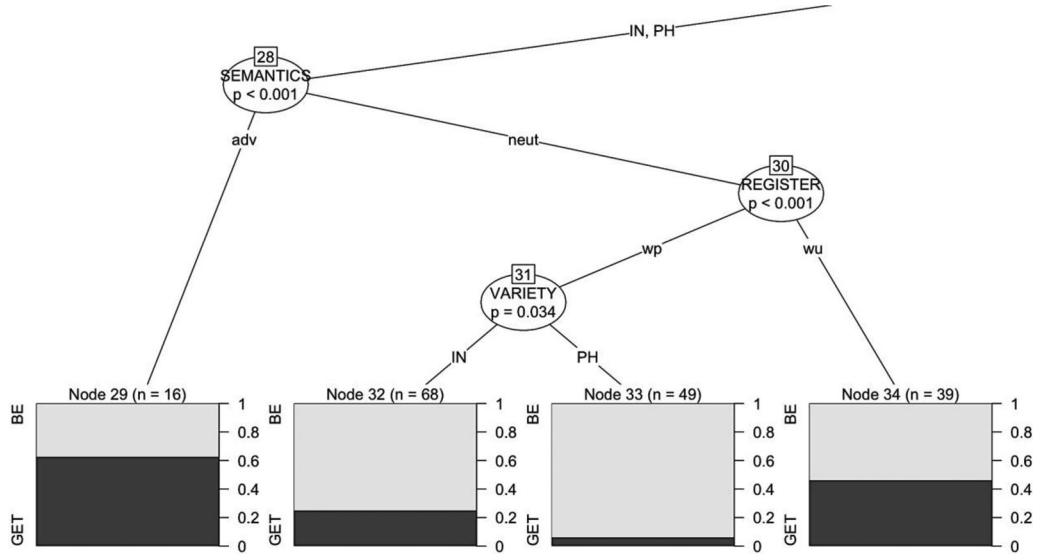

FIGURE 7b Partial glmertree for passives with inanimate subjects in written texts (IndE & PhilE).

Overall, the left-hand side of the tree with human and animate subjects attracts a lot of the GET-passives in our data, whereas the right-hand side of the tree features many of the contexts favouring BE-passives (such as more formal spoken or printed written production). Even though VARIETY is repeatedly returned as a significant predictor in the glmertree, there is no systematic grouping of varieties according to either variety type (IC vs. OC), region (Northern- vs. Southern Hemisphere, Asian vs. Non-Asian) or diachronic connection (BrE-based vs. AmE-based varieties).

Finally, turning to the random effect, we can take a look at the verbs that trend strongly towards the GET-passive and those that trend strongly towards the BE-passive (with positive and negative values in the glmertree, respectively). Table 4 lists the 10 verbs found at either end of the spectrum.[11] Surprisingly, only a few adversative verbs (*push* and *kick*) are found to strongly favour the GET-allostruction whereas a greater number (*force*, *remove*, *cancel*, *abuse* and *exclude*) are attracted to the variant with auxiliary BE. Our study thus lends further support to those studies (Fehringer, 2022; Bohmann et al., 2023) that have questioned the importance of semantic constraints for the choice between the two passive allostructions

5 | DISCUSSION

The RF analysis of the BE:GET-passive alternation in our ICE data shows that, with the exception of the *by*-phrase, all factors play a role in the choice of catenative, with REGISTER and VARIETY ranking highest. However, in a mixed model with V_LEMMA as a random effect, the semantics of the subject turns out to be the most important predictor variable. Thus, the GET-passive's more agentive origins are still reflected in a greater propensity to be used with human subjects. The BE-passive, in turn, combines more freely with inanimate subjects across all world Englishes in our ICE data. Both constructions repel co-occurrence with an external *by*-agent.

Somewhat surprisingly, our ICE data do not support the prediction that catenative GET combines more easily with a progressive passive, nor did we find a particular attraction of IndE towards progressive passives. Instead, CanE returned the highest overall number of progressive passives (17). In other words, we found neither the expected long shadow of prescriptive influence in a marked avoidance of progressive BE-passives nor any obvious substrate influence.

TABLE 4 Most strongly attracted verb lemmas in ICE (adversative verbs in italics).

V_LEMMA	Value	Catenative (dominant)
pollinate	2.2888924	GET
draft	2.0052142	GET
transcribe	2.00111002	GET
pay	1.9240397	GET
realise	1.82570428	GET
push	1.77421535	GET
kick	1.76461829	GET
select	1.72673255	GET
go	1.691396	GET
change	1.68440277	GET
….		
force	−2.4160464	BE
remove	−2.1532057	BE
cancel	−1.9478029	BE
allow	−1.8830197	BE
make	−1.7309546	BE
give	−1.6011761	BE
find	−1.5790581	BE
treat	−1.5534554	BE
abuse	−1.4764696	BE
exclude	−1.4711665	BE

The glmertree returns VARIETY as a much lower-ranking predictor variable than the RF. Our multifactorial modelling of the BE:GET-passive alternation across a broad range of world Englishes thus resembles previous research (Szmrecsanyi et al., 2016) in confirming that language-internal factors (in our case subject animacy and semantics of the construction) play a much greater role in the choice of passive construction than VARIETY. However, none of the splits in the glmertree reveal any predictable groupings by variety type (IC vs. OC), area (northern vs. southern hemisphere), or diachronic connection (BrE- or AmE-based). In other words, the prediction that IC varieties are, overall, more attuned to the colloquial origin of the GET-passive does not find support in our ICE data. Only two OC varieties – IndE and PhilE – turn out to be somewhat less sensitive to the GET-passive's colloquial connotations whereas the other OC varieties patterned with the IC Englishes in their tendency to avoid GET-passives in written production. Seeing that VARIETY turns out to be a rather low-ranking and unsystematic factor in the choice between the two passive allostructions, we would not like to take them as a sign of 'the emergence of a unique, region-specific grammar' (Szmrecsanyi et al., 2016: 133). In conclusion, what our case study supports instead is a view whereby the core grammar of most world Englishes is shared and develops in a lock-step fashion along a broadly similar trajectory, with stylistic or semantic localisation effects that are likely to be relatively transient in nature (see Mair, 2015: 124; Hansen, 2018). While the two passive allostructions are largely interchangeable, the main predictor variable (semantics of the subject) still reflects the different origins of the allostructions and surface in the GET-passives preference for human and animate subjects, a factor that plays out in all world Englishes investigated here. Seeing that our data come from a set of ICE components that were collected at the end of the twentieth century, it would be interesting to see whether more recent evidence, for example, from the GloWbE or NOW corpora support our results on a shared world Englishes

grammar of the passive alternation and the notion of converging grammars. We predict that such follow-up studies will show converging trends of language-internal predictors across varieties of world Englishes.

ACKNOWLEDGEMENTS

Work on this paper was mostly carried out during the autumn semester 2023 when the corresponding author was on research leave at the University of Kyoto. The support of both UZH and Kyoto University is gratefully acknowledged. We also acknowledge the constructive feedback of the two anonymous reviewers on an earlier version of the paper.

Open access funding provided by Universitat Zurich.

CONFLICT OF INTEREST

On behalf of my co-authors I hereby affirm that there is no conflict of interest.

ORCID

Marianne Hundt https://orcid.org/0000-0001-6045-2314

ENDNOTES

[1] See, for example, Hundt (2007: 69–72) or Mitkovska & Bužarovska (2012) for overlap between GET + participle and middle constructions.

[2] Access to ICE online is by registration and user password, see https://www.ice-corpora.uzh.ch/en/access.html for details.

[3] We initially only retrieved the data for the IC varieties, using a window of five words between the catenative and the participle, as the following regular expressions for the two passive variants show:

Regular expression for potential BE-passives: \b(be_VB|was_VBD|were_VBD|being_VBG|been_VBN|are_VBP|am_VBP|is_VBZ)\s+(\w+\s+){0,5}(\w+_VBN)

Regular expression for potential GET-passives: \b(get_VB|got_VBD|getting_VBG|got_VBN|gotten_VBN|get_VBP|gets_VBZ)\s+(\w+\s+){0,5}(\w+_VBN)

The manual post-editing of the data showed that a window of four words would suffice to-retrieve the relevant GET-passives, and the queries were adapted accordingly for the OC varieties.

[4] The emphasis in the corpus examples has been added throughout. They are given in their original spellings (for instance, all lower-case for the New Zealand component). The code following the corpus label provides information on the text file (see Greenbaum, 1996 for details).

[5] On download from the ICE-Dependency interface, the resulting concordance lines are automatically randomised, facilitating the analysis of a randomized subset of data.

[6] Note that abstract, inanimate *prospect* would not be a prototypical agent semantically, either.

[7] For a comparison of the standard approach to mixed modelling with the combined approach, see Fokkema et al. (2020).

[8] In fact, there is not a single occurrence of a modal BE-passive in the progressive in our data.

[9] See Hundt (2016) for a detailed discussion on the matter of the allography between *been* and *being* as perfect auxiliaries.

[10] We used bagging, that is, manually set mtry to include all predictor variables for splitting a node (for details, see https://epub.ub.uni-muenchen.de/59094/1/MA_Hatz.pdf).

[11] For a complete list, see https://osf.io/dx5bz/.

REFERENCES

Alexiadou, A. (2006). A note on non-canonical passives: the case of the *get*-passive. In H. Broekhuis, N. Corver, R. Huybregts, U. Kleinhenz & J. Koster (Eds.), *Organizing grammar: Linguistic studies in honor of Henk van Riemsdijk* (pp. 13–21). Berlin: de Gruyter. https://doi.org/10.1515/9783110892994.13

Alexiadou, A. (2012). Noncanonical passives revisited: Parameters of nonactive voice. *Linguistics*, *50*(6), 1079–1110. https://doi.org/10.1515/ling-2012-0036

Anderwald, L. (2016). *Language between description and prescription. Verbs and verb categories in nineteenth-century grammars of English*. Oxford: Oxford University Press.

Bergen, L. (2013). Early progressive passives. *Folia Linguistica Historica*, *34*, 173–208. https://doi.org/10.1515/flih.2013.007

Biewer, C. (2009). Passive constructions in Fiji English: A corpus-based study. In A. Jucker, D. Schreier & M. Hundt (Eds.), *Corpora: Pragmatics and Discourse* (pp. 361–377). Amsterdam: Rodopi.

Bohmann, A., Müller, J., Honkanen, M., & Neuhausen, M. (2023). A large-scale diachronic analysis of the English passive alternation. In B. Busse & I. Warnke (Eds.), *Language and linguistics in a complex world* (pp. 31–55). Berlin: de Gruyter. https://doi.org/10.1515/9783111017433-003

Brinton, L J. (1994). The differentiation of statives and perfects in Early Modern English: The development of the conclusive perfect. In D. Stein & I. T-B. van Ostade (Eds.), *Towards a standard English, 1600–1800* (pp. 135–170). Berlin: de Gruyter. https://doi.org/10.1515/9783110864281

Bruckmaier, E. (2016). Dialect contact influences on the use of *get* and the *get*-passive. In O. Timofeeva, A-C. Gardner, A. Honkapohja & S. Chevalier (Eds.), *New approaches to English linguistics: Building bridges* (pp. 117–140). Amsterdam: Benjamins. https://doi.org/10.1075/slcs.177.05bru

Bruckmaier, E. (2017). *Getting at GET in World Englishes*. Berlin: de Gruyter. https://doi.org/10.1515/9783110497311

Cappelle, B. (2006). Particle placement and the case for 'allostructions'. In D. Schönefeld (Ed.), *Constructions special volume 1 – constructions all over: Case studies and theoretical implications*. Retrieved from: www.researchgate.net/publication/31590515_Particle_placement_and_the_case_for_allostructions

Carletta, J. (1996). Assessing agreement on classification tasks: the kappa statistic. *Computational Linguistics*, *22*(2), 249–254. https://aclanthology.org/J96-2004.pdf

Carter, R., & McCarthy, M. (1999). The English *get*-passive in spoken discourse: Description and implications for an interpersonal grammar. *English Language and Linguistics*, *3*(1), 41–58. https://doi.org/10.1017/S136067439900012X

Chappell, H. (1980). Is the *get*-passive adversative? *International Journal of Human Communication*, *13*(3), 411–452. https://doi.org/10.1080/08351818009370504

Collins, P. C. (1996). *Get*-passives in English. *World Englishes*, *15*(1), 43–56. https://doi.org/10.1111/j.1467-971X.1996.tb00091.x

Coto-Villalibre, E. (2014). Get + past participle constructions in present-day spoken Indian English: Exploring the passive gradient. In P. Rodríguez-Puente, T. Fanego, E. Gandón-Chapela, S. Riveiro-Outeiral & M. L. Roca-Varela (Eds.), *Current research in applied linguistics: Issues on language and cognition* (pp. 10–30). Newcastle Upon Tyne: Cambridge Scholars Publishing.

Coto-Villalibre, E. (2015). Down the passive gradient: From agentive to borderline *get* + past participle constructions in Singaporean English. In C. Suárez-Gómez & E. Seoane (Eds.), *Englishes today: Multiple varieties, multiple perspectives* (pp. 81–102). Cambridge: Cambridge Scholars Publisher.

Coto-Villalibre, E. (2016). A look at participial constructions with *get* in Hong Kong English. In M. J. López-Couso, B. Méndez-Naya, P. Núñez-Pertejo & I. M. Palacios-Martínez (Eds.), *Corpus linguistics on the move. Exploring and understanding English through corpora* (pp. 204–226). Amsterdam: Brill. https://doi.org/10.1163/9789004321342_011

Downing, A. (1996). The semantics of *get*-passives. In R. Hasan, C. Cloran & D. G. Butt (Eds.), *Functional descriptions: Theory in practice* (pp. 179–205). Amsterdam: Benjamins. https://doi.org/10.1075/cilt.121

Fehringer, C. (2022). The *get*-passive in Tyneside English. A highly frequent but constrained variant. *English World-Wide*, *43*(3), 330–356. https://doi.org/10.1075/eww.21039.feh

Fischer, O. (1992). Syntax. In N. Blake (Ed.), *The Cambridge history of the English language, volume II, 1066–1476* (pp. 207–408). Cambridge: Cambridge University Press. https://doi.org/10.1017/CHOL9780521264754.005

Fokkema, M., Smits, N., Zeileis, A., Hothorn, T., & Kelderman, H. (2018). Detecting treatment- subgroup interactions in clustered data with generalized linear mixed-effects model trees. *Behavior Research Methods*, *50*(5), 2016–2036. https://doi.org/10.3758/s13428-017-0971-x

Fokkema, M., & Zeileis, A. (2019). Package 'glmertree'. https://cran.rproject.org/web/packages/glmertree/glmertree.pdf

Fokkema, M., Edbrooke-Childs, J., & Wolpert, M. (2020). Generalized linear mixed-model (GLMM) trees: A flexible decision-tree method for multilevel and longitudinal data. *Psychotherapy Research*, *31*, 1–13. https://doi.org/10.1080/10503307.2020.1785037

Fong, V. (2004). The verbal cluster. In L. Lim (Ed.), *Singapore English. A grammatical description* (pp. 75–104). Amsterdam: Benjamins. https://doi.org/10.2307/2910009

Givón, T. (1993). *English grammar. A function-based introduction*. Volume II. Amsterdam: Benjamins. https://doi.org/10.1075/z.engram2

Goldberg, A. E. (2013). Constructionist approaches to language. In T. Hoffmann & G. Trousdale (Eds.), *Handbook of construction grammar* (pp. 15–31). Oxford: Oxford University Press.

Greenbaum, S., ed. (1996). *Comparing English worldwide: The International Corpus of English*. Oxford: Clarendon Press. https://doi.org/10.1017/S1351324998211922

Hansen, B. (2018). *Corpus linguistics and sociolinguistics. A study of variation and change in the modal systems of World Englishes*. Leiden: Brill.

Hatcher, A. G. (1949). To get/be invited. *Modern Language Notes*, *64*(7), 433–446. https://doi.org/10.2307/2910009

Hundt, M. (1998). *New Zealand English grammar – fact or fiction?* Amsterdam: Benjamins. https://doi.org/10.1075/eww.20.2.10hol

Hundt, M. (2001). What corpora tell us about the grammaticalisation of voice in *get*-constructions. *Studies in Language*, *25*(1), 49–88. https://doi.org/10.1075/sl.25.1.03hun

Hundt, M. (2007). *English mediopassive constructions. A cognitive, corpus-based study of their origin, spread and current status* Amsterdam: Rodopi.

Hundt, M. (2009). How often do Things *get* V-*ed* in Philippine and Singapore English? A case study on the *get*-passive in two outer-circle varieties of English. In R. Bowen, M. Mobärg & S. Ohlander (Eds.), *Corpora and discourse – and stuff. Papers in honour of Karin Aijmer* (pp. 121–129). Göteborg: Acta Universitatis Gothoburgensis.

Hundt, M. (2016). Error, feature, (incipient) change – or something else altogether? In E. Seoane & C. Suárez-Gómez (Eds.), *World Englishes: New theoretical and methodological considerations* (pp. 37–60). Amsterdam: Benjamins. https://doi.org/10.1075/veaw.g57.03hun

Hundt, M. (2020). Corpus-based approaches to World Englishes. In D. Schreier, M. Hundt & E. W. Schneider (Eds.), *The Cambridge handbook of World Englishes* (pp. 506–533). Cambridge: Cambridge University Press. https://doi.org/10.1017/9781108349406

Hundt, M., Hay, J., & Gordon, E. (2008). New Zealand English: morphosyntax. In K. Burridge & B. Kortmann (Eds.), *Varieties of English: The Pacific and Australasia, handbook of varieties of English, volume III* (pp. 305–340). Berlin: de Gruyter. https://doi.org/10.1515/9783110208412.2.305

Hundt, M., Schneider, G., & Seoane, E. (2016). The use of the *be*-passive in academic Englishes: Local versus global usage in an international language. *Corpora*, *11*(1), 29–61. https://doi.org/10.3366/cor.2016.0084

Hundt, M., Röthlisberger, M., & Seoane, E. (2021). Predicting voice alternation across academic Englishes. *Corpus Linguistics and Linguistic Theory*, *17*(1), 189–222. https://doi.org/10.1515/cllt-2017-0050

Kim, C., & Sato, Y. (2013). More on *kena*-passives in Singapore English. *World Englishes*, *32*(3), 297–307. https://doi.org/10.1111/weng.12032

Leech, G., Hundt, M., Mair, C., & Smith, N. (2009). *Change in contemporary English: A grammatical study*. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511642210

Lehmann, H. M., & Schneider, G. (2012). Dependency bank. *LREC 2012 Conference Workshop "Challenges in the Management of Large Corpora"*. Istanbul, Turkey.

Mair, C. (2006). *Twentieth-Century English. History, variation and standardization*. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511486951

Mair, C. (2015). Response to Davies and Fuchs. *English World-Wide*, *36*(1), 29–33. https://doi.org/10.1075/eww.36.1.02mai

Matthews, R. (1993). *Papers on semantics and grammar*. Frankfurt am Main: Peter Lang.

Mitkovska, L., & Bužarovska, E. (2012). An alternative analysis of the English get-past participle constructions: Is get all that passive? *Journal of English Linguistics*, *40*(2), 196–215.

Petré, P. (2010). The functions of weorðan and its loss in the past tense in Old and Middle English. *English Language and Linguistics*, *14*(3), 457–484. https://doi.org/10.1017/S1360674310000158

Quirk, R., Greenbaum, S., Leech, G., & Svartvik, J. (1985). *A comprehensive grammar of the English Language*. London: Longman.

Rühlemann, C. (2007). Lexical grammar: The *get*-passive as a case in point. *ICAME Journal*, *31*, 111–127.

Schneider, E. W. (2007). *Postcolonial Englishes: Varieties around the world*. Cambridge: Cambridge University Press.

Schwarz, S. (2017). 'Like getting nibbled to death by a duck.' Grammaticalization of the *get*-passive in the TIME Magazine Corpus. *English World-Wide*, *38*(3), 305–335. https://doi.org/10.1075/eww.38.3.03sch

Sharma, D. (2009). Typological diversity in new Englishes. *English World-Wide*, *30*(2), 170–195. https://doi.org/10.1075/eww.30.2.04sha

Smith, N., & Rayson, P. (2007). Recent change and variation in the British English use of the progressive passive. *ICAME Journal*, *31*, 129–159. https://icame.info/icame_static/ij31/ij31-page129-160.pdf

Strobl, C., Hothorn, T., & Zeileis, A. (2009). Party on! A new, conditional variable-importance measure for random forests available in the party package. *The R Journal*, *1*(2), 14–17. https://journal.r-project.org/archive/2009-2/RJournal_2009-2_Strobl~et~al.pdf

Sussex, R. (1982). A note on the *get*-passive construction. *Australian Journal of Linguistics*, *2*(1), 83–95. https://doi.org/10.1080/07268608208599283

Svartvik, J. (1966). *On voice in the English verb*. The Hague: Mouton. https://doi.org/10.1017/S0022226700002565

Szmrecsanyi, B., Grafmiller, J., Heller, B., & Röthlisberger, M. (2016). Around the world in three alternations. Modeling syntactic variation in varieties of English. *English World-Wide*, *37*(2), 109–137. https://doi.org/10.1075/eww.37.2.01szm

Thompson, D., Ling, S. P., Myachykov, A., Ferreira, F., & Schaepers, C. (2013). Patient-related constraints on *get*- and *be*-passive uses in English: Evidence from paraphrasing. *Frontiers in Psychology*, *4*, 848. https://doi.org/10.3389/fpsyg.2013.00848

Thompson, D., Ferreira, F., & Scheepers, C. (2018). One step at a time: representational overlap between active voice, *be*-passive, and *get*-passive forms in English. *Journal of Cognition*, *1*(1), 35. https://doi.org/10.5334/joc.36

Vanrespaille, M. (1991). A semantic analysis of the English *get*-passive. *Journal of Applied Linguistics*, *5*(2), 95–112.

Wanner, A. (2009). *Deconstructing the English passive*. Berlin: de Gruyter. https://doi.org/10.1515/9783110199215

Werner, V. (2016). Rise of the undead? BE-perfects in World Englishes. In V. Werner, E. Seoane & C. Suárez-Gómez, (Eds.), *Re-assessing the present perfect* (pp. 259–294). Berlin: de Gruyter. https://doi.org/10.1515/9783110443530-012

Zaenen, A., Carletta, J., Garretson, G., Bresnan, J., Koontz-Garboden, A., Nikitina, T., Catherine O'Connor, M., & Wasow, T. (2004). Animacy encoding in English: Why and how. *Proceedings of the Workshop on Discourse Annotation*. Barcelona: Association for Computational Linguistics, 118–125. https://aclanthology.org/W04-0216.pdf

How to cite this article: Hundt, M., Dallas, B., & Nakanishi, S. (2023). The be- versus *get*-passive alternation in world Englishes. *World Englishes*, 1–24. https://doi.org/10.1111/weng.12633

APPENDIX

Generalized linear mixed model tree

Model formula:

CATENATIVE ~ 1|VARIETY + REGISTER + ANIMACY + SEMANTICS

Fitted party:
[1] root
| [2] ANIMACY in a, h
| | [3] REGISTER in sd
| | | [4] SEMANTICS in adv
| | | | [5] VARIETY in GB, HK: n = 16
| | | | (Intercept)
| | | | -0.00053731
| | | | [6] VARIETY in AU, CA, IN, IR, JA, NZ, PH, SG: n = 177
| | | | (Intercept)
| | | | 2.762841
| | | [7] SEMANTICS in neut: n = 243
| | | (Intercept)
| | | 1.054015
| | [8] REGISTER in sm, wp, wu
| | | [9] SEMANTICS in adv
| | | | [10] REGISTER in sm: n = 112
| | | | (Intercept)
| | | | 1.144378
| | | | [11] REGISTER in wp, wu: n = 113
| | | | (Intercept)
| | | | 0.007649042
| | | [12] SEMANTICS in neut
| | | | [13] REGISTER in wp: n = 119
| | | | (Intercept)
| | | | -1.234204
| | | | [14] REGISTER in sm, wu
| | | | | [15] VARIETY in AU, CA, GB, IN, IR, JA, NZ: n = 119
| | | | | (Intercept)
| | | | | 0.2387996
| | | | | [16] VARIETY in HK, PH, SG: n = 47
| | | | | (Intercept)
| | | | | -1.567275
| [17] ANIMACY in i
| | [18] REGISTER in sd, sm

| | | [19] SEMANTICS in adv: n = 105
| | | (Intercept)
| | | 0.683933
| | | [20] SEMANTICS in neut
| | | | [21] REGISTER in sd
| | | | | [22] VARIETY in AU, CA, IN, NZ: n = 110
| | | | | (Intercept)
| | | | | 0.1044477
| | | | | [23] VARIETY in GB, HK, IR, JA, PH, SG: n = 130
| | | | | (Intercept)
| | | | | -1.018943
| | | | [24] REGISTER in sm
| | | | | [25] VARIETY in AU, IN, IR, JA, NZ: n = 151
| | | | | (Intercept)
| | | | | -0.9601668
| | | | | [26] VARIETY in CA, GB, HK, PH, SG: n = 125
| | | | | (Intercept)
| | | | | -3.308755
| | [27] REGISTER in wp, wu
| | | [28] VARIETY in IN, PH
| | | | [29] SEMANTICS in adv: n = 16
| | | | (Intercept)
| | | | 0.8931954
| | | | [30] SEMANTICS in neut
| | | | | [31] REGISTER in wp
| | | | | | [32] VARIETY in IN: n = 68
| | | | | | (Intercept)
| | | | | | -1.385596
| | | | | | [33] VARIETY in PH: n = 49
| | | | | | (Intercept)
| | | | | | -2.617028
| | | | | [34] REGISTER in wu: n = 39
| | | | | (Intercept)
| | | | | -0.1647182
| | | [35] VARIETY in AU, CA, GB, HK, IR, JA, NZ, SG
| | | | [36] SEMANTICS in adv: n = 54
| | | | (Intercept)
| | | | -1.657665
| | | | [37] SEMANTICS in neut: n = 616
| | | | (Intercept)
| | | | -3.308332
Number of inner nodes: 18
Number of terminal nodes: 19
Number of parameters per node: 1
Objective function (negative log-likelihood): 712.8622