

# Auxiliary DO in Asian Englishes

Ninja Schulz

Julius-Maximilians-Universität Würzburg,  
Germany

## Correspondence

Ninja Schulz, Julius-Maximilians-Universität  
Würzburg, Department for English Linguistics,  
Am Hubland, 97074 Würzburg, Germany.  
Email: [ninja.schulz@uni-wuerzburg.de](mailto:ninja.schulz@uni-wuerzburg.de)

## Abstract

To indicate emphasis, auxiliary DO is used in affirmative contexts (DO+) when no other auxiliary is present. It is thus rooted in the grammatical system of do-support; however, DO+ does not always bear stress and can fulfil various discourse-marking functions (Nevalainen & Rissanen, 1986). Positioned at the intersection of grammar and discourse, DO+ constitutes an interesting study for its use in ‘non-native’ varieties of English since it can be assumed that the more salient grammatical functions are easier to master for learners. Focusing on Asian Englishes in contrast to Inner Circle varieties, this exploratory paper assesses the frequency and distribution of DO+ in the spoken and written parts of eight ICE components.

## 1 | INTRODUCTION

In Present-Day English (PDE), do-support refers to the use of auxiliary DO in negative, inverted or elliptic clauses, and in emphatic contexts:

|                    |  |
|--------------------|--|
| Negation:          | <i>I do not think this is relevant</i> |
| Inversion:         | <i>Do you think this is relevant?</i>  |
| Code/Ellipsis:     | <i>So do I.</i>                        |
| Emphasis/Contrast: | <i>I do think this is relevant.</i>    |

When no other auxiliary is present, this ‘dummy do’ is inserted to function as operator in the clause. In reference grammars of English, these grammaticalised uses of auxiliary DO are most commonly treated in connection with these ‘NICE properties’ of auxiliaries (Huddleston & Pullum, 2002, p. 93). In this paper, I will focus on the use of DO in affirmative contexts (henceforth DO+), which, in the context of the NICE properties, is either categorised as expressing ‘emphasis’ or ‘contrast’. According to Quirk et al. (1985), DO+ expresses either emotive or contrastive emphasis, with DO carrying

-----  
This is an open access article under the terms of the [Creative Commons Attribution](https://creativecommons.org/licenses/by/4.0/) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2021 The Authors. *World Englishes* published by John Wiley & Sons Ltd

stress which would otherwise fall on the main verb. From this perspective, DO+ results from a grammatical operation and its omission is unacceptable in standardised English.<sup>1</sup> However, studies have shown that DO+ is not always stressed and can, for example, be used to (re)introduce, end or elaborate a topic (Nevalainen & Rissanen, 1986) or to take explicit stance (Ranger, 2015):

Reintroducing a topic: *I did mention earlier that...*

Subjective endorsement: *I do believe that he is unfaithful.*

In the latter case, DO may be stressed but does not necessarily imply contrast: it is used parallel to *I really believe* (Ranger, 2015, p. 117). Thus, while the use of auxiliary DO in questions, negative and elliptic statements is obligatory in present-day standardised English, its use in affirmative contexts cannot be wholly covered by the labels 'emphasis' or 'contrast', nor can its omission be considered ungrammatical. In these contexts, its use is dependent on the speaker's choice and it serves some discourse-pragmatic function connected to the presentation of a proposition in relation to the previous discourse.

For learners of English, DO+ may pose particular difficulties due to its extreme versatility in function, which it shares with discourse markers. Thus, it can be expected that DO+ is less frequent in postcolonial varieties of English, at least in contexts where its use is not obligatory. To explore this for Asian varieties, this paper assesses the use and distribution of DO+ in eight components of The International Corpus of English (ICE), with British, American, New Zealand, and Canadian English for the Inner Circle and Indian, Hong Kong, Singapore, and Philippine English for the Outer Circle. The corpora, which provide a (largely) comparable set of different text types, are used in this paper to find out whether there are systematic differences in the use of DO+ between the Inner Circle and the Asian varieties in terms of frequency and distribution, whether specific functions are preferred or avoided in the postcolonial Englishes, and whether new functions have emerged. Distributional differences are analysed by calculating the normalised frequency of DO+ per 10,000 words (pttw) for spoken and written language as well as for different text types. Besides this general quantitative overview, the occurrences are analysed qualitatively for single text types and varieties in cases where the frequencies indicate variety-specific patterns.

Section 2 gives an overview of the distribution of DO+ in British English (BrE) and American English (AmE) to contextualise the patterns found in the Asian varieties. Section 3 focuses on the functional range of DO+, exploring its emphatic functions described in standard grammars as well as discursive and interpersonal functions, and discusses similarities to properties attributed to discourse markers. Approaching DO+ from the perspective of its discourse-marking functions will contribute to our understanding of its use and distribution in postcolonial Englishes. On the other hand, including a discourse-structuring device that is more clearly linked to its grammatical source construction might shed further light on the development and usage of discourse markers, not only in postcolonial Englishes but also in 'native' varieties. This paper can only be a first step into this direction since a fully-fledged functional categorisation of DO+ is still pending for any variety of English. My principle aim is to find out whether there are indeed systematic differences in frequency and distribution between Inner Circle and Asian varieties, which will offer first insights into potential differences in function.

## 2 | DISTRIBUTION OF DO+ IN BRITISH AND AMERICAN ENGLISH

Even though DO+ has received considerable attention in historical linguistics, there is a lack of corpus-based studies which explore its distribution in PDE. The only quantitative assessments of DO+ are, to my knowledge, a short section in Biber et al.'s (1999, pp. 24–35) grammar based on The Longman Spoken and Written English Corpus (LSWE Corpus), which contains material from BrE and AmE as well as from some other varieties of English and Nevalainen and Rissanen's (1986) study on its use in BrE. Based on the London-Lund Corpus (LLC) and the Lancaster-Oslo/Bergen Corpus (LOB) for spoken and written language, respectively, Nevalainen and Rissanen (1986, p. 47) found that DO+

**TABLE 1** Distribution of DO+ in spoken and written English

|             | Conversation | Fiction | Press | Academic prose |
|-------------|--------------|---------|-------|----------------|
| LOB/LLC     | 9.1          | 2.9     | 1.6   | 1.9            |
| LWSE Corpus | 4            | 3       | 1.5   | 1.5            |

(Based on Nevalainen & Rissanen, 1986, p. 47, Biber et al., 1999, p. 433)

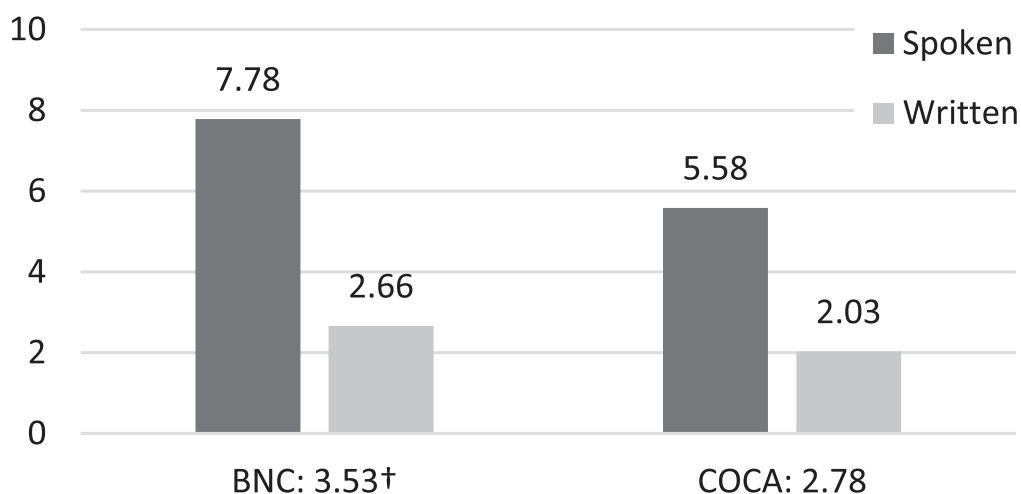
**TABLE 2** Number of words per sub-corpus from BNC and COCA

|         | BNC        | COCA        |
|---------|------------|-------------|
| Spoken  | 9,963,663  | 22,160,125  |
| Written | 48,969,398 | 82,046,457  |
| Total   | 58,933,061 | 104,206,582 |

is thrice as frequent in spoken than in written language and more common in fiction than in instructive writing. Biber et al. (1999, p. 433) find a similar distribution though the frequency of DO+ in conversations is considerably lower than in the LLC so that the difference between spoken and written language is much less marked here. Table 1 summarises these findings as normalised frequencies pttw.<sup>2</sup> However, as the corpora used are not comparable (especially in terms of time and variety), they cannot serve as a basis for comparing the distribution in BrE and AmE. It can only be speculated that DO+ is less common in conversations in AmE, which would be an explanation for the marked difference between the results for this category.

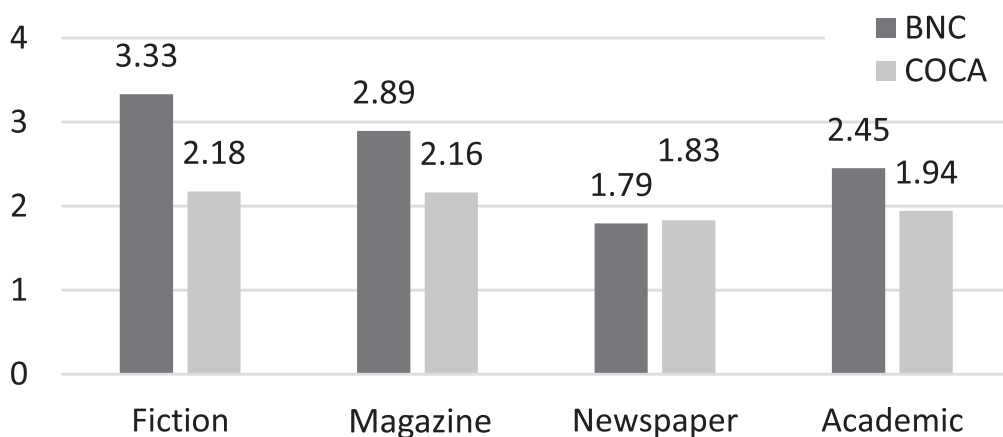
Thus, the following assessment of the distribution of DO+ relies on my own preliminary study based on the British National Corpus (BNC), the Corpus of Contemporary American English (COCA), and the corpora of the Brown-family.<sup>3</sup> Although the analysis within this paper includes ICE-GB and ICE-US, I consider it worthwhile to include a broader database to assess general differences between BrE and AmE and to establish genre- or medium-specific differences which are not solely based on the limited data found in ICE. The preliminary corpus analysis is based on subsets of the BNC and COCA, counting any instance of *do*, *does*, and *did* followed by an infinitive and allowing for an intervening adverb.<sup>4</sup> Only categories used in both corpora were included ('conversation', 'fiction', 'magazine', 'newspaper', and 'academic prose'). In addition, only the earliest period (1990–1994) of COCA was used to increase comparability with the BNC and with ICE. The number of words per sub-corpus, displayed in Table 2, is, though unbalanced, still sufficiently large to indicate general differences between the two varieties. The results in Figure 1 show, on the one hand, the expected tendency that DO+ is more frequent in spoken language in both varieties. On the other hand, it is generally more frequent in BrE, both in spoken and in written language. Figure 2 shows that this distribution is also found in all written sub-categories, except 'newspaper', so that the distribution in written language is not due to single text types, but a general difference between AmE and BrE. Looking at this difference in written language from a historical perspective on the basis of the corpora of the Brown-family shows that in the 1960s (Figure 3) there is no noticeable difference between either the different categories or the varieties.<sup>5</sup> The distribution found in COCA and BNC, however, is reflected in Frown and FLOB (Figure 4), where there is a clear tendency of DO+ to occur more frequently in fiction than in the more formal genres.

In AmE, there is a noticeable increase of DO+ in fiction, while it slightly decreases in the other categories, indicating the avoidance of DO+ in more formal language. This seems to suggest a process of colloquialisation, which is strongest in popular text types, especially fiction, while more formal text types are less affected (Biber & Gray, 2016, p. 140). In BrE, DO+ increases in all genres, but fiction is clearly in the lead of this development. The following observations are the most important for this paper: first, DO+ is more frequent in spoken language but seems to be spreading into written language (Hundt, 1998, p. 57). Second, DO+ is more frequent in BrE than in AmE in both spoken and written language. Third, in written language, this gap between AmE and BrE has only developed within the last 40 years: while DO+ has been increasing in BrE in all registers, there is even a slight decrease in AmE in formal language.



**FIGURE 1** Do+ in BNC and COCA (pttw)

† The numbers after the corpus indicates the frequency of DO+ pttw across all categories



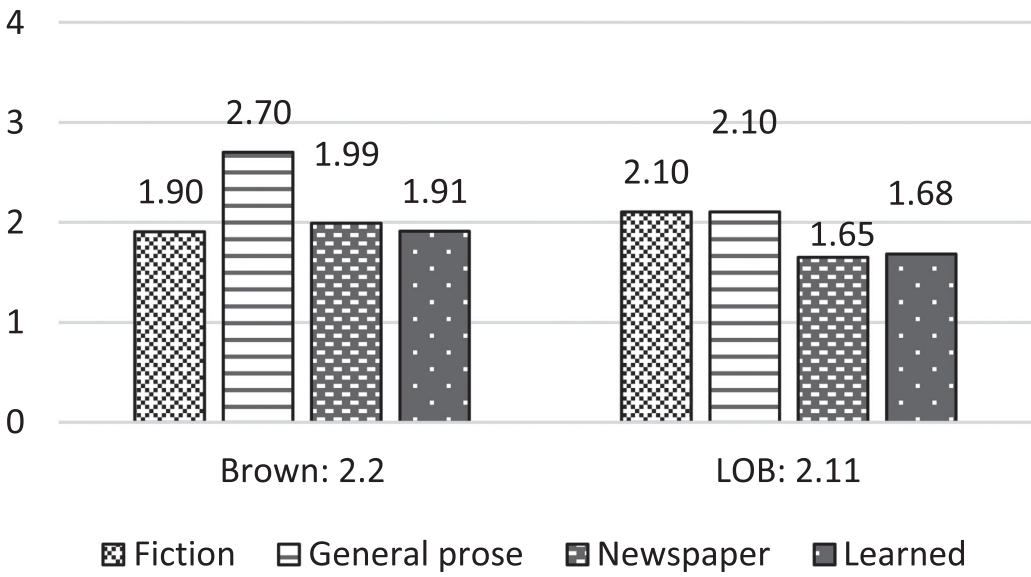
**FIGURE 2** Do+ in different categories of the written part of BNC and COCA (pttw)

### 3 | FUNCTIONAL RANGE OF DO+

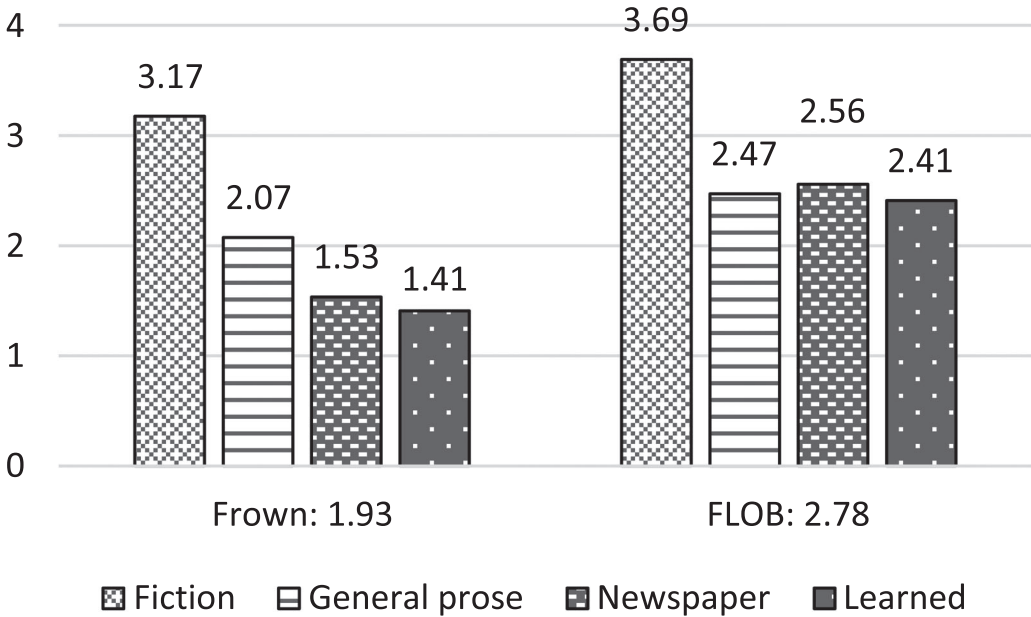
Quirk et al. (1985) distinguish between the 'emphatic positive', where DO is used 'to deny a negative which has been stated or implied', and 'emotive force', where no contrast is implied:

|                    |                                   |                     |
|--------------------|-----------------------------------|---------------------|
| Emphatic positive: | You should listen to your mother. |                     |
|                    | But I DÒ listen to her.           |                     |
| Emotive force:     | I DÒ wish you would lIsten.       | (1985, pp. 124–125) |

In cases of emotive force 'the speaker (in a style that is sometimes felt to be rather gushing and extravagant) is conveying enthusiasm', or some other emotion such as concern, sympathy or 'something like reproach or petulance'



**FIGURE 3** Do+ in Brown and LOB, 1960s (pttw)



**FIGURE 4** Do+ in Frown and FLOB, 1990s (pttw)

(Quirk et al., 1985, p. 1415). Though grammatically distinct, similar connotations are associated with the so-called emphatic or persuasive imperative.<sup>6</sup> I therefore include imperatives as in (1) in my conception of DO+ and thus also in the analysis of its use in different varieties of Englishes.<sup>7</sup>

- (1) B: Oh god oh my god oh my god. That got worse. I don't believe it. It was so embarrassing  
A: Oh **do** tell me (ICE-GB: S1A-091)

Though DO+ can express contrast in imperatives to refer back to a former non-compliance with a request, Davies (1986, p. 85) suggests that, in the majority of cases, DO+ emphasises 'the speaker's favourable attitude towards the possibility presented'. However, adding emotive force both to the imperative and to affirmative declaratives seems to be more frequent in fiction than in real conversations (Biber et al., 1999, pp. 434–435; Nevalainen & Rissanen, 1986, pp. 47–48). Accordingly, Nevalainen and Rissanen (1986, p. 48) suggest that 'emotive attitude and emphasis should perhaps be regarded as subordinate to more general aspects of cohesive discourse', which DO+ creates.

Regarding the emphatic positive, contrast can be created on various levels. The most detailed description can be found in Gerner (1996), but, for the purpose of this paper, I will only discuss the use of DO+ in concessions since it frequently occurs in this construction, and because it illustrates the difficulty of separating contrastive from emotive emphasis. In regard to discourse functions, concessions express, on the ideational level, a contrastive relationship between two propositions or their implications, expressed in a claim and a countermove, which often restricts the validity of the claim (Barth-Weingarten, 2003, pp. 163–170). On the interpersonal level, the acknowledgement of a proposition implied or expressed in the previous discourse is often used to counterweight the face-threatening potential of the countermove (Barth-Weingarten, 2003, pp. 172–181). In my data, the different moves are frequently signalled by concessive and adversative connectors, that is *although*, *though*, or *while* introducing the acknowledgement and *but*, *however*, or *nonetheless* (among others) introducing the countermove. Thus, contrast is expressed overtly, with or without DO+. Used in the countermove, as in (2), DO emphasises the truth value of the counterclaim in contrast to a previous claim, thus guiding the reader's interpretation (Aijmer & Simon-Vandenberg, 2011).

- (2) The meeting achieved little but it **did** contain two ingredients which were to dominate the politics of councils like Poplar. (ICE-GB: W2B-019)

If used in the acknowledgement, the speaker/writer does acknowledge the truth of some proposition but limits its scope in the countermove so that contrast is only created in retrospection as in (3) and (4).

- (3) It is true that the Franks **did** influence Brittany, but the consequent changes in Breton government [...] had little effect on the organisation of local communities and their self-regulation of their affairs. (ICE-GB: W1A-003)
- (4) At certain times during the period it **does** appear that Frankish control descended upon Brittany. However, this control was on the surface only and didn't exist in practical terms. (ICE-GB: W1A-003)

Thus, if the acknowledgement is not marked by a concessive connector, the use of DO prepares for a countermove which, however, may only follow in the next sentence as in (4). The frequent use of DO+ in these contexts might be the reason that even acknowledgements without following countermove retain an implicit reservation against the unrestrained applicability of the acknowledged proposition.<sup>8</sup> In the following conversation, no countermove follows, but the acknowledgment might still imply some disagreement.

- (5) A: Well haven't you noticed she's lost a little bit of weight  
 B: You mentioned  
 A: Well no but I mean can you not notice  
 B: Ya I ya you **do** notice (ICE-CAN: S1A-023)

In this way, the functions of rendering emotive emphasis (for example, sincerity, as in example (6)) and expressing contrast are not mutually exclusive:

- (6) Regarding the points made in respect of the policy benefits we **do** apologise for the confusion which arose in this instance, but would like to point out that your client's policy is tailor-made and therefore slightly different from the normal policy wording and conditions. (ICE-HK: W1B-023)

This double-edged function of DO+ might be the reason why Kleiner and Preston (1997) find an accumulation of DO+ in discourse disputes negotiating sensitive topics.

In addition to these emphatic functions, DO+ neither always bears nuclear stress nor is it always raised in pitch (Culicover, 2008; Gerner, 1996; Nevalainen & Rissanen, 1986). Nevalainen and Rissanen (1986) find a correlation between this unstressed DO+ and its function as a discourse-structuring device, for example for (re)introducing or ending a topic, elaborating on a topic, or summarising it. Furthermore, they observe that DO+ frequently either co-occurs with pragmatic particles or can be replaced by them, which underlines its function in discourse processing (Nevalainen & Rissanen, 1986, pp. 42–43). These uses separate DO+ from other contexts (negation, inversion, code) where its insertion is a purely grammatical operation – though even here it is perhaps not completely void of semantic meaning (Hirtle, 1997; Penhallurick, 1985). However, a clear-cut separation from emphatic uses is as problematic as trying to establish a categorical difference between contrastive and emotive emphasis. Nevalainen and Rissanen (1986, p. 38), though admitting that their categories of ‘explicit opposition’, ‘implicit opposition’, and ‘non-contrastive’ are non-discrete, report that 54 per cent of the instances of DO+ in spoken language do not imply ‘obvious opposition or contrast.’ The question is how ‘obvious’ is defined here, since Gerner (1996, pp. 150–152), in reassessing their data and using his own categorisation, reduces the non-contrastive functions of DO to 23 per cent. Thus, Bohnacker (2013, p. 175) claims that ‘the precise nature of “emotive force” [...] or discourse functions of *do* remains unclear from the discussions in the literature.’ This is perhaps less a reflection of a lack of description but of the context-dependent multifunctionality of DO+, a property shared with discourse markers for which ‘it often proves difficult in individual cases to tease apart the different uses, which may form a complex network of meanings’ (Brinton, 2017, p. 7).

Though DO+ assumes various discourse-pragmatic functions and adds procedural rather than propositional meaning, it is neither mobile (that is, able to occur in clause-initial, clause-medial, or clause-final position) nor does it ever form a separate tone unit. While DO+ is an integral part of the verb phrase, discourse markers are often characterised by their syntactic independence. Another formal criterion for identifying discourse markers is their potential for phonological reduction compared with the source form. If the source form is considered to be emphatic DO in the sense of the NICE properties, then the unstressed version might well fall into this category since Nevalainen and Rissanen (1986) have shown a clear correlation between unstressed DO and discourse-marking functions. This aligns with discourse markers often having ‘homophonous forms that function as standard parts of speech [...] from which they derive’ (Brinton, 2017, p. 7). In the case of DO+, the overlap of functions is perhaps the more prominent as the possibility of syntactic detachment is not given and the classification of its functions cannot be correlated with its position in the sentence.

As DO+ thus operates at the interface of grammar and discourse, it is an interesting construction to study in ‘non-native’ varieties of English since it can be assumed that the more salient grammatical functions are easier to master, while learners may struggle with structures that are not governed by grammatical principles (Müller, 2005, pp. 1–2). For the non-grammaticalised use of DO+ in L2 varieties, the learner is confronted with an additional obstacle, beyond its extreme versatility in function, which it shares with many discourse markers: the production of periphrastic DO in affirmative sentences where it does not serve a specialised emphatic or discourse-pragmatic function has been reported to occur in the acquisition process as a result of overgeneralisation from other contexts (Bohnacker, 2013; Schütze, 2013), which is partly triggered by priming, that is the repetition of a structure available from the previous discourse. While the omission of DO+ in suitable contexts (that is, where it does serve a discourse-pragmatic function) will probably not be noticed, the production of superfluous DO may be subject to overt comment. The internalisation of such a rule might result in the further avoidance of the construction even at an advanced level of proficiency.

## 4 | METHODOLOGY

The database for the general assessment of the frequency and distribution of DO+ across different varieties and genres consists of the ICE components in Table 3. The frequency of DO+ is assessed by counting all occurrences of *do*,

**TABLE 3** ICE components serving as database for the present study

| Inner Circle     | Asian varieties |
|------------------|-----------------|
| ICE-GB           | ICE-HK          |
| ICE-US (written) | ICE-SIN         |
| ICE-NZ           | ICE-IND         |
| ICE-CAN          | ICE-PHI         |

does, and *did* followed by a lexical verb (including lexical *do* and *have*) in the infinitive, allowing for (mostly adverbial) insertions in between, such as *in fact*, *indeed*, *however*, and so on. Any cases of repetition (*I do think... do think... do think that*), and unclear or defective constructions (as in (7) and (8)) are excluded:

- (7) Uhm **do** have you got a <,> dollar figure as to how much revenue you might have been lost as a result of all this piracy (ICE-HK:S1B-029)
- (8) So <,> it depend upon that <,> uh their environment <,> And even home ↔ **did** </> condition also <,> If they be in always... (ICE-IND:S1A-087)

I use normalised frequencies (pttw) in this paper, even though there is a substantial limitation to this approach: it is not suitable for a direct comparison between different genres since the frequency of DO+ is dependent on the number of finite verb forms. These are more frequent in informal and spoken language, while the more complex sentence structure of formal written language has considerably fewer finite verb forms per number of words. Nevalainen and Rissanen (1986, p. 49, note 6) mention that if the frequency is calculated as ratio of actual occurrences per potential contexts (that is DO+ per simple finite verb form), it is actually academic writing which shows the highest frequency of DO+. While normalised frequencies are thus not the best method for comparisons across genres, it is yet a suitable method to compare frequency across different varieties due to the fixed number of words per sub-category in the ICE components. Furthermore, it also allows for a comparison with other approaches (see Section 2), which use normalised frequencies.

Besides the overall frequency of DO+ within the different ICE components, its occurrence in the spoken and written parts as well as in different text types is assessed separately in order to document differences in distribution. Besides a broad distinction between the spoken and written parts, I also calculated the frequency for the sub-categories to see whether the overall frequency is overly influenced by single text types. However, as DO+ is a rather infrequent phenomenon, more fine-grained differences between varieties cannot be verified in this way as the total number of words in the sub-categories is too low to provide reliable results. The analysis for the spoken part is therefore restricted to the broader categories 'private dialogues' (S1A), 'public dialogues' (S2A), 'scripted monologues' (S2A), and 'unscripted monologues' (S2B). For the written part this caveat is even more severe: first, it is per se smaller than the spoken part (600,000 vs 400,000 words). Second, it is subdivided into more sub-categories, which cannot easily be grouped together in a meaningful way. For example, letters and student writing both fall into the category 'non-printed', but merging them into one category would veil the marked differences between these text types in, for instance ICE-NZ (Figure 8), ICE-IND, and ICE-SIN (Figure 9). Lastly, the frequency of DO+ is generally lower in written language, so that the absolute frequency for single text types is considerably lower than in the categories of the spoken part. To indicate where numbers are rather low, I included the absolute frequency ( $n = X$ ) in the figures in Section 5.2, even though these cannot be used to make comparisons between the categories within one variety as the total number of words for the sub-categories varies. Despite these difficulties, the results can still reveal text types with a particularly high or low frequency of DO+, which are used as a starting point for the qualitative analysis in Section 5.3.

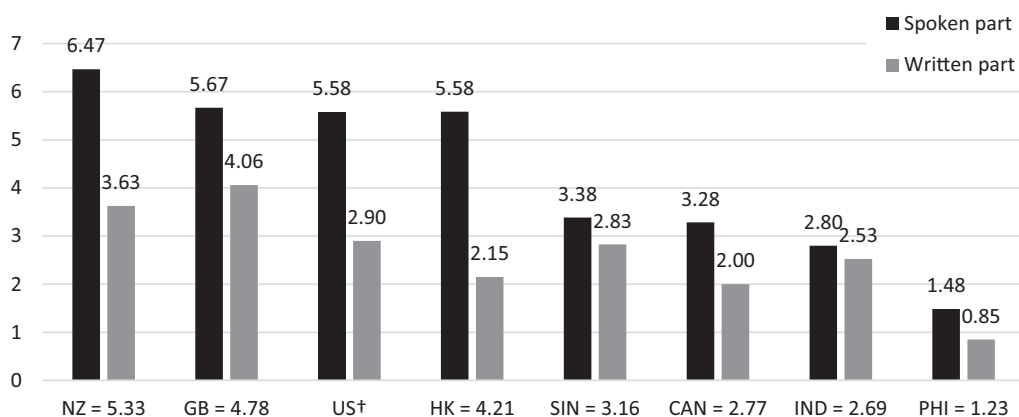


## 5 | ANALYSIS

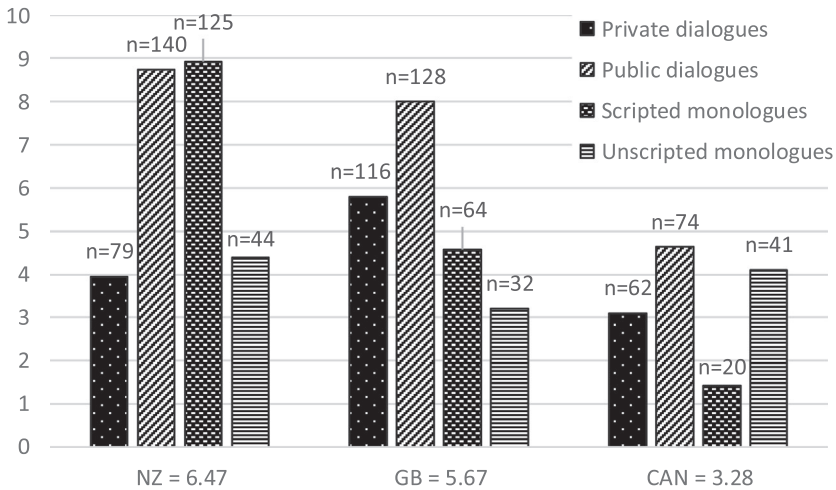
The analysis starts with the general distribution of DO+ across the spoken and written parts of the eight selected ICE components (Section 5.1), followed by a more detailed assessment of the occurrence in different genres (Section 5.2), which might influence the overall rate in the written and spoken parts. Finally, the use of DO+ in different functions in specific contexts will be taken into account focusing on a selection of observations (Section 5.3). Thus, the analysis overall might serve as starting point for further research in single varieties with a broader database than ICE can provide.

### 5.1 | Bird's eye perspective

The overall distribution of DO+ in spoken and written language across different varieties is displayed in Figure 5. The varieties are arranged according to the overall frequency of DO+ in the entire corpus, from ICE-NZ on the left with the highest frequency (5.33 pttw) of DO+ to ICE-PHI (1.23 pttw) to the right, with the lowest frequency both in the written and the spoken part. As ICE-US only contains written genres, the frequency for the spoken part is based on COCA (5.09 pttw) to provide a point of comparison (without claims about the general comparability of the corpora). It becomes obvious that even though two Inner Circle varieties, that is New Zealand English (NZE) and BrE, rank first, it is not generally true that DO+ is less frequent in Outer Circle varieties. The rates for AmE and Hong Kong English (HKE) are almost equal, and Canadian English (CanE) shows a lower rate than HKE and Singaporean English (SgE). While it is obvious that DO+ is more frequent in spoken than in written language in all varieties, the difference in SgE and Indian English (IndE) is rather small. Another conspicuous result is the high frequency of DO+ in the spoken part of ICE-HK. Furthermore, the frequency of DO+ in the written part of both ICE-GB and ICE-US is rather high (4.06 and 2.90 respectively) compared to both the results from the BNC (2.66) and COCA (2.03) (see Figure 1) and those from FLOB (2.78) and Frown (1.93) (see Figure 4). However, Figure 4 also shows that the rates for DO+ are highly genre dependent, which might suggest that single genres (over)represented in ICE are responsible for the higher overall frequency.



**FIGURE 5** Frequency of DO+ (pttw) in spoken and written parts of selected ICE components  
 † US spoken based on COCA (1990-1994)

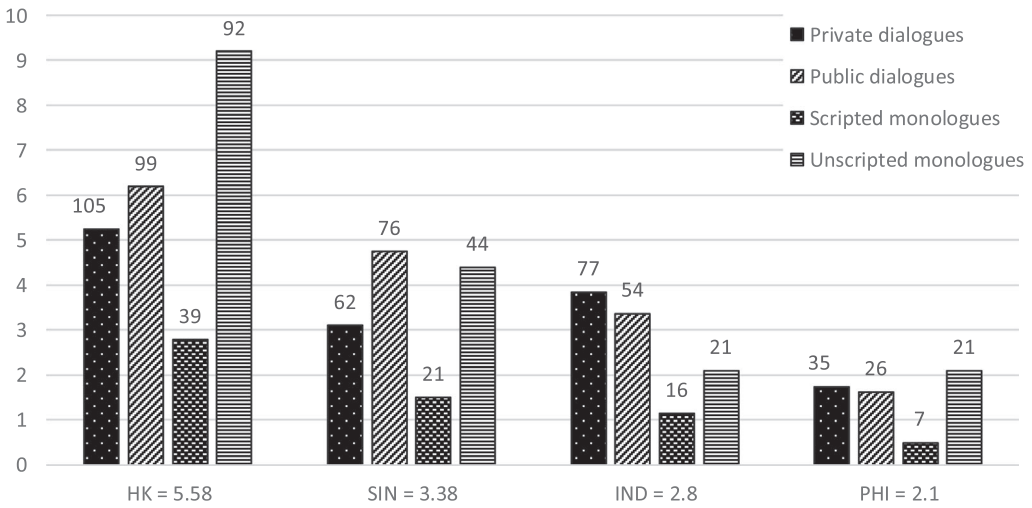


**FIGURE 6** Frequency of DO+ (pttw) in the spoken part of ICE: Inner Circle varieties

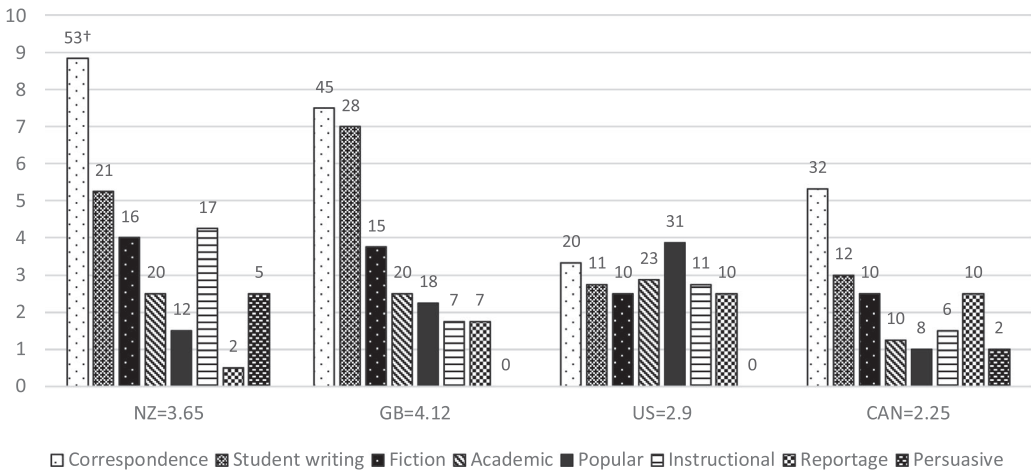
## 5.2 | Spoken vs written text types

Figure 6, displaying the results for the selected Inner Circle varieties except ICE-US, shows that the general ranking according to frequency for spoken language, that is NZ > GB > CAN, is not retained in all sub-categories, for example, in private dialogues, DO+ is most frequent in ICE-GB, and it is more frequent in ICE-CAN in unscripted monologues than in ICE-GB. A general tendency seems to be that DO+ is more frequent in public than in private dialogues. The results for scripted vs unscripted dialogues, however, are more diverse across the different varieties. Generally, the highest frequency for DO+ is found in public dialogues, which is only equalled in scripted monologues in ICE-NZ. This category also shows the most striking differences between the different varieties: in ICE-NZ the absolute frequency of DO+ ( $n = 125$ ) is twice as high as in ICE-GB ( $n = 64$ ) and even six times higher than in ICE-CAN ( $n = 20$ ). This suggests the avoidance of DO+ in CanE in a category which contains language produced in more formal contexts, while it is especially frequent here in NZE.

Regarding Asian Englishes (Figure 7), a general tendency appears to be the avoidance of DO+ in scripted monologues. This is consistent in all four Asian varieties, where the frequency of DO+ in this category rather patterns with the rates for written language, which corresponds to the distribution found in CanE. Another conspicuous result is the disproportionately high frequency of DO+ in unscripted monologues in ICE-HK, which even surpasses the rates found in the Inner Circle varieties. Figure 8 (Inner Circle) and Figure 9 (Outer Circle) show the frequency of DO+ in different text types in the written part of the selected ICE components. For the Inner Circle, the high frequency of DO+ in student writing in ICE-NZ with 13.25 is certainly the most striking result, but also the high rates in correspondence and student writing in ICE-GB stick out (ca. 7) compared with the general frequency of 4.06 for the written part of ICE-GB. Another noteworthy result is the marked difference between genres in ICE-NZ compared with the almost even distribution in ICE-US. With reference to the Outer Circle, the total frequency of DO+ is so low in the written part that hardly any conclusions can be drawn from its distribution in different text types. However, it is noteworthy that the highest rate can be found in the category 'correspondence' in all selected varieties. Furthermore, ICE-SIN and ICE-IND stick out with a particularly high frequency of DO+ in this category, equalling the rates found in ICE-NZ. Thus, the relatively high frequency of DO+ in the written part of ICE-IND and ICE-SIN is mostly due to its occurrence in letters.

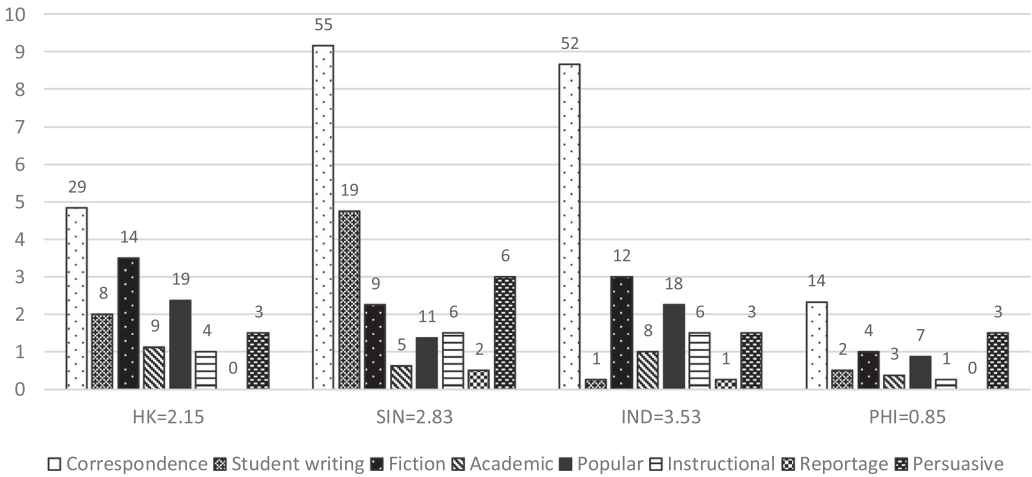


**FIGURE 7** Frequency of DO+ (ptw) in the spoken part of ICE: Asian Englishes



**FIGURE 8** Frequency of DO+ (ptw) in different text types in the written part of ICE: Inner Circle  
 † Numbers above the bars indicate the total number of hits (n)

Furthermore, the high rates for DO+ in student writing in ICE-NZ and ICE-GB, and to some extent in ICE-CAN, is only mirrored in ICE-SIN, while it is virtually absent in ICE-IND and ICE-PHI. However, it has to be mentioned here that DO+ is highly context-sensitive: as a device frequently used to juxtapose different claims, it is of course more frequent in argumentative essays. Thus, the frequency in student writing is to some extent task-related. Philippine English (Phile) again shows the lowest frequency of DO+ across all text types, but there is a marked avoidance of DO+ in the more formal genres in all Asian varieties, which does not correspond to the distribution found in the Inner Circle varieties. In fiction, however, there is no systematic difference between the Inner and Outer Circle.

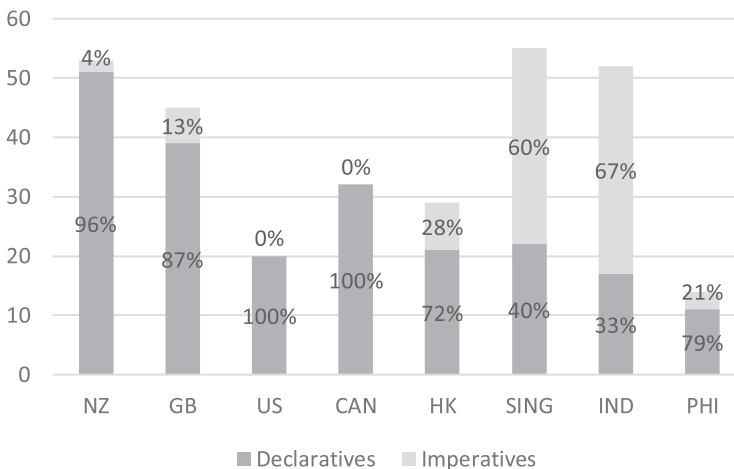


**FIGURE 9** Frequency of DO+ (pttw) in different text types in the written part of ICE: Outer Circle

### 5.3 | Focus on functions

There are three aspects I will focus on in this section: the high frequency of DO+ in the correspondence section in ICE-IND and ICE-SIN, its high frequency in unscripted monologues in ICE-HK, and potential learner effects in the usage of DO+.

**Correspondence:** Figure 9 shows that the high frequency of DO+ in the written part of ICE-IND and ICE-SIN is mostly due to its occurrence in letters, which suggests a special function not found in the other text types. Figure 10 separates uses of DO+ in affirmative declaratives and in the imperative, showing that the emphatic imperative is characteristic of SgE and IndE. It is noteworthy that this function is completely absent in the North American varieties and that, though used in NZE and BrE, it is clearly a feature of the Asian varieties not paralleled in any of the Inner Circle varieties. The examples found in ICE-SIN and ICE-IND reveal two different functions of the emphatic imperative: it is used to indicate emotional involvement, as in (9) and as a device to reinforce a request, as in (10).



**FIGURE 10** DO+ in declaratives and imperatives in 'correspondence' (total number of hits)

- (9) Do enjoy your stay there. (ICE-SIN:W1B-014)  
 (10) Please **do** suggest some topics in your field for the next conference. (ICE-IND:W1B-026)

The second use occurs more frequently in ICE-IND, though the overall results are too sparse to allow for generalisations.

**Unscripted monologues in ICE-HK:** The highest frequency of DO+ in ICE-HK is found in unscripted monologues, which mostly consists of sports reportage, which means, in the case of HK, horse-racing (ICE-HK s2a 001–018). The examples suggest a special function of DO+ which renders the reported events more lively:

- (11) [...] and he raced nicely clear Little Big John <,> getting into second Panda Power and that **did** jump the favourite quinella one and six <,> returning around seventy dollars ... (ICE-HK:S2A-004)  
 (12) [...] and on the outside of Beautiful Dancer here is Best Tiger now as they turn for home <,> you can see he **does** uh roll off half a horse here uh Beautiful Dancer Peter Lau ... (ICE-HK:S2A-012)

However, for this section the influence of compilation practices cannot be ruled out due to the inclusion of 'native' speakers from Australia.<sup>9</sup> But even if this is the case, it cannot explain the generally high frequency of DO+ throughout ICE-HK. A qualitative analysis of the occurrences did not indicate that any of the uses are 'non-canonical', which is not true for all Asian varieties.

**Learner effects:** There are single occurrences of DO+ which cannot be accounted for by the range of functions described in Section 3. These instances suggest mechanisms in 'non-native' language use which may lead to an increased frequency of periphrastic DO (that is, unstressed DO not connected to either emphasis or discourse-pragmatic functions). Example (13) shows an overuse of DO, possibly triggered by priming:

- (13) B: Shivaji is not exactly from Poona but we do have some uh historical events of those days it's uh recorded some where  
 C: Like like [...]  
 B: We **do** have like that  
 D: I don't know history very well  
 B: We **do** have [laughs]  
 A: We **do** have <O> one word </O> which is known for Peshwas  
 B: We **do** have a Peshwas yeah we **do** have Shanivarwada which is also known for Peshwas  
 D: Peshwas  
 B: And uh so many for uh recently we **do** have one snake project [...]  
 B: Yeah we **do** have observatory National Chemical Laboratory which is famous and even we **do** have a museum Mahatma Phule Museum (ICE-IND:S1A-008)

In the first turn, speaker B uses DO+ to express contrast, made explicit by the preceding *but*. However, most of the later turns lack this function. The use of periphrastic DO can also be found in learner English, for example, when it is included in an answer to a question formed with DO. On the other hand, the combination of DO and *have* is also very frequent in 'native' varieties (Biber et al., 1999, p. 434). Though HAVE as lexical verb does not frequently pattern with DO+ (compared with other verbs),<sup>10</sup> it is itself a high-frequency verb so that from the total occurrences of DO+ the percentage of *DO have* is rather high in all selected ICE components (Table 4). It is, unsurprisingly, more frequent in spoken discourse. In written discourse, its frequency is higher in the Inner Circle varieties than in Asian Englishes though the difference to BrE is not that prominent. I did not find clear examples of DO being used as a tense marker (Kortmann, 2004). Yet, awareness of oversupplied DO in the interlanguage (Selinker, 1972) may lead to its avoidance which reinforces insecurity about its more subtle functions:

**TABLE 4** Frequency of *DO have* (in per cent)

|     | Spoken | Written |
|-----|--------|---------|
| NZ  | 17.0%  | 12.3%   |
| GB  | 14.1%  | 7.9%    |
| US  | n/a    | 15.5%   |
| CAN | 26.4%  | 10.0%   |
| HK  | 15.8%  | 4.7%    |
| SIN | 12.8%  | 6.2%    |
| IND | 13.7%  | 4.0%    |
| PHI | 15.7%  | 5.9%    |

(14) A: You seem to be versatile man with varied interests

B: Thank you

A: I couldn't have guessed as much from

B: I **did** try it <,> Tried it

A: Tried it of course (ICE-IND:S1A-026)

In the context of the conversation, speaker B's use of *DO+* can be interpreted as contrastive to the former statement that he did not succeed. The self-correction suggests that the speaker is insecure about this use. Whether the repetition of it by speaker B is an affirmation of the assumed correct linguistic form or an instance of back-channelling has to remain speculation.

## 6 | DISCUSSION

The analysis of the distribution of *DO+* in various Asian Englishes compared to Inner Circle varieties has revealed various points of interest that invite further investigation. This exploratory paper has focused on frequency to test whether *DO+* is used less in Outer Circle varieties and to establish differences in frequency across different genres. This might indicate differences in function since it has been shown that contrastive functions, which are in line with the use of *DO* as emphatic operator, are prevalent in written discourse while non-contrastive uses are more frequent in spoken discourse (Nevalainen & Rissanen, 1986). The assumption that the grammatical functions are easier to master is to some extent supported by the fact that the gap between Inner and Outer Circle varieties is lower in written language than in spoken language (see Figure 5). However, it was also shown that it is not generally true that *DO+* is used less in the Asian varieties as HKE seems to pattern with BrE, outpacing CanE in the use of *DO+*.

The generally lower frequency in AmE is certainly one of the factors which is responsible for the low frequency in PhilE. However, the almost categorical avoidance of *DO+*, especially in formal language, suggests that PhilE is already taking its own path in this respect. Still, the association of *DO+* with informal language might be the initial impetus for this development. Studies of other phenomena in PhilE have shown that speakers generally tend to choose more formal variants (see for example Dayag, 2016, on preposition stranding vs pied-piping). Thus the avoidance of *DO+* in written as well as spoken discourse might be another instantiation of 'stylistic underdifferentiation' (Gonzalez, 2012, p. 334) in PhilE. While AmE and CanE also seem to be more conservative in the use of *DO+* in written genres (in contrast to BrE, where a spread into all written registers is visible (see Section 2)), NZE seems to go into the opposite direction, which is in line with observations that more colloquial forms prevail in NZE (Peters, 2009, p. 397).

Another interesting finding is the use of the emphatic imperative in ICE-IND and ICE-SIN, which exceeds by far the use in any of the 'native' varieties. Biber et al.'s (1999, p. 435) observation that this form is rather used in fiction could be an indication of its stylised character that is seldom found in actual language usage. As this use is, however, avail-

able in 'native' English varieties, especially in BrE, it could have been extended in IndE and SgE. However, diachronic developments in BrE might play a role here, too.<sup>11</sup> The emphatic use has supposedly been available as early as the 15th century (Denison, 1993, pp. 266–267) and was prevalent in the 18th (Nurmi, 2018, p. 119). The lexical associations of the construction attested for the late 17th and 18th centuries, most notably the types of verbs (Nurmi, 2018, p. 131) and adverbs (Bækken, 2002, p. 328), indicate a drift towards a specialised function associated with emphasis, for instance as politeness marker or to emphasise stance or epistemicity. This use of adding emotive force is exemplified in the following examples from The Corpus of Early English Correspondence (CEEC). There is no indication that DO+ was contrastive then, indicated by the lack of adversative connectors co-occurring with it,<sup>12</sup> while it is frequently used with intensifiers.

(15) Which being soe, ..., I **doe** wholly concurr with you that ... (pepys\_059; 1679)

(16) And I **do** earnestly intreate you to move hir Majesty ... (stuart\_058; 1610)

(17) ... and I **doe** heartily pray God to blesse you in your proceedings... (petty\_035; 1679)

That this use of DO+ was more frequent in earlier stages of English might be a factor contributing to its use in post-colonial Englishes indicating the retention of an earlier function ('extraterritorial conservatism' or 'colonial lag'). Thus, it is worth analysing whether non-contrastive emphatic uses (which Nevalainen & Rissanen, 1986, pp. 47–48, find to be a feature of fiction) are also more frequently used in SgE and IndE than in other varieties.

That contrastive emphasis was not a prevalent function in the 17th century implies a functional change during the 18th and 19th centuries, which would be in line with the tendency that emphatic elements often acquire adversative meaning (Traugott, 2012). A further development from a contrastive meaning to discourse-marking functions is suggested by the fact that some underlying contrast is still implied though not necessarily made explicit (Nurmi, 2018, p. 123). This is also paralleled by the development of adversative adverbs and connectors (Traugott, 2012, p. 555). This would imply an overall development from emotive emphatic to contrastive emphatic to discursive. While all functions are still preserved in PDE, their diachronic development in terms of frequency as well as their distribution across different genres certainly affects the use of DO+ in postcolonial Englishes. Besides the influence of the superstrate, influence can also be expected from learner effects, which seem to shine through in my data in single examples. However, whether the use of DO+ is indeed influenced by an awareness of its oversupply in learner language cannot be established by a corpus-linguistic approach. The general tendency that DO+ is less frequent in Outer Circle varieties might also be due to the difficulty to master its versatile functions. For other discourse markers a tendency to overuse specific functions has been reported (Cheng & Warren, 2001). To verify this for DO+, a more fine-grained categorisation of its functions based on objective criteria is necessary, which is still pending even for BrE and AmE.

## 7 | CONCLUSION

In this paper, I explored the use of DO+ in four Asian varieties compared to four Inner Circle varieties of English in terms of frequency and distribution across different text types. The underlying hypothesis was that DO+ is less frequent in postcolonial Englishes due to its versatile discourse-pragmatic functions and the use of periphrastic DO as result of overgeneralisation or priming. Though there is a general tendency for its less frequent use in the Asian varieties, this cannot be generalised since HKE shows a higher frequency of use than CanE. A more detailed assessment of the distribution of DO+ in specific varieties would certainly require a broader database than was used here. This is not only because DO+ is a relatively infrequent phenomenon, but also because its use is also highly text type dependent. Thus, differences in genres compiled for the different ICE sections may have a greater effect on the results than variety-specific differences. In addition, work with authentic spoken data is necessary to include the prosodic properties of DO+ into the assessment of its functions.

Yet, this exploratory paper can certainly be used as a starting point for further research, not only on single varieties, but also on the variable studied in the larger context of discourse markers. The differences in distribution across Inner

Circle varieties seem to be connected to a recent spread of DO+ into more formal genres, which is also reflected in the Asian varieties. Another important influence seems to be the historical development of DO+ from emphatic to discursive uses. As these functions are distributed unevenly over different text types, more systematic approaches to DO+ are necessary to gain a more accurate insight into the influence of the superstrate on new varieties of English. Such studies would contribute further to our understanding of the development of discourse-marking functions, not only in postcolonial Englishes but also in the history of English: DO+ has been shown to take a linking position between grammatical and discursive functions, so as soon as it ceases to be treated as a primarily grammatical phenomenon, it might become an interesting subject of study for the interrelatedness of grammar and discourse.

## ACKNOWLEDGEMENT

Open access funding enabled and organized by Projekt DEAL.

## NOTES

- <sup>1</sup> At least if a positive-negative opposition is supposed to be expressed. Nuclear stress on the main verb rather implies a choice among other verbs (Bohnacker, 2013, p. 43, note 4).
- <sup>2</sup> While Nevalainen and Rissanen (1986) calculate the frequency per 1,000 words, Biber et al. (1999) calculate it per million words. For the sake of comparability, I have converted both into hits per ten thousand words, a measure I am using throughout this paper. Furthermore, I used only parts of the results from LOB and LLC to directly compare them to the categories 'conversation', 'fiction writing', 'news writing', and 'academic prose' used in Biber et al. (1999).
- <sup>3</sup> The corpora are designed to allow for comparisons between AmE and BrE. The Brown (AmE) and LOB (BrE) corpora contain data from the 1960s, while the data in Frown (AmE) and FLOB (BrE) were collected in the 1990s.
- <sup>4</sup> More-word insertions, as in *He did in fact call*, are not covered. The reason for this is that the most frequent insertions are single adverbs (while insertions are in general not frequent in this position in PDE). The inclusion of more complex search terms, on the other hand, would further increase the number of false hits. The emphatic imperative is included in the results (see Section 3).
- <sup>5</sup> The only exception is the category 'general prose' in Brown. This is mostly due to the high frequency of DO+ in section G containing belles lettres, biographies, memoirs, and so on.
- <sup>6</sup> For example, in the imperative, DO can be used with *be*, as in *Do be careful*, which is not possible (in 'standardised' BrE or AmE) in declaratives (Quirk et al., 1985, p. 1415, note a). The label 'emphatic' or 'persuasive' already implies that these imperatives constitute emotive appeals, urgent requests, or entreaties rather than commands (Davies, 1986, p. 76).
- <sup>7</sup> Bold print was added by the author of this paper to highlight the use of DO+ in the examples.
- <sup>8</sup> Ranger (2015, p. 117) claims that in these cases 'the speaker's subjective endorsement is made explicit' and that this subjective position stands 'in contrast with other potentially opposing perspectives.' He thus suggests that even cases of emotive emphasis are underlyingly contrastive, though he admits that further studies based on authentic oral corpora are needed.
- <sup>9</sup> Though ICE Australia was not included into the analysis, New Zealand English, which shows various similarities to Australian English, especially regarding colloquialism (Peters, 2009, p. 397), shows the highest rates for DO+ in all text types.
- <sup>10</sup> Biber et al. (1999, p. 433) claim that '[t]here are a few verbs that typically occur with emphatic *do*, and none of these is particularly common', though they do not indicate the verbs in question. My own analysis of the BNC revealed that *apologise* and *exist* score highest in this respect (Schulz, 2017).
- <sup>11</sup> Though there are various accounts on the origin and emergence of DO in affirmative contexts (Denison, 1993), scholars have focused on a grammatical assessment of the construction rather than on its functions. There are attempts to describe the functions of DO+ in Early Modern English (Bækken, 1999). However, since its use cannot be pinned down to a core function, scholars have taken retreat to the claim that it is a semantically empty alternative to the simple verb form (see Klemola, 1998, p. 26; an opinion, which is, for example, still reflected in Schütze, 2013, p. 20).
- <sup>12</sup> Even in the last decades covered by the CEEC (1631–1680) only 12% of all instances of DO+ can be considered contrastive, while the percentage lies at 54% in the correspondence section of ICE-GB and 57% for ICE-US (Schulz, 2018).

## REFERENCES

- Aijmer, K., & Simon-Vandenberg, A. M. (2011). Pragmatic markers. In J. Zienkowski, J.-O. Östman, & J. Verschueren (Eds.), *Discursive pragmatics* (pp. 223–308). Amsterdam: John Benjamins.
- Bækken, B. (1999). Periphrastic *do* in Early Modern English. *Folia Linguistica Historica*, XX, 107–128.
- Bækken, B. (2002). *Yet this follie doth many times assault the brauest minds*: Affirmative declarative *do* in 17th-century English. *Nordic Journal of English Studies*, 1, 317–337.
- Barth-Weingarten, D. (2003). *Concession in spoken English: On the realisation of a discourse-pragmatic relation*. Tübingen: Narr.



- Biber, D., & Gray, B. (2016). *Grammatical complexity in academic English: Linguistic change in writing*. Cambridge: Cambridge University Press.
- Biber, D., Johansson, S., Leech, G., Conrad, S., & Finegan, E. (1999). *Longman grammar of spoken and written English*. Harlow: Pearson Education.
- Bohnacker, U. (2013). Reflections on dummy 'do' in child language and syntactic theory. In C. D. de van Craats-Oosterwold, J. Verhagen, & W. B. T. Blom (Eds.), *Dummy auxiliaries in first and second language acquisition* (pp. 171–208). Berlin: Mouton de Gruyter.
- Brinton, L. J. (2017). *The evolution of pragmatic markers in English: Pathways of change*. Cambridge: Cambridge University Press.
- Cheng, W., & Warren, M. (2001). The functions of actually in a corpus of intercultural conversations. *International Journal of Corpus Linguistics*, 6, 257–280.
- Culicover, P. W. (2008). The rise and fall of constructions and the history of English *do*-support. *Journal of Germanic Linguistics*, 20, 1–52.
- Davies, E. (1986). *The English imperative. The Croom Helm linguistics series*. London: Croom Helm.
- Dayag, D. T. (2016). Preposition stranding and pied-piping in Philippine English: A corpus-based study. In *Communicating with Asia: The future of English as a global language* (pp. 102–119). Cambridge: Cambridge University Press.
- Denison, D. (1993). *English historical syntax: Verbal constructions*. London: Longman.
- Gerner, J. (1996). *Untersuchungen zur Funktion des emphatischen "do" im Englischen*. Frankfurt am Main: Peter Lang.
- Gonzalez, A. (2012). Stylistic shifts in the English of the Philippine print media. In J. Cheshire (Ed.), *English around the world* (pp. 333–363). Cambridge: Cambridge University Press.
- Hirtle, W. (1997). *DO auxiliary – a meaningful support and operator. Lingua*, 100, 111–149.
- Huddleston, R., & Pullum, G. (2002). *The Cambridge grammar of the English language*. Cambridge: Cambridge University Press.
- Hundt, M. (1998). *New Zealand English grammar, fact or fiction? A corpus-based study in morphosyntactic variation*. Amsterdam: John Benjamins.
- Kleiner, B., & Preston, D. R. (1997). Discourse disputes: How come you do do like you do. *Folia Linguistica*, 31, 105–131.
- Klemola, J. (1998). Semantics of DO in southwestern dialects of English English. In I. Tiekens-Boon van Ostade (Ed.), *Do in English, Dutch and German: History and present-day variation* (pp. 25–51). Amsterdam: Stichting Neerlandistiek VU.
- Kortmann, B. (2004). Do as a tense and aspect marker in varieties of English. In B. Kortmann (Ed.), *Dialectology meets typology: Dialect grammar from a cross-linguistic perspective* (pp. 245–275). Berlin: Mouton de Gruyter.
- Müller, S. (2005). *Discourse markers in native and non-native English discourse*. Amsterdam: John Benjamins.
- Nevalainen, T., & Rissanen, M. (1986). Do you support the do-support? Emphatic and non-emphatic DO in affirmative statements in present-day spoken English. In S. Jacobson (Ed.), *Papers from the third Scandinavian Symposium on Syntactic Variation*. Stockholm, May 11–12, 1985 (pp. 35–50). Stockholm: Almqvist & Wiksell.
- Nurmi, A. (2018). Periphrastic DO in eighteenth-century correspondence: Emphasis on no social variation. In T. Nevalainen, M. Palander-Collin, & T. Säily (Eds.), *Patterns of change in 18th-century English: A sociolinguistic approach* (pp. 117–135). Amsterdam: John Benjamins.
- Penhallurick, J. M. (1985). The semantics of auxiliary DO. *Studies in Language*, 9, 311–333.
- Peters, P. (2009). Epilogue. In A. Smith, P. Peters, & P. Collins (Eds.), *Comparative studies in Australian and New Zealand English grammar and beyond* (pp. 387–399). Amsterdam: John Benjamins.
- Quirk, R., Greenbaum, S., Leech, G., & Svartvik, J. (1985). *A comprehensive grammar of the English language*. London: Longman.
- Ranger, G. (2015). Parameters for unexpected (and expected) meanings: Auxiliary *do* in affirmative contexts. *Canadian Journal of Linguistics/Revue Canadienne De Linguistique*, 60, 107–120.
- Schulz, N. (2017). Aspectual functions of auxiliary DO in world Englishes. Unpublished conference paper presented at the ISLE Summer School, *Variation in world Englishes: Advanced issues in theory and methodology*, University of Regensburg, 4–7 October 2017.
- Schulz, N. (2018). The functional development of auxiliary *do* in affirmative sentences in Early Modern English. Unpublished conference paper presented at the 20th International Conference of English Historical Linguistics (ICEHL 20), University of Edinburgh, 26–31 August 2018.
- Schütze, C. (2013). Superfluous 'do' and comparison of spell-outs. In C. D. de van Craats-Oosterwold, J. Verhagen, & W. B. T. Blom (Eds.), *Dummy auxiliaries in first and second language acquisition* (pp. 11–38). Berlin: Mouton de Gruyter.
- Selinker, L. (1972). Interlanguage. *IRAL-International Review of Applied Linguistics*, 10, 209–231.
- Traugott, E. (2012). Pragmatics and language change. In K. Allan & K. M. Jaszczolt (Eds.), *The Cambridge handbook of pragmatics* (pp. 549–565). Cambridge: Cambridge University Press.

**How to cite this article:** Schulz N. Auxiliary DO in Asian Englishes. *World Englishes*. 2021;40:502–518.

<https://doi.org/10.1111/weng.12537>