



# *Eh* Across Englishes: A Corpus-Pragmatic Analysis of the Corpus of Global Web-Based English

Michael Westphal<sup>1</sup>

Received: 4 August 2023 / Accepted: 29 October 2023 / Published online: 27 November 2023  
© The Author(s) 2023

## Abstract

This paper presents an analysis of the pragmatic marker *eh*, which is typical of spoken discourse, in written online discourse from nine varieties of English using the Corpus of Global Web-based English. The analysis focuses on sentence-final *eh* and considers variation in terms of variety, punctuation, text type, and function. This paper also includes a variationist analysis of *eh* in contrast to *huh*. Although there are cross-variety differences, *eh* is used across all nine varieties in similar ways. *Eh* is mostly combined with a question mark, it is more frequent in blogs than in general websites, and emphatic functions dominate over narrative and interrogative uses. A qualitative analysis of the indexicalities demonstrates that *eh* mainly signals orality and informality in online writing but also has specific local meanings. The variationist analysis shows that *eh* is preferred over *huh* in the Canadian and New Zealand components. This preference is even more pronounced for the British and Philippine components. In contrast, *huh* dominates in the US component. These results show that *eh* is well integrated into online writing and can be characterized as a translocal pragmatic marker as it is used globally but has developed local characteristics.

**Keywords** Pragmatic markers · GloWbE · World Englishes · Translocality

## Introduction

Pragmatic markers, such as *like*, *innit*, or *eh*, are integral components of spoken discourse as they fulfill a wide range of conversational functions. Speakers use them to express stance, encode politeness, structure conversations, and guide utterance interpretation (Aijmer, 2013; Beeching, 2016). However, pragmatic markers are difficult to define in terms of formal criteria because they are syntactically optional, are only loosely linked to the syntax of an utterance, are often short and phonologically reduced, and have little to no propositional meaning (Brinton, 2017).

---

✉ Michael Westphal  
westphal@anglistik.uni-kiel.de

<sup>1</sup> English Department, Kiel University, Kiel, Schleswig-Holstein, Germany

Pragmatic markers are also used in written discourse, but Brinton (2017) argues that alternative sets of markers are employed, such as structuring devices, like *however*, and even if spoken pragmatic markers are used, their function likely differs, and they are commonly evaluated negatively. The advent of computer mediated communication (CMC) has led to the emergence of new written genres that approximate spoken discourse and incorporate pragmatic markers more readily: Emoticons are typical pragmatic markers of CMC (Tagg, 2012), but traditional spoken pragmatic markers are also employed in interactive CMC, such as texting (Tagg, 2012), and in asynchronous genres, like blogs (Millar, 2015). Spoken pragmatic markers remain marked features across most (offline and online) written genres, and they are often employed to index orality and informality (McCarthy, 2013) or for metaphorical purposes, such as signaling humor. Tagg (2012: 122) therefore describes their use in texting as “part of a deliberate performance of a particular persona”.

The heightened indexical loading of many pragmatic markers enables speakers and writers to employ them for identity work. In the English language, many pragmatic markers are typical for specific varieties of English, like *innit* for British English (Beeching, 2016), while others are used globally, such as *OK*. The pragmatic marker *eh* is both local and global. Denis (2013: 1) describes *eh* as “the quintessential Canadian stereotype”, but it is also typical of New Zealand (Meyerhoff, 1994), Caribbean (Allsopp, 2003; Westphal, 2021), and Philippine English(es) (Bautista, 2011; Westphal, 2021). Furthermore, *eh* has been shown to be used in other Englishes, including British, Nigerian, and Singaporean English (Columbus, 2009, 2010; Westphal, 2022). Hence, *eh* can be characterized as a translocal pragmatic marker as it is used globally and has developed locally specific patterns of use and meanings.

This paper investigates the translocal dynamics of *eh* by analyzing its use and functions in online written discourse from nine varieties of English using the Corpus of Global Web-based English (GloWbE; Davies, 2013), a 1.9-billion word corpus based on written data in English from general websites and blogs. This analysis focuses only on the American (US), British (UK), Canadian (NZ), Indian (IN), Jamaican (JA), New Zealand (NZ), Nigerian (NG), Philippine (PH), and Singaporean (SG) components, since *eh* has been analyzed in these (or related) varieties in previous research (see Section “[Eh as a Translocal Pragmatic Marker](#)”). In these data, I investigate variation in the frequencies of *eh* (in sentence-final position) across the nine varieties, as well as the constraints of text type and punctuation. I illustrate the different functions of *eh* and analyze how function is constrained by variety, text type, and punctuation. Since *eh* is commonly viewed as a Canadian alternative to *huh* (Denis, 2013; Norrick, 1995), I also include a variationist cross-variety comparison of these two pragmatic markers. I address the following research questions:

- How do variety, text type, and punctuation affect the frequency of sentence-final *eh*?
- How do variety, text type, and punctuation affect the pragmatic function of sentence-final *eh*?

- Which functions does sentence-final *eh* serve in online writing?
- How do variety, text type, and punctuation affect the selection of sentence-final *eh* in contrast to sentence-final *huh*?

This paper adds a global perspective to research on the pragmatic marker *eh*. On a methodological level, I illustrate an innovative approach of a corpus-pragmatic analysis of pragmatic markers in huge online corpora. On a theoretical level, this paper aims to contribute to the understanding of spoken pragmatic markers in online written discourse.

The rest of the paper is structured as follows: In Section “[Eh as a Translocal Pragmatic Marker](#)”, I give a brief overview of previous research on *eh*. I then describe GloWbE and the specific corpus-pragmatic approach of my analysis in Section “[Data and Methods](#)”. Section “[Eh in GloWbE](#)” presents the results on the use of *eh* across the nine varieties. In Section “[Discussion: Eh as a Global Pragmatic Marker in Online Writing](#)”, I discuss these results in relation to previous research, and in Section “[Conclusion: Methodological Implications](#)”, I illustrate the methodological implications of my analysis.

## **Eh as a Translocal Pragmatic Marker**

*Eh* is commonly described as a pragmatic marker (Denis, 2020; Schweinberger, 2018)—or more specifically as an invariant question tag (Columbus, 2009, 2010; Westphal, 2021, 2022)—that is predominantly used in utterance/sentence-final position. The discourse functions of *eh* have been conceptualized in various ways. Avis (1972: 96; cited in Meyerhoff, 1994: 369) views *eh* as an interrogative tag, used “to verbalize enquiries and/or to reinforce them; as such it is a kind of articulated question mark”. Similarly, Denis (2020) describes the main function of *eh* to seek confirmation. Meyerhoff (1994) rejects this view of *eh* as an interrogative tag by showing a general lack of verbal responses to utterances with *eh*. Columbus (2010) illustrates that *eh* may fulfill a wide range of pragmatic functions but is mainly used to add emphasis, as a narrative tag, and for interrogative functions. Westphal (2021, 2022) illustrates that there is a preference for emphatic functions, but *eh* also functions as a narrative and interrogative question tag. The three-way distinction into emphatic, narrative, and interrogative functions is also apparent in Gold and Tremblay’s (2006) list of contexts: *Eh* may be used to add emphasis to statements of opinions or facts, commands, exclamations, antagonistic statements, and fixed expressions. As an interrogative tag, *eh* may be used to ask for information, confirmation, or repetition. *Eh* is also used as a narrative tag to signal that the speaker continues with their story.

Besides being used for these pragmatic functions, *eh* is employed to do interpersonal and identity work. Meyerhoff (1994) argues that *eh* signals solidarity in New Zealand, especially among Māori men. Vine and Marsden (2016) also show that *eh* is important for indexing a Māori identity, but it may additionally signal youth, masculinity, and working class. Besides these social types, they argue that *eh* generally indexes solidarity, interactiveness, informality, and laid-backness. In Canada,

*eh* signals a Canadian national identity, which is enregistered, for example, through T-Shirts or mugs that show a maple leaf in combination with *eh*. Gold and Tremblay (2006) argue that the indexicality of *eh* for Canadianness is much more pronounced in written representations. In the Caribbean, *eh* does not do identity work for a specific social type but is indexical of informality and Caribbean English Creoles (Allsopp, 2003). In the Philippines, *eh*, which is spelled <e> in the International Corpus of English (ICE) for the Philippines, is viewed as a Filipino (i.e., the co-official language of English) marker that is integrated into English (Bautista, 2011).

Despite the localized salience of *eh*, it can be considered a global pragmatic marker since it is used across a wide range of Englishes in spoken discourse. Columbus (2009, 2010) shows that *eh* is present in British, Hong Kong, New Zealand, and Singaporean English, albeit to different degrees, while *eh* is absent from Indian English. *Eh* occurs most frequently in New Zealand English (1460 tokens per one million words [pmw]) but is much rarer in British (30 pmw), Singaporean (25 pmw), and Hong Kong English (5 pmw). Westphal (2021, 2022) attests to the use of *eh* in Trinidadian (465.2 pmw), Philippine (334.8 pmw), and Nigerian English (295.7 pmw), with *eh* being used most frequently in the former and least frequently in the latter variety.

In addition to these cross-variety differences in terms of frequency, *eh* has also developed local patterns of use. Tagliamonte (2006) and Denis (2020) show that the use of *eh* is constrained by age and time in Canada. Both studies illustrate diachronic change: i.e., *eh* is more frequent in older data and among older speakers. In New Zealand, several studies have shown significant variation with regard to age, social class, gender, and ethnicity (Meyerhoff, 1994; Schweinberger, 2018): i.e., *eh* is used most frequently among younger and working-class speakers, males, and Māoris. For the Philippines, Trinidad, and Nigeria, Westphal (2021, 2022) shows that *eh* is more frequent in informal private dialogues than in more formal public ones.

This brief overview has shown that *eh* is a translocal pragmatic marker. This means that *eh* is used globally and generally signals informality but has specific local meanings and patterns of use. Most previous research on the use of *eh* has worked with rather small data sets, except for Columbus' (2009, 2010) and Westphal's (2021, 2022) corpus-pragmatic studies, using data of 250,000 and 230,000 words from individual components of the ICE, respectively. In addition, all these studies are based on spoken data, and there has been no analysis of *eh* in written discourse, except for the discussion of individual examples of *eh* in advertisements and websites by Gold and Tremblay (2006). Furthermore, most research on *eh* has taken a local perspective, analyzing *eh* in Canada and New Zealand, while other varieties or cross-variety differences remain under-researched.

In contrast to the plethora of studies on *eh*, there is little to no research on *huh*. Biber et al., (1999: 1089) include *huh* as an example for invariant question tags and Leech et al., (2009: 243) name *huh* (alongside *right*) as the main invariant alternative to variant question tags in American English. However, *huh* and other spelling variants are not included in concordance lists of corpus-pragmatic studies on invariant tags (e.g., Columbus, 2009, 2010). Norrick (1995) illustrates the discourse functions of *huh* (spelled *hunh*), which are very similar to *eh*. He states that *huh* is

**Table 1** Overview of data

Component	Status of English	Abbreviation	Total size [words]	Blog size [words]	General size [words]
Great Britain	ENL	GB	387,343,392	131,671,002	255,672,390
USA		US	386,597,335	133,061,093	253,536,242
Canada		CA	134,661,559	43,814,827	90,846,732
New Zealand		NZ	81,324,412	22,625,584	58,698,828
India	ESL	IN	96,343,062	28,310,511	68,032,551
Philippines		PH	43,215,533	13,457,087	29,758,446
Singapore		SG	42,940,598	13,711,412	29,229,186
Nigeria		NG	42,619,321	11,996,583	30,622,738
Jamaica <sup>a</sup>	ESD	JM	39,629,689	11,124,273	28,505,416

<sup>a</sup>Jamaica is used as a Caribbean alternative to Trinidad.

commonly used in the USA and explicitly aligns it with *eh*, which he describes as an alternative to *huh* in Australia, Canada, and England.

## Data and Methods

For the analysis of *eh* in online written discourse across different varieties of English, I use GloWbE. This large online corpus has a size of 1.9 billion words based on texts from 1.8 billion websites from 20 different countries where English serves an official function. This analysis only uses the American, British, Canadian, Indian, Jamaican, New Zealand, Nigerian, Philippine, and Singaporean components of GloWbE. This dataset has a size of 1,254,674,901 words. Table 1 shows an overview of the data, sorted according to size and the status of English, i.e., English as a first language (ENL) and English as a second language/dialect (ESL/ESD).

While all data represent non-interactive written online discourse, there is a distinction between general websites and blogs. Davies and Fuchs (2015) hypothesize that blogs represent more spoken-like informal language than general websites. However, they caution that a clear distinction between general and blog texts is not possible (Davies & Fuchs, 2015), and Loureiro-Porto (2017) illustrates that blog texts cannot be equated with spoken language. GloWbE has been used as a larger alternative to the smaller ICE corpora for research on World Englishes (e.g., Unu-abonah et al., 2021), but the GloWbE data need to be treated with some caution in terms of representativeness of different Englishes. Davies and Fuchs (2015) tried to ensure that websites were correctly associated with the specific countries in the corpus compilation. The correct mapping of a website to a country theoretically means that the language use on this website is representative of an online register of a particular variety. However, Davies and Fuchs (2015) find that in practice this mapping does not work perfectly. Polzenhagen (2022) describes one problematic example for Nigerian English. While 66 texts (258,169 words) from the blog *bluechampions*.

com are part of the Nigerian component of GloWbE, he shows that the authors are in fact from India and the UK with no affiliation to Nigeria. Such issues of (un)tidiness are often inherent in huge corpora, but this caveat is somewhat compensated for with their sheer size (see Hansen, 2018).

I use GloWbE for a corpus-pragmatic analysis of the pragmatic marker *eh*. A corpus-pragmatic approach typically combines vertical reading of all corpus texts in a quantitative concordance analysis with horizontal reading of texts for qualitative analyses of individual concordances (Aijmer & Rühlemann, 2015). For such a mixed approach, typically smaller corpora, such as the ICE, are used (e.g., Columbus, 2009, 2010; Westphal, 2021, 2022). While it is possible to attest the presence of specific pragmatic markers and to compare their frequencies across the different national components of GloWbE, further qualitative analyses in this online corpus are somewhat problematic with the online concordance tool because the context of a concordance for horizontal reading is restricted and many links to the respective websites do not work.

Nevertheless, GloWbE has been used for research on pragmatic markers but not for cross-variety corpus-pragmatic analyses. For example, Unuabonah et al. (2021) show that pragmatic markers borrowed from indigenous Nigerian languages (i.e., *abeg*, *sef*, and *na*) are used in written online discourse of the Nigerian component of GloWbE, but they occur to a much lesser degree than in ICE-Nigeria. They illustrate that the three markers serve similar functions in spoken and written discourse, and they do not highlight an intensified indexical loading of the pragmatic markers in GloWbE.

For the analysis of *eh* in GloWbE, I used the online concordance tool offered by [www.english-corpora.org](http://www.english-corpora.org). On the one hand, *eh* and *huh* may be used as pragmatic markers, but on the other hand, they may also be used as a filler lacking the discursive functions described above or as onomatopoeic expressions of feelings like pain or sorrow. Columbus (2010: 305) and Norrick (1995: 689) address this distinction and imply that sentence-final uses are indicative of *eh* and *huh* as pragmatic markers. Hence, I only analyzed sentence-final *eh* (and *huh*), searching for *eh* followed by a full stop, an exclamation point, or a question mark. This approach does not exclude all onomatopoeic uses of *eh*, for example when *eh* is not attached to a sentence but stands on its own, and potentially excepts many uses of *eh* as a pragmatic marker since manually checking all instances is impossible in huge corpora, like GloWbE. However, this approach aims for better comparability to previous research on *eh* as an utterance-final pragmatic marker (Columbus, 2010; Denis, 2020; Schweinberger, 2018). I also searched for *eh* in general to illustrate the share of sentence-final *eh* across the nine varieties. To discover whether, and if so, to what degree the spelling of *eh* is conventionalized in GloWbE, I also searched for spelling variants <e>, <ee>, <eeh>, and <ehh> in sentence-final position. To examine the effects of variety, text type, and punctuation on the use of sentence-final *eh*, I illustrate variation regarding these three variables with descriptive statistics. All normalized results are presented as tokens per one million words (pmw).

For the analysis of the pragmatic functions of *eh*, I combined a quantitative with a qualitative approach. For the quantitative analysis of the pragmatic functions of *eh*, I took a sample of thirty tokens of sentence-final *eh* from each national component,

i.e., ten tokens of <eh.>, <eh!>, and <eh?>, respectively. If possible, the sample was also balanced for text type. From the concordances fulfilling these criteria the tokens for the sample were selected randomly. As *eh* followed by a full stop only occurs seven times in the Indian and six times in the Jamaican component, the overall sample only includes 263 tokens. Cases in which *eh* is used onomatopoeically were removed from the sample and replaced with tokens of *eh* that function as a pragmatic marker.

I coded each of these tokens in terms of their pragmatic function, using a three-way distinction into emphatic, narrative, and interrogatory tags, which have emerged as the three main categories from previous descriptions (Columbus, 2010; Gold & Tremblay, 2006; Westphal, 2021, 2022). Emphatic *eh* is mainly writer-centered and is used stylistically to add emphasis to (evaluative) opinions (1) or fixed expressions (2), for example. Emphatic *eh* can also be more addressee-centered when *eh* is added to emphasize antagonistic statements or (pseudo-) commands (3):

- (1) some men are really morons eh! (US\_43)
- (2) Never mind eh! (GB\_248)
- (3) Anyway good luck to everyone, and KG your \$hit together KG and help Santa out here eh! (CA\_25)

Columbus (2010) describes narrative tags as the opposite of minimal responses in the sense that they signal that the speaker is continuing with their turn. In online written discourse, narrative *eh* is inserted into narrative passages as an informal structuring device (4), which also integrates readers into the discourse, as *eh* implies an addressee:

- (4) So let's start with the make up bits, eh. These are ones that I've not really tried before, but since having them in my life I have found myself using them on a daily basis. (GB\_115)

Interrogative *eh* is mostly addressee-centered and often a particular person or the reader is addressed directly (5). In theory, interrogative *eh* asks for a confirmation or an explanation, and an answer is required. However, it is not always the case that there is a degree of insecurity about the issue at hand on the part of the writer. As a direct response is not possible in the non-interactive online discourse represented in GloWbE, many interrogative cases of *eh* are rather rhetorical and often express the writer's stance. However, there are also cases of reported speech in the corpus and if *eh* is used as an interrogative tag in such cases, a response is often provided (6).

- (5) Look at us and finish us. Criminals, what did we do to you, eh? (NG\_86)
- (6) "So, 5 husbands, eh? Lucky you!" Sita said suddenly to lift the mood. "It was exhausting!" Draupadi said rolling her eyes. (IN\_109)

I show the distribution of functions via descriptive statistics across the nine varieties. To investigate in how far variety, text type, and punctuation influence the function of *eh*, I ran three regression models with function as dependent variable (reduced to a binary distinction) and variety, punctuation, and text type as predictor

**Table 2** Sentence-final *eh* distribution across varieties

Component	GB	US	CA	NZ	IN	PH	SG	NG	JM
<i>eh</i> raw	4547	3199	1077	818	303	636	571	298	196
Sentence-final <i>eh</i> raw	3387	2073	860	580	194	348	258	147	113
%-sentence-final <i>eh</i>	74.49	64.80	79.85	70.90	64.03	54.72	45.18	49.33	57.65

variables. In the first model, emphatic is the application value (compared to both other functions), the second model compares narrative functions to both other functions, and the third model uses interrogative as the application value. The approach of treating function as a dependent variable and punctuation as a predictor variable reflects the researcher's perspective of interpreting the data and not necessarily of what happens during language production. This quantitative account of the pragmatic functions of *eh* is enriched with further (purely) qualitative analyses of individual examples, showing the indexicalities (beyond the immediate pragmatic functions) of *eh* in online writing.

For the variationist analysis of *eh* in contrast to *huh*, I searched for sentence-final *huh* in the nine varieties. To illustrate the patterns of variation for this distinction, I ran a binary regression model with sentence-final pragmatic marker (*eh* vs. *huh*) as dependent variable and variety, punctuation, and text type as predictor variables. It is important to keep in mind that *eh* and *huh* are not perfectly synonymous and there is a wide range of alternative forms that may be used instead of these two pragmatic markers (see Pichler, 2010). Hence, this comparison only provides a limited picture but helps to provide an additional perspective on variation of *eh* across Englishes.

## ***Eh* in GloWbE**

### **Variation of *eh* with regard to variety, punctuation, and text type**

The analysis attests to the presence of sentence-final *eh* in online discourse of all nine varieties. Overall, there are 7960 occurrences (6.34 pmw) of sentence-final *eh*, which corresponds to 68.36% of all *eh* tokens (11,645; 9.28 pmw) in the data. Table 2 depicts the raw frequencies of *eh* and sentence-final *eh*, as well as the share of *eh* in sentence-final position of all *eh* tokens for each national component. Figure 1 shows the normalized frequencies of *eh* and sentence-final *eh* across the nine components.

Figure 1 shows that sentence-final *eh* occurs most frequently in the British component, followed by the Philippines, New Zealand, Canada, Singapore, and the USA. In the Jamaican, Nigerian, and Indian components, sentence-final *eh* occurs less than half as frequently as in the British component. To validate the high frequency of sentence-final *eh* in the British component, I scanned the data qualitatively and did not find frequent patterns of sentence-final *eh* where it does not function as a



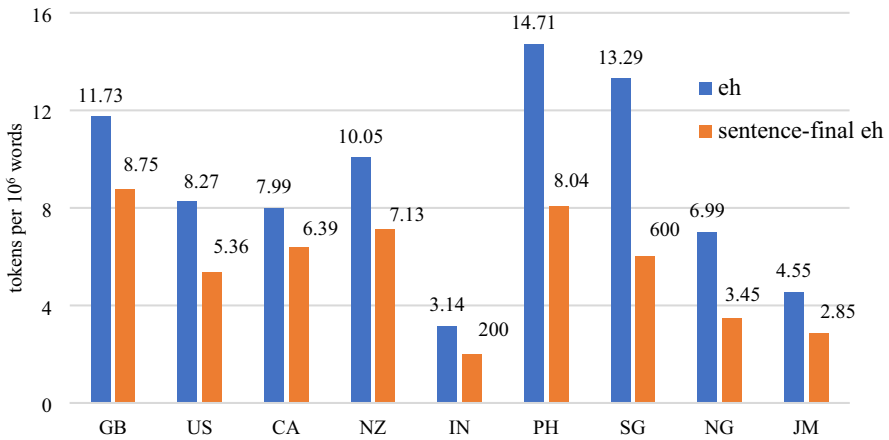


Fig. 1 Eh across the nine components of GloWbE

pragmatic marker. The high frequency of *eh* in the Philippine component is due to the high degree of code-switching to Filipino. This fact also explains the relatively low share of sentence-final *eh* in the Philippine data, as *eh* occurs frequently in Filipino sentences (7) and hybrid mixes of English and Filipino (8). In these cases, *eh* is more integral to the syntax and propositional meaning of the sentences than in sentence-final positions in sentences that are dominantly written in English.

- (7) I show her how to pee sa toilet bow. Eh ang mga bata gaya gaya di ba? ('Children are like that, aren't they?') (PH)
- (8) After our late lunch, we went to the beach to bum. Hahahaha! 😊 Ang predictable eh no ('It's predictable isn't it') (PH)

The lowest shares of *eh* in sentence-final position are evident in the Nigerian and Singaporean data, where *eh* is frequently used onomatopoeically, for example as an expression of sorrow (9) or as a sign of hesitation (10), and when (cat)calling other people (11). The share of *eh* in sentence-final position is highest in the Canadian, British, and New Zealand components, where cases of *eh* not used as a pragmatic marker seemed rare when scanning the data qualitatively. These cross-variety differences should be treated with some caution because they do not necessarily show that *eh* is substantially more typical of British English in contrast to all other Englishes, but these differences may also indicate different levels of informality in online writing.

- (9) eh eh eh may God help us o. (NG)
- (10) and the sun radiation isn't so strong.. Eh... how can it not go away (SG)
- (11) Did you think to leave without saying goodbye? Eh, Dunni? (NG)

The spelling of *eh* is very consistent in the data. Other spelling variants are rare, and <e> and <ee> do not seem to be used as pragmatic markers. Both forms are

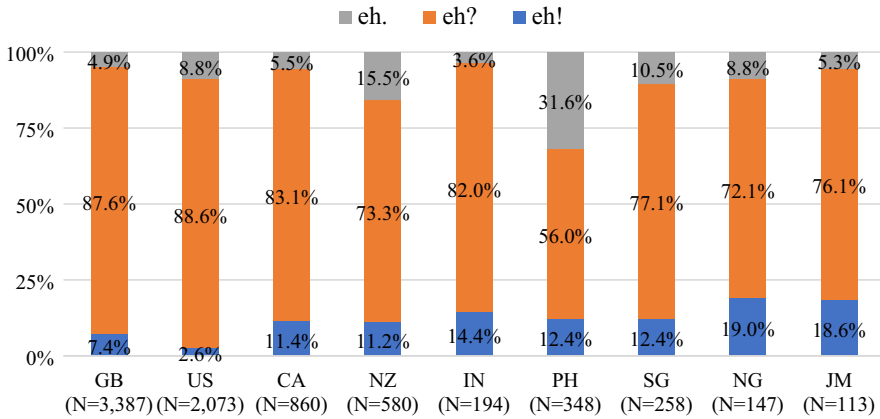


Fig. 2 Punctuation after sentence-final *eh*

mainly used for abbreviations or appear when there were problems with non-Latin script in the data compilation process. In the Philippine component, there are individual cases of <e> being used a pragmatic marker (12). This spelling is rare but corresponds to the transcription conventions in ICE-Philippines (Bautista, 2011). However, Filipino online writers prefer the spelling <eh> over <e>. While there are instances of <ehh> being used a pragmatic marker, onomatopoeic uses are more frequent (13).

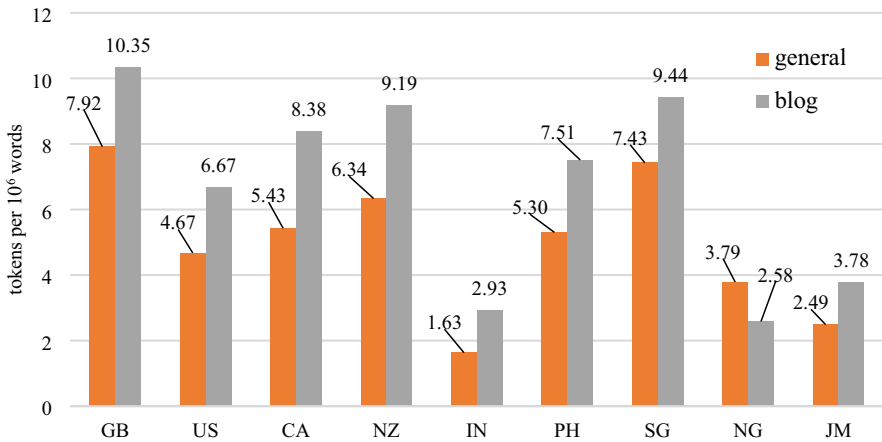
- (12) but people should still be free to voice opinions. kaya nga opinions, e! ('that's why they are opinions') (PH)
- (13) Used to love him. Now I'm like ehh. (US)

In contrast to <ee>, <e>, and <ehh>, <eeh> in sentence-final position (39; 0.031 pmw) is mostly used as a pragmatic marker. However, *eeh* seems to be mainly characteristic of the Jamaican data (14).

- (14) Cho! Oonoo boring eeh? ('You are boring') (JM)

Patterns of punctuation are relatively consistent across the nine GloWbE components. Figure 2 shows the frequencies of the different punctuation patterns as percentages for each national variety. The combination of *eh* with a question mark is most frequent, accounting for 84.0% of all occurrences of sentence-final *eh*. *Eh* combined with a full stop (8.2%) and with an exclamation point (7.8%) are substantially less frequent. This combination mirrors the interrogative function of *eh* in contrast to the other two punctuation variants. The combination of *eh* with a full stop is most frequent in the Philippine component (31.6%). In many of these cases, the sentences ending with *eh* are strongly marked for Filipino, as in (15).

- (15) Parang puppy love lang eh. ('It's just like puppy love') (PH\_46)



**Fig. 3** Text type variation of sentence-final *eh*

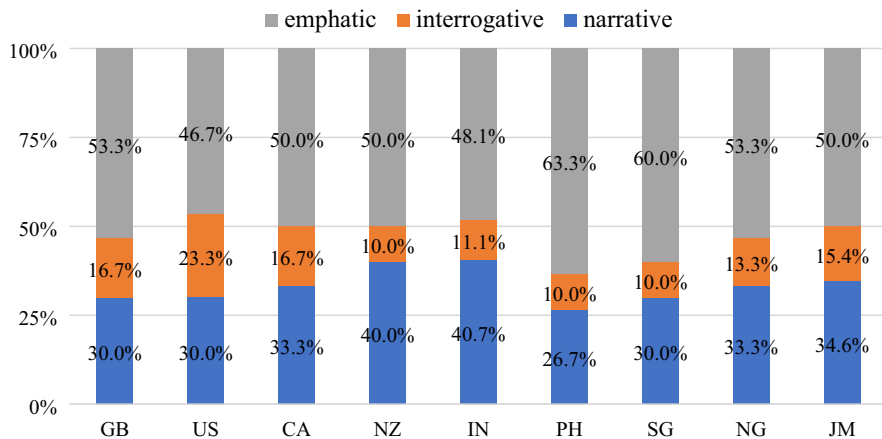
Text type variation is also consistent across the nine components. Figure 3 shows the normalized frequencies of sentence-final *eh* in blogs and general websites for each component. Except for the Nigerian component, *eh* is used more frequently in blogs than in general websites. These findings confirm Davies and Fuchs' (2015: 26) descriptions of blogs representing more informal spoken-like language than texts from general websites.

### Functions of *eh* in GloWbE

The analysis of the sample of 263 *eh* tokens illustrates the distribution of pragmatic functions and allows filtering out tokens that do not function as a pragmatic marker. From the original sample, nine cases of *eh* being used onomatopoeically as an exclamation of sorrow or disapproval, such as (16), were replaced. Three of these nine cases stem from the Nigerian component. This shows that the clear majority of concordances of sentence-final *eh* (i.e., 96.5% in the sample) are pragmatic markers.

- (16) It was around this time that, when I visited, my grandmother would answer my "How are you?" with the phrase, "Eh. I'm already dead and my body just doesn't know enough to lay" (CA)

The results of the quantitative analysis of the functions are presented in Figure 4 and illustrate that *eh* is mainly used emphatically (52.8%), followed by narrative uses (33.1%), while interrogative uses are least frequent (14.1%). This distribution of functions is relatively consistent across the nine components, while there are some cross-variety differences. For example, emphatic uses are most dominant in the Philippine component (63.3%), narrative functions have the highest share in the New Zealand (40.0%) and Indian (40.7%) components, and interrogative functions occur comparatively most frequently in the US (23.3%) component. However, the overall results and the cross-variety differences in the sample need to be treated with



**Fig. 4** Distribution of functions of sentence-final *eh*

**Table 3** Results of the regression analyses

Model	Application value	Input probability	<i>p</i> -variety	<i>p</i> -text type	<i>p</i> -punctuation	R <sup>2</sup>
1	Emphatic	0.526	0.893	0.202	<0.001	0.11
2	Narrative	0.331	0.932	0.416	0.006	0.06
3	Interrogative	0.085	0.819	0.448	<0.001	0.35

some caution because the sample is comparatively small, the differences between the components are influenced by chance, and punctuation influences function.

The results of the regression analyses demonstrate that variety and text type have no significant effect on the function of sentence-final *eh*. Punctuation, by contrast, influences the variation in function significantly in all three models. Table 3 gives an overview of the input probabilities of the application values, *p*-values of each predictor variable, and R<sup>2</sup>-values. The direction and effect size of the individual levels of the predictor variable punctuation (exclamation point, full stop, question mark) are shown as (centered) factor weights, which are plotted in Figure 5. The values are connected with lines to facilitate interpretation although they do not represent an evolution but independent results. Each line represents the results of one model with the respective application value. Factor weights range from 0 to 1. Values above 0.5 indicate a statistical preference for the application value and values below 0.5 a statistical dispreference. The steeper the line and/or the higher the range of the factor weight values the more pronounced the effect of punctuation on the specific application value.

These results illustrate that there is a preference for emphatic uses for *eh* with an exclamation point, whereas there is only a very marginal dispreference for *eh* combined with the other two punctuation marks. While the model shows a significant effect of punctuation on emphatic uses, the curve is relatively flat, the range of factor weight values is comparatively low ( $\Delta = 0.28$ ), and the model only explains 11% of

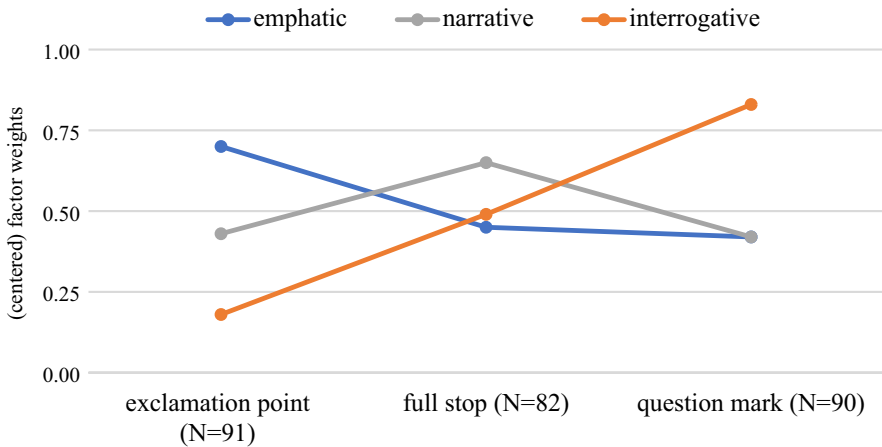


Fig. 5 Results regression analyses of functions

the variance in the sample. There is a comparative statistical preference for emphatic uses with a punctuation but only a very marginal dispreference for full stop and question mark. For narrative functions, the effect of punctuation is still significant but the least pronounced of all three models, illustrated via the comparatively flattest curve and the lowest range ( $\Delta=0.23$ ). In addition, the model only explains 6% of the variation in the sample. There is a slight statistical preference for narrative uses of *eh* combined with a full stop, and a very minor dispreference for the combination of *eh* with the other two punctuation marks. For interrogative functions, the effect of punctuation is most pronounced, shown via the comparatively steepest curve and the highest range ( $\Delta=0.65$ ). In addition, the model is also the best from all three, explaining 35% of the variance. Interrogative functions are strongly preferred for *eh* followed by a question mark and strongly dispreferred for *eh* with an exclamation point. *Eh* combined with a full stop does not show an effect on interrogative functions.

While this quantitative account gives a first overview of the general patterns of the pragmatic functions of *eh* in the online discourse of GloWbE, the analysis works on a very high level of abstraction and cannot illustrate the context-specific indexicalities of *eh*. In addition, the functional classification is at times problematic because there is very little context given and many websites were not accessible. To a certain degree all *eh* tokens that function as pragmatic markers index informality in writing (McCarthy, 2013) because *eh* is associated with informal spoken discourse (e.g., Schweinberger, 2018: 6; Vine & Marsden, 2016: 401).

The emphatic category is relatively well defined through the different contexts emphasized with *eh*. *Eh* is mainly used to emphasize evaluative opinions expressed in verbless expressions, such as *Awesome, eh!* or *Crazy, eh!*. In the context of exclamations, *eh* is frequently part of sayings, such as *Time flies, eh?*. Fixed expressions in the sense of conventionalized speech acts, such as thanking, are marginal in the sample. Only *Never mind eh!* was coded as a fixed expression.

In many cases where *eh* serves emphatic functions, it also adds an antagonistic tone to the sentence (17) or is used when reluctance towards something is expressed (18). In these cases of antagonism, the onomatopoeic quality of *eh* is additionally exploited by the writers.

- (17) Liberal Atheist go back to Canada eh. (US\_19)  
 (18) Ross and Moira have blocked me from their Facebook page. Can't handle the heat eh? (NZ\_77)

The indexicality of orality of sentence-final *eh* is very pronounced in these cases of antagonism but indexing orality is also a general quality of *eh*, evident in many cases. For example, *eh* is frequently used in contexts of direct speech, as in (19).

- (19) Fadi: Oh yeah? Olivia: Not in that way... he's just a friend... Fadi: "Just a friend", eh? Olivia: Are you jealous? Fadi: Get out of here! Course I'm not. (GB\_141)

The oral quality of *eh* is also pronounced when a specific person or the reader is addressed directly (20), often with second-person pronouns (21) or first-person plural pronouns (22). Such direct addresses imply a dialogic situation in the asynchronous CMC discourse of GloWbE.

- (20) Houellebecq's novel is a clone (ha, get the joke? Eh?) (JM\_18)  
 (21) Belief is not enough, even the demons believe. Ironic that demons believe in God, but you people don't, eh? (US\_1376)  
 (22) OK really when one weighs the pros & cons Otherwise what would really happen if We All saw Eye to Eye with each other eh! (IN\_22)

In addition to orality, *eh* also indexes informality in the online data. This indexicality is evident through the higher frequencies of *eh* in the more informal text type of blogs (Figure 3) but also through the co-occurrence of *eh* with a wide range of other linguistic features that index informality. For example, in the Jamaican component, *eh* frequently co-occurs with Jamaican Creole, as in (23). The writer not only employs the Jamaican Creole third-person plural pronoun *them* but also uses stylized spelling to imitate spoken speech: <g> in *smoking* is left out to imitate the pronunciation of (ING) as [ɪn], *enough* is spelled <nuf> to imitate the Jamaican pronunciation [nɒf], and *that* is spelled <dat> to imitate TH-stopping, which is also typical of a Jamaican pronunciation. In addition, the writer uses a very informal register, for example, *cokin* is used as a verb that describes the act of consuming cocaine. This co-occurrence with an informal register and stylized spelling that imitates an informal pronunciation is not restricted to the Jamaican component but it is evident in many cases across the entire GloWbE data.

- (23) Its the smokin and the cokin dat getting to dem brain so they lose their senses and are no longer men with sense to see the degradation of their ways. "Nice rap eh!" nuf said. (JM\_15)

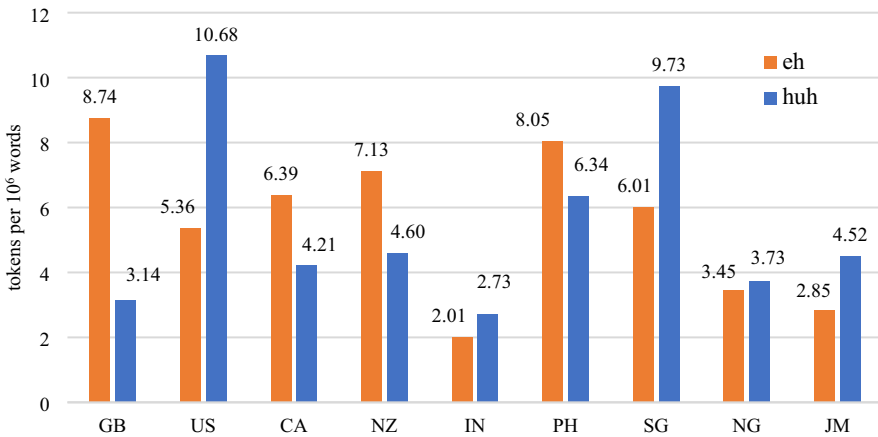
In the Philippine component, *eh* frequently co-occurs with code-mixing of English and Filipino or switches to sentences entirely in Filipino, as in (12) or (24), where the author also provides a translation to Standard English. The translation of *eh* with *you know* indicates that the writer views *eh* as Filipino and not as English.

- (24) My son then intervened and said, “Sumbong kita sa manager nyo. Customer kami eh.” (“I will report you to the manager. This is not the way to treat a customer, you know”) (PH\_3)

In the Nigerian component, *eh* frequently co-occurs with pragmatic markers that are typical of Nigerian English/Pidgin, such as *o* or *na*, which both mainly serve emphatic functions (Westphal, 2022; Unuabonah et al., 2021). *Eh*, *o*, and *na* are phonologically reduced pragmatic markers and have a rather pronounced onomatopoeic quality. This co-occurrence of *eh* with other pragmatic markers that have an onomatopoeic quality, the low share of *eh* in sentence-final position (in relation to the raw frequencies of *eh*) (Table 2), as well as the exclusion and subsequent replacement of several *eh* tokens that did not function as pragmatic markers in the quantitative analysis of the functions of *eh* suggest a comparatively heightened onomatopoeic quality of *eh* in the Nigerian data. Example (25) illustrates the co-occurrence of *eh* with *o* and *na*. This example is also characterized by ‘grassroots spelling’ (Blommaert, 2008): All letters are lower case, apostrophes (<im> for I’m) and individual graphemes are absent (<o> in *know*), and the spelling generally imitates pronunciation (<u> for *you*, <den> for *then*). These linguistic features are not only characteristic of an informal register but have also been described as typical of online discourse (e.g., Tagg, 2012). Hence, *eh* as a pragmatic marker also characterizes an informal online register that imitates spoken discourse.

- (25) im a gud guy o! buh she got to knw of a fling i had buh still there for me..... if she cooks for u eh!...  
den check her beauty out na..... (NG\_27)
- ‘I’m a good guy. But she got to know of a short-term sexual relationship I had but she is still there for me. If she cooks for you. Then check her beauty out.’

These co-occurrences of *eh* with other linguistic features typical of Jamaican, Philippine, and Nigerian English illustrate the local indexicalities of *eh* in these three contexts. The Canadianness and New Zealandness of *eh* and the local indexicalities in these two locations are less evident in the GloWbE data. Nevertheless, there are individual examples from the Canadian and other components which show that *eh* is used to index Canada or Canadianness. For example, the author of a travel blog, which is part of the New Zealand component, uses the heading “Canada, eh?” in their blog entry on their travels through Canada. The writer in (26) uses *eh* in a rant blog post about Canadian politics. *Eh* serves to express their aversion to Canadian politics and highlights the Canadianness of the problem they describe in their blog post.



**Fig. 6** Sentence-final *eh* versus *huh*

(26) Now that's true governance Canadian style, eh. (CA\_32)

In the New Zealand component, there are individual examples that highlight the indexicality of *eh* for a Māori identity. One blog author uses “Indigenous eh?” and “Maoris, eh?” derogatively in their posts about Māori people.<sup>1</sup> Besides these individual examples, the general indexicalities of orality and informality are much more prominent in the Canadian and New Zealand components.

### ***Eh* in Contrast to *huh***

The variationist comparison of *eh* and *huh* (in sentence-final positions) provides an additional perspective on cross-variety differences in the use of *eh*. With 7578 (49.68 pmw) occurrences, sentence-final *huh* is approximately as frequent as sentence-final *eh* in the nine components of GloWbE. Other spelling variants, such as *hunh* (Norrick, 1995) are extremely rare. However, there are substantial differences regarding the distribution across the nine components. Figure 6 shows the normalized frequencies of sentence-final *huh* and *eh* for each national component. *Huh* is more frequent than *eh* in the US, Singaporean, Indian, Jamaican, and Nigerian components, whereas this pattern is reversed for the British, Canadian, New Zealand, and Philippine components. The US component sticks out, with *huh* being used by far most frequently and almost twice as frequently as *eh*. In contrast, the difference in favor of *eh* is most pronounced in the British component, where *eh* occurs almost three times more frequently than *huh*.

Besides these cross-variety differences, the patterns of variation for *huh* in GloWbE regarding punctuation and text type are very similar to *eh*. Figure 7 shows the distribution of the punctuation patterns of sentence-final *huh* across the nine

<sup>1</sup> Examples are excluded to avoid reproducing racist statements.



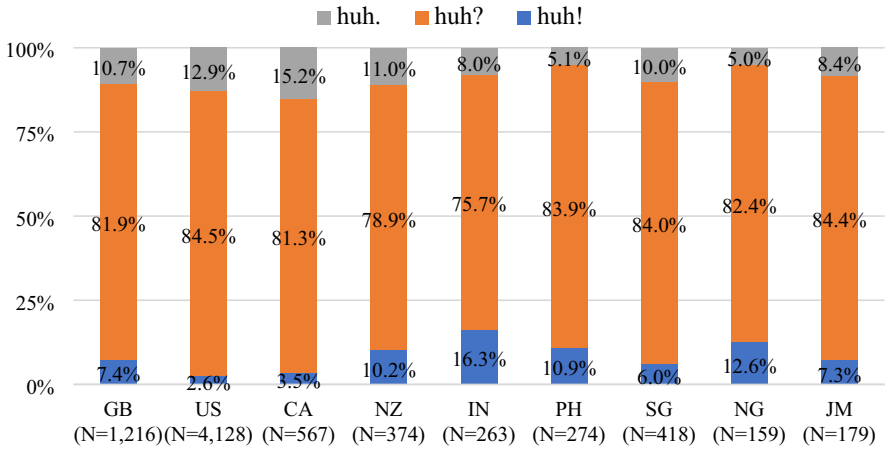


Fig. 7 Punctuation of sentence-final *huh*

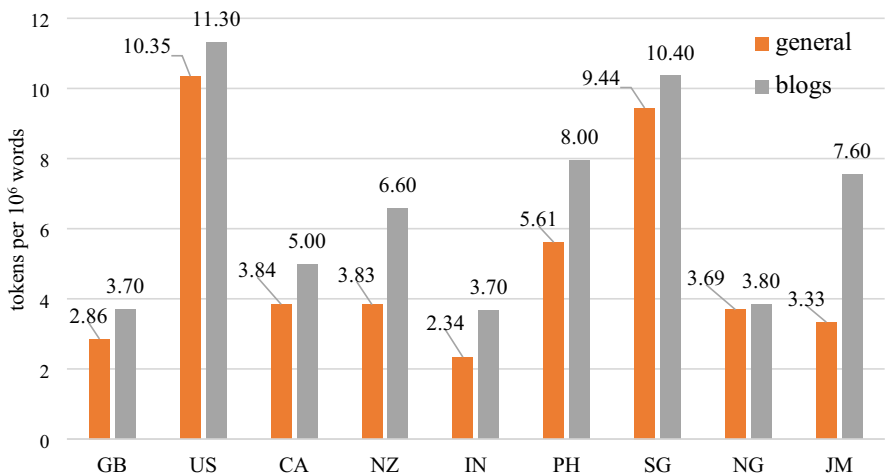


Fig. 8 Text type variation of sentence-final *huh*

components, and Figure 8 illustrates the patterns for text type variation. The combination of *huh* with a question mark dominates by far over the combination with a full stop and an exclamation point. Of all 7578 tokens, 83.2% are combined with a question mark, 11.7% with a full stop, and 5.1% with an exclamation point. In terms of text type variation, *huh* is used more frequently in blogs than in general websites.

The regression analysis of sentence-final *eh* in contrast to *huh* confirms the cross-variety differences shown in the descriptive statistics and illustrates more fine-grained differences between the two pragmatic markers in online writing concerning punctuation and text type variation. The results of the regression analysis with *eh* as the application value are shown in Table 4. All three predictor variables are shown

**Table 4** Results regression analysis *eh* versus *huh*; application value=*eh*; input probability-*eh*=0.52;  $R^2=0.15$ 

Predictor variable	Level	<i>N</i>	%- <i>eh</i>	(Centered) factor weight
Variety $p < 0.001$	GB	4603	73.3	0.73
	NZ	954	60.8	0.60
	CA	1427	60.3	0.59
	PH	622	55.9	0.56
	NG	365	56.4	0.54
	IN	457	42.5	0.41
	SG	676	38.2	0.37
	JM	292	38.7	0.37
	US	6201	33.4	0.33
Punctuation $p < 0.001$	Exclamation point	1063	63.7	0.58
	Question mark	12,997	51.5	0.50
	Full stop	1537	42.2	0.42
Text type $p < 0.001$	Blog	6104	53.5	0.52
	General	9493	50.1	0.48

to have a significant effect on the distribution. However, due to the large token count even very fine-grained differences reach the level of significance. It is thus important to consider the effect sizes of the individual levels more closely. While the model illustrates differences between *eh* and *huh* regarding punctuation and text type, the effect size differences are small to marginal. In terms of punctuation, there is a slight preference for *eh* in combination with an exclamation point, no effect for the combination with a question mark, and a slight statistical dispreference for the combination with a full stop. Regarding text type variation, the model shows a very marginal preference for *eh* in contrast to *huh* in blogs, and a reversed pattern for general web-sites. This may indicate that *eh* is slightly more informal than *huh*. The cross-variety differences are more pronounced in the data. The model shows a strong preference for *eh* in contrast to *huh* for the British component and a reversed effect for the US component. *Eh* is preferred in the New Zealand, Canadian, Philippine, and Nigerian components, whereas there is a statistical dispreference in the Indian, Singaporean, and Jamaican components. Despite these marked differences concerning variety, it is important to keep in mind that variety differences in GloWbE need to be interpreted with much caution and there is still much unexplained variation in the data, as this model only explains 15% of the variance.

## Discussion: *Eh* as a Global Pragmatic Marker in Online Writing

In contrast to previous research that has analyzed *eh* as a local pragmatic marker, mostly in Canada (Avis, 1972; Denis, 2013, 2020; Tagliamonte, 2006) and New Zealand (Meyerhoff, 1994; Schweinberger, 2018; Vine & Marsden, 2016), this

analysis illustrates that *eh* is a global pragmatic marker used in written online discourse across nine Englishes. In contrast to other spoken pragmatic markers in writing, such as *na* in Nigerian English (Unuabonah et al., 2021), the spelling of *eh* is strongly conventionalized across the nine components. In comparison to previous corpus-pragmatic research on *eh* in spoken discourse (Columbus, 2009, 2010; Westphal, 2021, 2022), *eh* is used less frequently in online writing. Like in spoken discourse, there are also substantial differences between the varieties in GloWbE. While *eh* is used frequently in the New Zealand and Canadian components, the results show that *eh* is used most frequently in the British component followed by the Philippines. *Eh* occurs most rarely in the Indian component and comparatively infrequently in the Nigerian and Jamaican components. The US and Singaporean components occupy mid-positions. However, these variety differences need to be treated with much caution due to issues of (un)tidiness and representativeness of GloWbE (Loureiro-Porto, 2017; Polzenhagen, 2022). Variation across the components may also represent regional differences in online writing practices. The use of spoken pragmatic markers, such as *eh* and *huh*, in online writing is indicative of colloquialization, i.e., the incorporation of informal spoken features into written language (Mair, 1997). The comparatively lower frequencies of *eh* (and *huh*) in the Indian and Nigerian components may be indicative of comparatively more formal writing styles in these two ESL countries, while the high frequencies of *eh* in the British component and of *huh* in the American component could point towards a higher degree of colloquialization of online writing in these two ENL countries.

The variationist analysis of *eh* in contrast to *huh* shows that *huh* seems to be particularly typical of American English and *eh* of British, Philippine, Canadian, New Zealand, and Nigerian English sampled in GloWbE. Except for British English, previous research has illustrated the frequent use of *eh* in spoken discourse in these varieties (Columbus, 2009, 2010; Westphal, 2021, 2022). This finding points towards a high potential of research on *eh* in British English, which has been under-researched so far. Columbus (2009, 2010) shows that *eh* is rare in spoken British English, but her data from ICE-GB represents Standard English usage, with many speakers being from the South of England. The use of *eh* in comedy performances that depict a stylized (working-class) northern British accent, such as the Monty Python sketch on four Yorkshiremen<sup>2</sup> or a sketch from the BBC on the reaction of a posh family from London to their northern nanny<sup>3</sup>, point towards northerness and working-class connotations of *eh*. The general lack of research on *huh* and the high frequency of the form in GloWbE point towards an additional research gap.

Despite these cross-variety differences, writers use sentence-final *eh* in very similar ways in the nine components. *Eh* is predominantly combined with a question mark, and it is used more frequently in blogs, which confirms the comparatively greater informality of this text type (Davies & Fuchs, 2015). While *eh* may fulfill interrogative and narrative functions as in spoken discourse (Gold & Tremblay, 2006), there is a preference for emphatic functions in the written data, similar to

<sup>2</sup> Online at: <https://www.youtube.com/watch?v=ue7wM0QC5LE>

<sup>3</sup> Online at: <https://www.youtube.com/watch?v=g-BVgPeZR-Y>

Columbus' (2009) and Westphal's (2021, 2022) findings for spoken data. *Eh* signals informality and orality across all components, which can be regarded as global indexicalities of *eh*—similar to other spoken pragmatic markers in writing (McCarthy, 2013). As *eh* has also been shown to index informality in spoken discourse (Vine & Marsden, 2016), these results indicate an overlap in the functions of *eh* in speech and writing. However, in contrast to previous descriptions of *eh* in written discourse as highly marked (Gold & Tremblay, 2006; see also Brinton, 2017), the comparatively high frequency of *eh* in GloWbE indicates that this spoken pragmatic marker seems to be well integrated into (informal) online writing.

The global indexicalities of orality and informality of *eh* in online writing seem to overwrite the local indexicalities, such as Canadianness or a Māori ethnic identity, and its potential for a construction of a specific online persona (Tagg, 2012). However, local indexicalities are not absent and they are most pronounced in the Philippine component, where there is a strong link of *eh* to Filipino, which is also evident in spoken discourse from the Philippines (Westphal, 2021). In the Jamaican and Nigerian components, *eh* is indexical of Creole and Pidgin, respectively. Hence, *eh* is often part of (written) multilingual language use in ESL/ESD contexts. The local indexicalities in the Canadian and New Zealand components are not as pronounced, but this does not mean that writers may not use *eh* for various local purposes. More detailed qualitative analyses of longer stretches of discourse would be necessary to investigate these more localized practices of *eh*. Analyses of memes that contain metalinguistic commentary on *eh* or online ethnographies of specific New Zealand or Canadian online communities may be more suitable to examine the local indexicalities of *eh* in online discourse.

## Conclusion: Methodological Implications

The current approach shows that large online corpora, such as GloWbE, may be used for studying spoken pragmatic markers in written discourse. In contrast to traditional written contexts, spoken pragmatic markers seem to be more common and more smoothly integrated into this more informal written discourse. In addition, spoken pragmatic markers seem to stand out less than in traditional written contexts, where they are often used in stereotypical ways (Gold & Tremblay, 2006), which impedes systematic analyses of patterns of use. However, the approach taken in the present paper comes with many methodological caveats. Issues of (un)tidiness are a serious concern for GloWbE (Hansen, 2018), and although I focused the search queries on *eh* in sentence-final position (i.e., the typical position of *eh* as a pragmatic marker), the concordances still include many onomatopoeic uses of *eh*. Qualitative filtering of all concordances is not possible due to the sheer size of the corpus. Hence, a sample was taken for further analysis to draw conclusions about the entire data. However, due to the restricted context, this approach is also not ideal. Hence, researchers willing to work on pragmatic markers in GloWbE are advised to get access to the full texts.

The purely quantitative corpus-linguistic approach of counting frequencies of sentence-final *eh* in different components and text types comes with additional

caveats. The validity of these findings is somewhat limited because variation in the data may be caused by cross-variety differences in the use of *eh* and online writing practices. The variationist approach, which analyzes the effects of text type, punctuation, and variety on the function of *eh* and on the use of *eh* in contrast to *huh*, is able to illustrate more in-depth patterns of use. This approach can also illustrate the Canadianness and New Zealandness of *eh* in contrast to *huh*. However, there are many alternatives to *eh* and *huh*, and the two pragmatic markers are not exactly synonymous, a problem faced by all variationist studies using pragmatic phenomena (Pichler, 2010). Hence, the comparative study is far from ideal, and my approach only provides a limited picture of spoken pragmatic markers in online writing. Researchers interested in future research on spoken pragmatic markers in GloWbE are advised to focus on forms that are (almost) exclusively used as pragmatic markers, such as *inmit* in British English, or focus the search queries even more to reduce the amount of non-pragmatic marker tokens.

Despite these caveats, my analysis has shown that spoken pragmatic markers are part of written online discourse, which provides a valuable source for studying the dynamics of pragmatic markers in writing. Previous descriptions of pragmatic markers in written discourse need to be revised (or extended) as writers in CMC contexts draw on features of spoken language, such as pragmatic markers, more readily and not only use them for very stereotypical purposes. Other online genres, such as Tweets, might provide even more informal contexts and may show higher frequencies of spoken pragmatic markers than the websites and blogs sampled in GloWbE. While writers generally seem to use *eh*—and potentially other markers—to index informality and orality, they may also employ them for metaphorical purposes, such as signaling antagonism, and for (local) identity work. Much previous research on pragmatic markers has focused strongly on their local indexicalities and local patterns of use, but since many markers are used globally, future research should pay more attention to the translocal qualities of pragmatic markers.

**Funding** Open Access funding enabled and organized by Projekt DEAL.

## Declarations

**Conflict of interest** On behalf of all authors, the corresponding author states that there is no conflict of interest.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## References

- Aijmer, K. (2013). *Understanding pragmatic markers in English*. Edinburgh University Press.
- Aijmer, K., & Rühlmann, C. (Eds.). (2015). *Corpus pragmatics*. Cambridge University Press.
- Allsopp, R. (2003). *Dictionary of Caribbean English usage*. University of the West Indies Press.
- Avis, W. (1972). So eh? is Canadian eh? *Canadian Journal of Linguistics*, 17(2/3), 89–104.
- Bautista, L. (Ed.). (2011). *Studies of Philippine English: Exploring the Philippine component of the International Corpus of English*. Anvil.
- Beeching, K. (2016). *Pragmatic markers in British English*. Cambridge University Press.
- Biber, D., Johansson, S., Leech, G., Conrad, S., & Finegan, E. (1999). *The Longman grammar of spoken and written English*. Longman.
- Blommaert, J. (2008). *Grassroots literacy: Writing, identity and voice in central Africa*. Routledge.
- Brinton, L. (2017). *The evolution of pragmatic markers in English*. Cambridge University Press.
- Columbus, G. (2009). A corpus-based analysis of invariant tags in five varieties of English. In A. Renouf & A. Kehoe (Eds.), *Corpus linguistics: Refinements and reassessments* (pp. 401–414). Rodopi.
- Columbus, G. (2010). “Ah lovely stuff, eh”—invariant tag meaning and usage across three varieties of English. In S. Gries, S. Wulff, & M. Davies (Eds.), *Corpus-linguistic applications: Current Studies New Directions* (pp. 85–102). Rodopi.
- Davies, M. (2013). *Corpus of Global Web-Based English: 1.9 billion words from speakers in 20 countries (GloWbE)*. <https://www.english-corpora.org/glowbe/>
- Davies, M., & Fuchs, R. (2015). Expanding horizons in the study of World Englishes with the 1.9 billion word Global Web-based English Corpus (GloWbE). *English World-Wide*, 36(1), 1–28.
- Denis, D. (2013). The social meaning of eh in Canadian English. In *Proceedings of the 2013 annual conference of the Canadian linguistics association*.
- Denis, D. (2020). How Canadian was eh?: A baseline investigation of usage and ideology. *The Canadian Journal of Linguistics*, 65(4), 583–592.
- Gold, E., & Tremblay, M. (2006). Eh? and hein?: Discourse particles or national icons. *The Canadian Journal of Linguistics*, 51(2/3), 247–263.
- Hansen, B. (2018). *Corpus linguistics and sociolinguistics: A study of variation and change in the modal systems of World Englishes*. Brill.
- Leech, G., Hundt, M., Mair, C., & Smith, N. (2009). *Change in contemporary English: A grammatical study*. Cambridge University Press.
- Loureiro-Porto, L. (2017). ICE Vs. Glowbe. *World Englishes*, 36(3), 448–470.
- Mair, C. (1997). Parallel corpora: A real-time approach to the study of language change in progress. In M. Ljung (Ed.), *Corpus-based studies in English* (pp. 195–209). Rodopi.
- McCarthy, M. (2013). Spoken discourse markers in written text. In G. Fox, M. Hoey, & J. Sinclair (Eds.), *Techniques of description: Spoken and written discourse* (pp. 170–182). Taylor and Francis.
- Meyerhoff, M. (1994). Sounds pretty ethnic, eh?: A pragmatic particle in New Zealand English. *Language in Society*, 23(3), 367–388.
- Millar, S. (2015). Blathering beauties: The use of pragmatic markers on an Irish beauty blog. In P. Amador-Moreno, K. McCafferty, & E. Vaughan (Eds.), *Pragmatic markers in Irish English* (pp. 292–317). Benjamins.
- Norrick, N. (1995). Hunh-tags and evidentiality in conversation. *Journal of Pragmatics*, 23, 687–692.
- Pichler, H. (2010). Methods in discourse variation analysis: Reflections on the way forward. *Journal of Sociolinguistics*, 14(5), 581–608.
- Polzenhagen, F. (2022). Critical reflections on the use of corpora for cross-varietal metaphor research. In U. Schröder, M. de Oliveira, & A. Tenuta (Eds.), *Metaphorical conceptualizations: (Inter)cultural perspectives* (pp. 23–31). De Gruyter.
- Schweinberger, M. (2018). The discourse particle eh in New Zealand English. *Australian Journal of Linguistics*, 38(3), 395–420.
- Tagg, C. (2012). *The discourse of text messaging: Analysis of SMS communication*. Continuum.
- Tagliamonte, S. (2006). “So cool, right?”: Canadian English entering the 21st century. *The Canadian Journal of Linguistics*, 51(2/3), 309–331.
- Unuabonah, F., Oyebola, F., & Gut, U. (2021). “Abeg na! We write so our comments can be posted!”: Borrowed Nigerian Pidgin pragmatic markers in Nigerian English. *Pragmatics*, 31(3), 455–481.
- Vine, B., & Marsden, S. (2016). Eh at work: The indexicality of a New Zealand English pragmatic marker. *Intercultural Pragmatics*, 13(3), 383–405.

- Westphal, M. (2021). Question tags across New Englishes. *World Englishes*, 40(4), 519–533.
- Westphal, M. (2022). The multilingual pragmatics of New Englishes: An analysis of question tags in Nigerian English. *Frontiers in Communication*, 6, 1–16.

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.