Question Tags in Philippine English

Michael Westphal¹

Received: 12 November 2019 / Accepted: 5 February 2020 / Published online: 11 February 2020
© The Author(s) 2020

Abstract

This study investigates the use of question tags (QTs) in a subcorpus of dialogues from the Philippine component of the International Corpus of English. It takes into account the full range of QT forms used in Philippine English, including English variant QTs as well as English and Tagalog invariant forms. The analysis investigates the effects of text type and pragmatic function on the selection of particular forms. The results show that Filipino speakers use English and Tagalog forms to almost equal proportions, but invariant forms dominate by far over variant ones. Text type has a strong effect on the overall frequency of QTs and on the distribution of individual forms. In addition, function is shown to be a significant factor that influences QT use: speakers preferentially use specific QTs over others for particular functions in specific contexts. The results show that it is beneficial to analyze the full range of QTs to describe the characteristics of Philippine English and to illustrate variety-internal variation. Furthermore, the analysis of English and Tagalog QTs shows that variation in discourse-pragmatic features can provide valuable insights into language contact situations. In conclusion, the study highlights the benefits of small diverse corpora for corpus-pragmatic research as they allow studying pragmatic phenomena in a range of different contexts.

Keywords Question tags · Philippine English · International Corpus of English · Text type variation

Introduction

The Philippines is a highly multilingual archipelago in South East Asia: according to Ethnologue (Eberhard et al. 2019), 183 living languages coexist in the country. The two principal languages are English, introduced to the Philippines via American colonization starting in 1889, and Tagalog (also called Filipino), the country's national language. Tagalog is primarily spoken in the largest and most populous

✉ Michael Westphal
michael.westphal@wwu.de

¹ English Seminar, University of Münster (NRW), Münster, Germany

island Luzon, which is also home to the country's political and economic capital Manila. In Metro Manila, the combination of English and Tagalog is common in many private and public domains (Thompson 2003). Philippine English (PhiE), the country's emerging standard variety of English, is strongly influenced by this multilingual embedding (Tupas 2004). Most previous corpus-based descriptions of PhiE have focused on morpho-syntax (Lim and Borlongan 2012) but first corpus-pragmatic studies (Bautista 2011a; Lim and Borlongan 2011) have noted the high presence of Tagalog discourse-pragmatic features in PhiE. Thus, a pragmatic perspective seems promising to assess multilingual variation in PhiE.

This paper analyzes question tags (QT), i.e. a specific set of discourse-pragmatic features, in the Philippine component of the International Corpus of English (ICE-PHI), which aims to depict Standard English in the Philippines and in particular Metro Manila (Bautista 2011b). I do not analyze the use of QTs in the entire corpus but focus on seven dialogical text types from ICE-PHI: conversations, phonecalls, classroom lessons, legal cross-examinations, broadcast interviews, broadcast discussions, and parliamentary debates. The analysis takes into account the full range of QT forms, including so-called variant QTs, such as *isn't it* or *don't you*, and invariant ones, such as *right* or *eh*. In PhiE, the repository of invariant QTs includes English-based forms as well as Tagalog forms, such as *di ba* or *ha*. Apart from describing the variation among these forms, this study shows that the selection of one form over the others is not only a matter of individual preferences but is constrained by the context (operationalized as text type) and the pragmatic function speakers want to achieve. This analysis contributes to the description of PhiE and advocates an inclusion of pragmatic phenomena in descriptions of varieties of English (e.g. Kortmann and Schneider 2008). The analysis also provides a pragmatic perspective on language contact situations by analyzing English and Tagalog QT forms. In particular, the study addresses the following research questions:

- Which QT forms are used in PhiE as represented by ICE-PHI?
- How does text type constrain the distribution of individual QT forms?
- What is the form-function relationship of the most frequent QT forms?
- Which forms do speakers select in specific contexts for a particular function?

The rest of the paper is structured as follows: The section “[Previous Research on Question Tags](#)” discusses previous research on QTs in different varieties of English. In “[Data and Method](#)”, the data and method used in the current analysis are presented. The section “[Descriptive Statistics on QT Use in ICE-PHI](#)” illustrates the results on the variation of QT use in ICE-PHI. The methodological implications of the results are discussed in the section “[Inferential Statistics on QT Use](#)”.

Previous Research on Question Tags

This study analyzes QTs as one set of discourse-pragmatic features, which is defined by certain formal but mainly functional properties. Pichler (2013: 4–6) defines discourse-pragmatic features as formally heterogeneous and syntactically optional

elements that are used by speakers to express stance, to guide utterance interpretation, or to structure the discourse (see also Schiffrin 1987: 1–30). Pichler (2013) uses the term discourse-pragmatic-feature as an umbrella term for pragmatic markers and discourse markers, which are often treated as different categories but are difficult to delimit (see also Fedriani and Sansò 2017). As discourse-pragmatic features, QTs are mainly defined on functional-pragmatic grounds but there are certain formal characteristics, which have been central to most previous descriptions and analyses of QTs.

Most English grammar books (e.g. Biber et al. 1999: 208–210) focus on variant QTs, whose structure depends on the adjacent main clause. Variant QTs consist of an auxiliary verb and a pronoun (Biber et al. 1999: 208) and the auxiliary verb is identical to the auxiliary in the main clause (excerpt 1); or if there is no auxiliary in the main clause *do* is used (excerpt 2). The auxiliary also agrees with the tense, aspect, and mood of the verb in the main clause (excerpts 1 and 2). However, variant QTs are often not strictly modeled after the clause they are attached to: very commonly, speakers use *isn't it* 'invariantly' as in excerpt (3).

- (1) <#>Maybe at this point you have made a decision already *have you* (S1B-014)¹
- (2) <#>The reason why I can understand stuff like this like Cosmopolitan magazine heck they they provide jobs *don't they* (S1A-068)
- (3) <#>You can easily go to McDonald and eat your yourself away but that will not give you peace *isn't it* (S1B-005)

In contrast to variant QTs, invariant QTs have a fixed form and are often only discussed in passing in grammar books (e.g. Biber et al. 1999: 1089). Invariant QTs include single words, such as *right*, multi-word units, such as *you know*, or phonological sequences, such as *uh* (excerpts 4–6). PhiE invariant QTs also include Tagalog forms, such as *di ba* (excerpt 7). Some invariant QTs have become shibboleths of a particular variety of English, for example *eh* for Canadian English (Gold and Tremblay 2006) or *innit* for Multicultural London English (Pichler 2016), or have heightened indexical loading: for example, in Trinidadian English *not so* is considered a polite alternative to Creole QTs (Wilson et al. 2017: 735).

- (4) <#>Claude Van Damme movies are considered A films *right* (S1B-037)
- (5) <#>Well it's not just about Manila *you know* (S1A-096)
- (6) <#>You guys kind of agree *uh* (S1B-035)
- (7) <#>Did you see it *di ba* (S1A-004)

Most previous corpus-based research on QTs has focused on variant ones with a strong focus on their exact formal realization. For example, Tottie and

¹ All examples are from ICE-PHI and text codes are given in brackets. The ICE markup is used in the excerpts: <#> marks the beginning of an utterance; <\$A> marks speaker turns; <[> and </(> mark overlaps; <&> and </&> mark comments.

Hoffmann (2006) compare British and American English using the BNC and the Longman Spoken American Corpus, Barron et al. (2015) study variant QTs in ICE-GB and ICE-Ireland, Axelsson (2018) analyzes variant QTs in the BNC2014, and Kimps (2018) investigates their use in ICE-GB, the London Lund Corpus, and the Bergen Corpus of London Teenage Language. Beyond this wide body of research on English varieties spoken as a native language (ENL), several studies have also investigated variant QTs in countries where English functions as a second language (ESL) as is the case in the Philippines. For example, Wong (2007) analyzes variant QTs in Hong Kong, Borlongan (2008) in the Philippines, and Parviainen (2016) as well as Takahashi (2016) in four Asian Englishes, including PhiE. All four studies use the ICE corpora, compare their findings to variant QT use in British, American, or—in Takahashi’s (2016) study—Canadian English, and highlight that the Asian Englishes under investigation are characterized by a high frequency of invariant uses of *isn’t it* and/or *is it (not)*.

Previous studies on invariant QTs have often focused on individual forms in a particular variety: for example, *innit* in Multicultural London English (Pichler 2016) or *na* in Indian English (Lange 2012: 195–234). In contrast, Columbus (2009, 2010) studies a wide range of invariant QT forms across British, New Zealand, Indian, Singapore, and Hong Kong English using the ICE corpora. She shows a high frequency of invariant QTs in all five varieties and substantial regional variation with regard to individual forms. All of these studies on variant and invariant QTs have focused on conversations, while other text types are underresearched. Exceptions include Barron (2015), who studies variant QTs in service encounters and highlights genre-specific uses of QTs. Furthermore, the studies by González (2018) and Wilson et al. (2017) analyze both variant and invariant QTs in a range of different spoken text types: González (2018) shows that variant QTs are more frequent than invariant ones (717 vs. 141 occurrences) in the spoken component of ICE-GB, and she highlights functional differences between the two sets of QTs. In contrast, Wilson et al. (2017) demonstrate that variant QTs are marginal in contrast to invariant ones (11 vs. 1015 occurrences) in spoken Trinidadian English. Both studies emphasize the importance of text type for QT use: González (2018: 138) concludes that “discourse setting and speakers’ roles, rather than their gender and/or age seem to be determinant factors for the choice” of QTs, and Wilson et al. (2017: 740) state that the communicative setting of a text type constrains QT frequency, form, and function.

Previous research on Tagalog particles in ICE-PHI illustrates their high frequency and highlights considerable variation across text types and their functional diversity. Bautista (2011a) demonstrates an extremely high frequency (788 occurrences) of the Tagalog particle *’no*, which can function as a QT, in the spoken component of ICE-PHI. She shows substantial variation with regard to text type: *’no* is very frequent in classroom lessons and unscripted speeches, while it is (almost) absent from very formal text types, such as legal cross-examinations, or parliamentary debates; private dialogues take a middle position. Lim and Borlongan (2011) describe the use of the Tagalog particles *na*, *pa*, *ba*, and *’no* in ICE-PHI. The latter two particles function as QTs. Whereas *ba* mainly functions in an informative way as a yes–no question, the functions of *’no* are more diverse: it mainly expresses attitudinal (or emphatic)

and facilitative (i.e. integrating the listener into the conversation) functions but is also used to ask for confirmation (Lim and Borlongan 2011: 61–62, 70–74).

This brief discussion of corpus-pragmatic studies on QTs has shown that there is a strong focus on variant QTs and a dearth of studies that analyze both variant and invariant ones. There is also the need to study QTs in text types other than only conversations. While there is a growing body of research on QTs in ESL varieties, the vast majority of studies has analyzed ENL varieties. Furthermore, previous research on PhiE has shown a high productivity of Tagalog particles in ICE-PHI but has mainly investigated them in isolation. Thus, a study of variant and invariant QTs in PhiE, which includes Tagalog and English-based forms, is suitable to address these research desiderata.

Data and Method

ICE-PHI is one component of the ICE project (Greenbaum 1996), which aims to represent Standard Englishes in ENL and ESL contexts. Standard English in the ICE project is defined via the speakers: they have to be 18 years or older and have to have completed secondary education in English (Greenbaum 1996: 6). The speakers in ICE-PHI fulfill these requirements but there is a strong bias towards speakers from Metro Manila, and especially De La Salle University (Bautista 2011b), a highly prestigious university in Manila. In addition, 61.4% of speakers who stated their L1 named Tagalog. Thus, ICE-PHI does not depict Standard English spoken in the entire Philippines but rather Standard English spoken by highly educated (or elite) speakers in Metro Manila whose L1 tends to be Tagalog. In Metro Manila, most private or public dialogues are dominated by Tagalog or code-switching between English and Tagalog (Thompson 2003; Bautista 2011b: 7–12). The ICE-PHI team tried to select texts which are almost exclusively in English with little code-switching to Tagalog, which is nevertheless quite frequent in many text types.

With a total size of 1,000,000 words each, the ICE corpora are by today's standards relatively small but have a diverse design (Nelson 1996), covering 300 spoken and 200 written texts (2000 words each) from overall 28 (15 spoken; 13 written) different text types, encompassing formal (e.g. legal cross-examinations or academic writing) and informal situations (e.g. conversations or personal letters). This study makes use of this diverse design and focusses on seven spoken and unscripted dialogic text types, which differ in their communicative setting and level of formality: conversations, phonecalls, classroom lessons, broadcast discussions, broadcast interviews, legal cross-examinations, and parliamentary debates. The only dialogues from ICE-PHI not included in the analysis are business transactions, as texts from this text type are not coherent in terms of the communicative setting and the level of formality: for example, they include doctor-patient interactions, informal conversations in a private business, and a university staff meeting (see Bautista 2011b: 10). See Table 1 for an overview of the data.

Both conversations and phonecalls are informal private dialogues and there are no predefined speaker roles. In contrast to conversations, phonecalls are long-distance conversations of two people restricted to oral communication. In the statistical

Table 1 ICE-PHI dialogues

Text type	Text codes	Texts	~ Size (words)
Private dialogues			
Conversations	S1A-001 to S1A-090	45 ^a	90,000
Phonecalls	S1A-091 to S1A-100	10	20,000
Public dialogues			
Classroom lessons	S1B-001 to S1B-020	20	40,000
Broadcast discussions	S1B-021 to S1B-040	20	40,000
Broadcast interviews	S1B-041 to S1B-050	10	20,000
Parliamentary debates	S1B-051 to S1B-060	10	20,000
Legal cross-examinations	S1B-061 to S1B-070	10	20,000

^aOnly a random selection of 45 conversations is included in the analysis

analysis, they are conflated to private dialogues. The other five text types are all public dialogues and have a higher level of formality—legal cross-examinations and parliamentary debates being the most formal. Broadcast talks and interviews vary substantially with regard to the level of formality as some texts are about formal topics, such as politics or finances, while others are about entertainment (Bautista 2011b: 8–9). Furthermore, speakers have fixed roles in the five public text types: in ICE-PHI, classroom lessons are largely monologues (Bautista 2011b: 8) where a university lecturer gives a talk and wants their students to follow the line of argumentation. In broadcast interviews, a host asks questions to the interviewee, and in broadcast discussions, a host guides a discussion between several people. In the statistical analysis, the latter two text types are merged to broadcasts. In legal cross-examinations, an attorney questions a witness. In parliamentary debates, politicians discuss various issues, and the discussion is led by the speaker of the house. Thus, lecturers, hosts, attorneys, and the speaker have a certain control of the conversation.

Similar to most corpus-pragmatic phenomena, there is a form-function mismatch for QTs, which requires a close analysis of each token (Aijmer and Rühlemann 2015: 9–13). A top-down concordance search for a range of forms that potentially function as QTs generates many tokens which do not function as QTs, and other forms not included in this search are overlooked. To overcome this problem, the current study takes a corpus-driven approach (Anderson 2016) as the researcher read through all 125 texts and identified all QT tokens in context. The classification of a specific form as a QT is based on the following criteria: QTs are discourse-pragmatic features and are attached to utterances. QTs are neither fillers (i.e. forms surrounded by repetitions or fillers are excluded), nor entire utterances on their own (e.g. *right* used as a backchannel), nor items used in their full literal sense (e.g. the right choice), nor part of fixed expressions (e.g. right now). Crucially, QTs fulfill an informative, punctuational, and/or facilitative function. This three-way distinction of the pragmatic functions of QTs builds on previous classifications of the pragmatic functions of QTs summarized by Tottie and Hoffmann (2006: 297–301): they distinguish between confirmatory, facilitating, attitudinal, informational, peremptory, and aggressive tags. A clear distinction between confirmatory and informational

functions as well as between attitudinal, peremptory, and aggressive functions often proved impossible for the PhiE data. Furthermore, in Tottie and Hoffmann's (2006) analysis, confirmatory, facilitating, and attitudinal functions accounted for over 90% of the total QTs. Thus, the current functional distinction (i.e. informative, punctuational, facilitative) correspond to these three main functional types but there is no distinction between confirmatory and informational functions (i.e. summarized as informative) and between attitudinal, peremptory, and aggressive types (summarized as punctuational). The three-way distinction is thus less fine-grained than previous functional classifications (Tottie and Hoffmann 2006; also Columbus 2010; Wilson et al. 2017), which allows using function as a variable in statistical modeling.

Speakers use QTs with an informative function when they are unsure of the content of the utterance, when they request information or a confirmation, and when an answer is expected. For example, in excerpt (8) the speaker adds *'no* to find out whether Cara is really the bridesmaid. QTs with a punctuational function are used for stylistic purposes, mainly to add emphasis. Speakers are sure about the content of the utterance and no answer is expected. For example, in excerpt (9) the speaker adds *will you* to the directive for emphasis. Speakers use QTs in a facilitative way to integrate their listeners more closely into the discourse either by signaling that they are willing to hand over their turn or to invite (verbal or non-verbal) backchanneling. For example, in excerpt (10) the lecturer uses *OK* to check whether the students have understood the concept of a correlation coefficient and gives them the chance to intervene. A clear distinction between these functions is sometimes difficult as many QTs are multifunctional. For example, in excerpt (11) the speaker uses *di ba* to emphasize her empathy for the problems of her friend Kris but also to invite her interlocutors to backchannel or to comment on the issue.

(8) <#>Cara's your bridesmaid *'no* (S1A-070)

(9) <#>Victor cool it *will you* (S1A-088)

(10) <#>Now the size of a correlation coefficient ranges from negative one to positive one *OK* (S1B-009)

(11) <#>But in Kris' case I think she's I don't know uh <,> she's very intelligent <#>It's such a waste *di ba* (S1A-006)

Each QT was coded for its function based on these criteria. If a token did not fulfill one or two of these functions it was not included in the analysis. The codings were checked by a research assistant who is a speaker of PhiE and Tagalog. The decision whether a QT form is Tagalog or not is based on the marking of all Tagalog words and utterances by the Filipino transcribers in ICE-PHI as 'indigenous' (<indig>). This qualitative identification and coding for function of each QT token is the basis for the quantitative analysis, which gives an overview of the QT forms and their frequencies in the subcorpus of ICE-PHI and analyzes the distribution across text types and the form-function relationship of the most frequent QT forms. For the descriptive statistics on frequencies of QTs across different text types, the results are normalized to tokens per 50,000 words due to the unequal sizes of

Table 2 QT forms in ICE-PHI

English QTs		Tagalog QTs	
Form	Frequency (%)	Form	Frequency (%)
<i>alright</i>	13 (1.0)	<i>ano</i>	18 (1.4)
<i>OK</i>	246 (18.5)	<i>ba</i>	11 (0.8)
<i>right</i>	206 (15.5)	<i>hindi/ba</i>	94 (7.1)
<i>uh</i>	28 (2.1)	<i>e</i>	77 (5.8)
<i>yeah</i>	16 (1.2)	<i>ha</i>	55 (4.1)
<i>you know</i>	133 (10.0)	<i>'no</i>	353 (26.6)
Other invariant QTs ^a	26 (2.0)	<i>oo</i>	6 (0.5)
Variant QTs	45 (3.4)		

^aEnglish invariant QTs with a frequency below 10 (*correct*, *yes*, *or no*, *or not*, *or what*)

the four text types. The effects of text type and function on the selection of one QT form over the others is analyzed in detail via binary logistic regression modeling: form is the dependent variable, which is reduced to binary distinctions, such as *OK* (as the application value) versus all other forms; text type (private dialogues vs. classroom lessons vs. broadcast dialogues vs. legal cross-examinations) and function (facilitative vs. informative vs. punctuational) are fixed predictor variables; speaker is included as a random factor (random intercept) in order to reduce the risk that individual speaker preferences bias the results with regard to fixed group effects (Johnson 2009: 363–365). Multifunctional QTs are excluded from the statistical analysis and if there is a categorical absence of a QT form in a particular text type, this text type is excluded from the regression model. Statistical analyses were carried out with Rbrul (Johnson 2009). Rbrul reports whether a predictor variable has a significant effect on the distribution via *p*-values and the direction and effect size of the individual levels of a predictor variable via centered-factor weights, which range from 0 to 1. Values above 0.5 indicate a preference for the application value and values below 0.5 a dispreference. In a last step, I provide a qualitative perspective on QTs, which aims at highlighting how speakers employ certain QT forms in particular situations for specific functions.

Descriptive Statistics on QT Use in ICE-PHI

The speakers in ICE-PHI use a wide range of English and Tagalog QT forms. Table 2 gives an overview of the different QT forms as raw token frequencies and the proportion of the total QT count. Of the overall 1327 QTs, 713 (53.7%) are English forms and 614 (46.3%) are derived from Tagalog. *'No* is the most frequent QT, followed by *OK*, *right*, and *you know*; all other forms have a frequency below 100 occurrences. With only 45 occurrences, variant QTs play a marginal role for QT use in PhiE. Of the variant QTs, 19 are used invariantly, which means that the auxiliary does not agree with the verb in the main clause. The most frequent variant QT forms

Fig. 1 Normalized QT frequencies in the six text types

used invariantly are *isn't it* (17), *is it* (6), and *is it not* (6). One speaker alone is responsible for all six occurrences of the latter form, which means that *is it not* is not an innovative form of PhiE but an idiolectal peculiarity.

Text type has a strong effect on the overall QT frequency (Fig. 1): most QTs can be found in classroom lessons (511; norm: 638.8), followed by phonecalls (122; norm: 305), conversations (454; norm: 252.2), broadcast discussions (168; norm: 210), broadcast interviews (45; norm: 112.5), and legal cross-examinations (25; norm: 62.5). With only two occurrences (norm: 5), QTs are extremely rare in parliamentary debates and hence this text type and the two tokens are excluded from further analysis. Parliamentary debates are very formal and the speaker closely controls the dialogue. The politicians prepare their arguments and present them in a highly structured way. In cases where speakers seek a confirmation, they use very formal constructions, such as “Is that correct Mr. President” (S1B-056). Thus, speakers generally do not use QTs—or discourse-pragmatic features in general—to organize these political discussions. This is in contrast to classroom lessons, where lecturers use a high frequency of QTs as an educational strategy. Hosts in broadcast discussions and interviews also use QTs as a strategy to organize this public discourse. Attorneys use QTs rather rarely to elicit information from witnesses and largely rely on other questioning strategies. QTs are more frequent in phonecalls than in conversations: the difference in frequencies can be explained with the exclusive reliance on verbal tools to structure dialogues in phonecalls.

The subordinate role of variant QT forms is evident in all text types. Legal cross-examinations seem to deviate slightly from this overall trend as seven out of the 25 tokens are variant QTs. However, as six of them are used by one person alone this seeming exception is due to individual speaker preferences. In contrast to variant QTs, Tagalog QTs are very frequent in all text types and the fairly balanced ratio of English and Tagalog QT forms is relatively stable across all text types. Thus, Tagalog QTs are integrated into PhiE in all text types analyzed in this study, which range from informal conversations to formal legal cross-examinations. The two QT occurrences in parliamentary debates are *'no* and *ha*; this suggests that Tagalog QTs are even acceptable in this highly formal political discourse.

While the overall distribution of Tagalog and English QTs is quite stable across the different text types, a closer analysis of the distribution of individual QTs reveals

Fig. 2 Tagalog QTs in the six text types

that most Tagalog forms are used more frequently in informal private dialogues than in more formal public ones. Figure 2 shows the text type distribution of the most frequent Tagalog QTs as percentages by text type based on the normalized frequencies of each QT to account for the unequal text type sizes. The descriptive statistics show that *ano*, *ba*, *hindildi ba*, *e*, and *ha* are predominantly used in informal private dialogues. *Ano*, *ba*, *hindildi ba*, and *ha* are categorically absent from legal-cross examinations and *e* is used only once in this highly formal text type. However, these Tagalog QTs are not absent from public conversations: a considerable proportion of *ano*, *ba*, *hindildi ba*, *e*, and *ha* can be found in broadcast discussions and interviews. In addition, there are 12 (norm: 15) occurrences of *hindildi ba* in classroom lessons, which corresponds to a share of 19.4% of all normalized *hindildi ba* tokens. In contrast to these five Tagalog QTs that show a preference for informal private dialogues, *'no* is used frequently by speakers in all text types: there are 34 tokens (norm: 18.9) in conversations, 15 (norm: 37.5) in phonecalls, 213 (norm: 266.3) in classroom lessons, 70 (norm: 87.5) in broadcast discussions, 18 (norm: 45) in broadcast interviews, and even two (norm: 5) in legal cross-examinations. Despite this text type versatility, the largest share of *'no* tokens in the ICE-PHI subcorpus is produced by lecturers in classroom lessons. Bautista (2011b: 81) describes *'no* as a “verbal tick” among some speakers, which is also evident in the analysis of *'no* as a QT: for example, the lecturer in S1B-001B produces *'no* 61 times as a QT and the lecturer in S1B-008 57 times. However, it is not an idiolectal feature of a few Filipino lecturers: *'no* occurs in 12 out of 20 classroom lessons. The second highest proportion of *'no* as a QT can be found in broadcast discussions where it is used by hosts as well as guests. Thus, *'no* seems to be a special case of the Tagalog QTs due to the high proportion among lecturers and as it transcends across all text types.

Overall, the most frequent English QTs (*OK*, *right*, variant QTs, and *you know*) seem to be spread out across the six text types more evenly than the six Tagalog QTs (see Fig. 3). With the exception of *right* being absent from legal cross-examinations, *right* and *you know* are used by speakers in all text types fairly equally. Nevertheless, private dialogues and classroom lessons show the overall highest proportions of *right* and *you know*, with more than 20% of the normalized token counts. For variant QTs, the largest share can be found in legal cross-examinations, but this result is

Fig. 3 English QTs in the six text types

Table 3 Relative entropy of QT functions

<i>hindi/di ba</i>	<i>right</i>	<i>ano</i>	<i>ba</i>	Variant QT	<i>ha</i>	<i>you know</i>	<i>'no</i>	<i>OK</i>	<i>e</i>
0.96	0.95	0.94	0.94	0.90	0.77	0.73	0.66	0.48	0.37

strongly influenced by individual speaker preferences as six of the seven variant QTs are used by one attorney. *OK* is the only English QT where text type has a distinct effect on the distribution: lectures alone use *OK* 201 times, which corresponds to 86.6% of the normalized *OK* count.

The functional diversity of QTs is operationalized via three categories in this study: facilitative, informative, and punctuational QTs. Overall, 658 (49.6%) QTs are facilitative, 427 (32.1%) punctuational, and 176 (13.3%) informative. In addition, 65 (5.0%) QTs fulfill two functions: 63 multifunctional QTs are facilitative and punctuational, one token serves an informative and punctuational functional, and one QT functions in a facilitative and informative way. These multifunctional QTs are excluded from further analysis.

The analysis of the functional diversity of the most frequent individual QTs shows that there is no one-to-one form function mapping: except for *e*, which is categorically not used in an informative way, each QT form analyzed can fulfill each function. However, there are preferential differences in the form-function mapping and the dispersion of the three functions varies between the ten QT forms. The dispersion is described as relative entropy (H_{rel}), which describes the average amount of uncertainty and varies from zero to one (Gries 2010: 8–9). The larger the H_{rel} the more random a distribution is and the closer to zero the more focused it is. Table 3 gives an overview of the H_{rel} values sorted from highest to lowest from left to right. Figure 4 shows the ratio of functions in percentages for each of the 10 most frequent QT forms. These two descriptive statistics together provide a good account of the form-function relationship.

The dispersion of *hindi/di ba*, *right*, and *ba* is above 0.90 and the ratio between all three functions is fairly even as no functional share is above 50%. Variant QTs are slightly more focused with a preference for informative functions (54.8%) over the other two functions. *Ha*, *you know*, and *e* are predominantly used in a punctuational way, with *e* showing the lowest functional dispersion. *'No* and *OK* also have a strongly focused functional profile with a substantial preference for facilitative uses. Thus, speakers prefer certain QTs over others for specific functions and specific QTs

Fig. 4 Functions of the most frequent QTs (multifunctional QTs are excluded)

have a more focused functional profile, such as *OK* and *e*, while others are more variable in their functional use, such as *hindi/di ba*. The overall frequency of a QT does not seem to play a role for the form-function relationship. Highly frequent forms, such as *right*, as well as low frequency forms, such as *ano* and *ba*, show a high functional variability, whereas highly frequent *'no* and *OK* as well as the less frequent *e* have a rather focused functional profile.

These descriptive statistics show that both text type and function have an effect on the use of specific QT forms. However, these results are potentially biased due to low token frequencies. Especially infrequent QTs, such as *ano* and *ba*, are more strongly influenced by individual speaker preferences. This bias tends to be more problematic in text types with a low number of texts and thus a lower raw token frequency, such as in legal cross-examinations. In order to account for idiosyncratic variation, the regression models include speaker as a random factor and thus aim to substantiate the variation in the sample, depicted in this descriptive part.

Inferential Statistics on QT Use

The regression models, which investigate the effects of text type² and function on the use of a particular QT form in contrast to all other forms, show that most QTs, except for *hindi/di ba*, have a specific functional profile. Text type has a significant effect on the use of individual QT forms, except for *e*, *ha*, and variant QTs. However, text type variation is vulnerable to low token frequencies in some text types. For *e* (Table 4) and *ha* (Table 5) there is a strong preference for punctuational uses, while facilitative ones are dispreferred. For *e* there is a categorical absence of informative

² Broadcast discussions and interviews are merged to broadcast; conversations and phonecalls are joined to private dialogues.

Table 4 *e* ($N=72$)

	$R^2=0.77$	N	%- <i>e</i>	FW
Function ^a ; $p < .001$	PUNC	420	14.5	0.74
	FACI	655	1.5	0.26
Text type; $p = .007$				
^a INFO excluded				

Table 5 *ha* ($N=53$)

	$R^2=0.91$	N	%- <i>ha</i>	FW
Function; $p < .001$	PUNC	420	8.6	0.73
	INFO	155	4.1	0.41
	FACI	655	1.5	0.35
Text type ^a $p = .296$				
^a Legal cross-examinations excluded				

Table 6 *hindi/di ba* ($N=91$)

	$R^2=0.54$	N	%- <i>hindi/di ba</i>	FW
Text type; $p = .002$	Private dialogue	545	12.1	0.71
	Broadcast	212	5.7	0.40
	Classroom lesson	473	2.5	0.38
Function ^a $p = .411$				
^a Legal cross-examinations excluded				

functions and for *ha* this function is dispreferred. Text type does not have a significant effect on the use of both QTs but the models are somewhat biased: there is a categorical absence of *ha* in legal cross-examinations. In addition, one of the 25 QTs in legal cross-examinations happens to be *e*, which then has a strong effect on the model due to the low QT frequency in this text type.

For *hindi/di ba* (Table 6) only text type has a significant effect: there is a preference for *hindi/di ba* in private dialogues, categorical absence in legal cross-examinations, and a dispreference in the remaining public text types. Function does not have a significant effect on the use of *hindi/di ba* in contrast to the other QT forms. This finding corroborates the functional flexibility of *hindi/di ba* described with the high H_{rel} .

'No' (Table 7) and OK (Table 8) have very similar functional profiles: there is a preference for both forms for facilitative uses, a slight preference for punctuational functions, and a dispreference for informative uses. With regard to text type, there is a preference for both QTs in classroom lessons and a dispreference in private dialogues. However, the effects of legal cross-examinations and broadcast interviews and discussions differ: there is a preference for 'no' and a dispreference for OK in broadcast interviews and discussions. For legal cross-examinations, there is dispreference for 'no' but no clear effect for OK.

Table 7 'no' ($N=333$)

	$R^2=0.55$	N	%-'no'	FW
Text type; $p < .001$	Broadcast	212	41.5	0.80
	Classroom lesson	473	41.4	0.64
	Legal cross-ex	25	8.0	0.34
	Private dialogue	545	8.4	0.22
Function; $p = .002$	FACI	657	35.9	0.63
	PUNC	425	19.3	0.55
	INFO	173	8.1	0.32

Table 8 OK ($N=228$)

	$R^2=0.67$	N	%-OK	FW
Text type $p < .001$	Classroom lesson	473	39.3	0.90
	Legal cross-ex	25	4.0	0.48
	Private dialogue	545	5.9	0.28
	Broadcast	212	3.8	0.21
Function $p < .001$	FACI	657	27.9	0.74
	PUNC	425	9.9	0.65
	INFO	173	1.2	0.16

Table 9 right ($N=200$)

	$R^2=0.62$	N	%-right	FW
Text type; $p < .001$	Private dialogue	545	23.7	0.75
	Classroom lesson	473	9.7	0.40
	Broadcast	212	11.3	0.34
Function ^a ; $p < .001$	INFO	155	45.8	0.83
	FACI	655	13.4	0.42
	PUNC	420	9.5	0.22

^aLegal cross-examinations excluded

Table 10 variant QTs ($N=42$)

	$R^2=0.80$	N	%-variant QT	FW
Function; $p < .001$	INFO	173	12.7	0.78
	FACI	657	1.8	0.40
	PUNC	425	1.6	0.29
Text type; $p = .57$				

Right (Table 9) and variant QTs (Table 10) have a similar functional profile in ICE-PHI: there is a preference for informative uses, while facilitative and punctuational functions are dispreferred, the latter more strongly than the former. Text

Table 11 *you know* ($N = 128$)

	$R^2 = 0.56$	N	%-you know	FW
Text type; $p < .001$	Legal cross-ex	25	12.0	0.80
	Private dialogue	545	16.3	0.62
	Broadcast	212	12.3	0.49
	Classroom lesson	473	2.1	0.15
Function; $p < .001$	PUNC	425	18.1	0.76
	FACI	657	7.0	0.58
	INFO	173	2.9	0.19

type only has a significant effect for *right*: it is used preferentially over other QT forms in private dialogues and there is a dispreference for all other text types, as *right* is categorically absent from legal cross-examinations.

The functional profile of *you know* (Table 11) contrasts with that of *right* and variant QTs: punctuational uses are strongly preferred for *you know* and facilitative ones only slightly, but informative functions strongly disprefer *you know*. With regard to text type, there is a preference for *you know* in legal cross-examinations and private dialogues, no clear effect for broadcast discussions and interviews, as well as a dispreference in classroom lessons.

Qualitative Insights on QT Use

The quantitative analysis of the most frequent QTs in ICE-PHI shows preferential differences in the functional profiles of the individual QTs and their distribution across the different text types. However, while the operationalization of the functional diversity of QTs as facilitative, informative, and punctuational allows for a quantitative analysis, it nevertheless works on a high level of abstraction. Furthermore, the text type analysis generalizes all speakers in one text type and does not show how speakers with different roles use QTs. This last results section provides a qualitative perspective on QT use in ICE-PHI in order to highlight their functional complexity and the importance of specific speaker roles for QT use.

Out of the dialogues from ICE-PHI, QTs have the highest frequency in classroom lessons. The lecturers in this text type use QTs very frequently as an educational strategy. All texts are lecture-type dialogues with long monologic stretches by the lecturers. During these long turns, the lecturers want their students to follow along. To ensure this, the lecturers frequently use 'no and OK in a facilitative way. For example, in excerpt (12), a lecturer explains the benefits of an entrepreneurial society in a business lecture.

- (12) <#>That will be made possible if we have more and more businesses being set up and from which taxes will be paid to the government *OK* <#>So the government can only produce such uh activities which are supposed to be of social benefits to us '*no* (S1B-004)

These QTs function as quasi-rhetorical questions checking whether the content of the utterance is clear. Students do not backchannel verbally but they most likely express their continued efforts to follow along by visual cues, such as looking at the lecture attentively or nodding. Hosts use facilitative QTs similarly in broadcast discussions and interviews during longer turns, which typically introduce a topic or summarize arguments, to include their interlocutors into the conversation. For example, in excerpt (13), the host introduces the topic of the interview and inserts '*no* to integrate the interview guests.

- (13) <#>And I'd like to talk about that because it has everything to do with economic policy and how it affects the people particularly those that are less uh less privileged in life '*no* (S1B-044)

In private dialogues, speakers also at times use QTs to check whether their interlocutors are still following them in longer turns, but here the listeners often backchannel and many facilitative tags in conversations and phonecalls also signal that the speaker is handing over the turn. For example, in excerpt (14) speaker A does not necessarily seek a confirmation as she is not unsure about the content of her utterance but tries to integrate the other person into the conversation by adding *di ba*.

- (14) <\$A><#>I mean you can't say when someone falls in love *di ba*
<\$B><#>Uh huh <#>That's true yeah well that's true <&>laughter</&> (S1A-006)

In ICE-PHI, attorneys never use QTs in facilitative way. In the rare cases that they use QTs, they do so in leading questions, to prompt a witness to confirm or disclose information as in excerpt (15), where the witness has no real choice but to confirm that they would not engage in any illegal activities.

- (15) <\$A><#>You would not engage in anything that is illegal *is it not*
<\$B><#>No (S1B-069)

These informative uses of QTs also have an antagonistic meaning and attorneys use them to put the witnesses under pressure and reinforce their position of authority; or put differently: attorneys use QTs to 'do power'.

Apart from legal cross-examinations, antagonistic uses only appear in private dialogues but are extremely rare. Excerpt (16) shows part of a discussion about cell phones, which becomes rather heated as the two participants disagree. Speaker B interrupts A and uses *OK* to emphasize her disagreement. Thus, *OK* is a punctuational QT with an additional antagonistic meaning.

- (16) <\$A><#>We only need <[/>dual or single</[/> band
 <\$B><#><[/>Thirty-two fifty is not</[/> <#>Eighty-two fifty is not tri-band it's just dual band
 OK <#>So the only tri-band phone here in the Manila right now is the eighty-eight nine TM this
 Motorola thing (S1A-041)

In contrast to legal cross-examinations, speakers use informative QTs in private dialogues and classroom lessons in cases of true uncertainty to receive a confirmation or to acquire new information. The speakers in excerpts (17) and (18) genuinely seek information from their interlocutors.

- (17) <#>You have a daughter and a son *right* (S1A-002)
 (18) <#>You're already I believe of voting age *are you not* (S1A-093)

Hosts in broadcast discussions and interviews use QTs in an informative way as a moderation strategy. They are not truly uncertain about the content of the utterance but use informative QTs as a questioning strategy and to focus the discussion on particular topics. For example, in excerpt (19), the host A focusses the interview on the family drama of her interview guest B.

- (19) <\$A><#>Uh I understand that for example [...] your ex uh has been trying to to contact uh your
 daughter '*no*
 <\$B><#>Well (S1B-046)

In contrast to hosts, who mainly use facilitative and informative QTs, the guests in broadcast dialogues mainly use punctuational QTs to emphasize their arguments, particularly in broadcast discussions. In excerpt (20), a politician uses *e* to add emphasis to his contribution to the discussion about the role of the Filipino Senate on a particular political matter.

- (20) <\$B><#>Well yes dahil ('because') it is really the fundamental question *e* <#>Are we a voice or
 are we an echo (S1B-024)

Speakers in private dialogues also frequently use QTs in a punctuational way for emphasis. For this function, they mostly rely on *e*, *ha*, and *you know*. Punctuational QTs are also added to imperatives, which are rare and mainly occur in private dialogues (excerpt 21).

- (21) <\$A><#>Hold it *ha* (S1A-091)
 <\$B><#>Sure

However, the punctuational QTs added to imperatives are mostly not antagonistic but also help to soften the face-threatening act of a demand. Similarly, speakers frequently use punctuational QTs to add a humorous tone to an utterance. For example, in excerpt (22) speaker A adds the QT *or what* to his potentially face-threatening act when he challenges speaker B about her prejudices against students of De La Salle University.

- (22) <\$A<#>Well when I was in college I lived with my parents and uh by the way I I took up uh Engineering <[/>Industrial</> Engineering
 <\$B><#><[/>Where</>
 <\$A><#>At uh La Salle one of those uh
 <\$B><#>La Salle
 <\$A><#><&>laughter</&> Why you have uh a notion for La Sallites *or* <[/>what</>
 <\$B><#><[/>Uhm</> well-off uh students <&>laughter</&> (S1A-084)

This brief qualitative perspective on QT use has used the three functional categories as a basis to describe the extended functional diversity of QTs. This analysis has also shown that the specific role of a speaker in a particular communicative setting has a strong effect on how they use QTs as a conversational strategy. Speakers use QTs to structure various types of conversations in different ways and to express a range of different stances. Therefore, it is of utter importance to analyze not only the form of QTs—and by extension other discourse-pragmatic features—but also their function, and to study variation on both levels across different text types.

Discussion and Conclusion

The analysis of QTs in dialogues from ICE-PHI has shown that Filipino speakers use a wide range of different forms and combine nearly even proportions of English and Tagalog QTs. This shows that Tagalog QTs are an integral part of PhiE and are thus particularly characteristic of the variety. In contrast, English variant QT forms only play a marginal role in dialogues in ICE-PHI as they make up only 3.4% of all QTs. While variant QTs have a rather focused functional profile as they mainly function informatively, they are not telling of text type variation in the subcorpus. Thus, variant QTs cannot be assumed to be especially characteristic for PhiE contrary to previous conclusions on QT use in PhiE (Borlongan 2008; Parviainen 2016; Takahashi 2016). The overwhelming focus on variant QTs in research on English QTs is not necessarily justified by actual usage patterns in PhiE. Wilson et al. (2017) have also shown that invariant QTs dominate by far over variant ones in Trinidadian English. In contrast to these two ESL varieties, González (2018) shows a clear preference for variant ones in British English. There is thus reason to presume that in ESL varieties of English, invariant QTs are much more characteristic of the different varieties and are more salient for internal variation. Future research on QT use in ESL and ENL varieties should include both variant and invariant forms (see also Barron 2015: 224–225) to further substantiate this hypothesis. Future descriptions of varieties of English (e.g. Kortmann and Schneider 2008) need to include invariant QTs and their locally specific usage profiles instead of dismissing them as “garden variety” features of ESLs (Mesthrie 2008: 30).

The high frequency of Tagalog QTs indicates that on this level of linguistic variation Tagalog is well integrated into PhiE. Except for *'no*, there is a tendency that Tagalog QTs are slightly more frequent in informal private dialogues. However, there are substantial differences in the usage pattern of individual forms: *'no* is

found to be the overall most frequent QT, corroborating the high prevalence of *'no* in ICE-PHI described by Bautista (2011a). *Hinid/di ba*, *e*, and *ha* are relatively frequent with 50 to almost 100 occurrences. *E* has previously not been discussed as a discourse-pragmatic feature in PhiE. However, it is questionable whether it is a Tagalog form as marked in the corpus or similar to *eh*, which has been attested for several varieties of English (Columbus 2009, 2010; Gold and Tremblay 2006). The remaining Tagalog forms, *ano*, *ba*, and *oo*, are rare in the subcorpus with less than 20 tokens each. This variation in frequency suggests that some forms are more easily integrated into English than others, which are restricted to specific speakers, functions, and contexts. With regard to the invariant English forms, *OK*, *right*, and *you know* are highly frequent, while all other forms are rarely used by Filipino speakers, such as *yes* or *yeah*. *OK*, *right*, and *you know* seem promising candidates for cross-variety comparisons of QT use as their relative high frequency has been attested in other varieties of English as well (Columbus 2009, 2010; Wilson et al. 2017).

The analysis has further shown that text type is a crucial factor for the use of QTs. The overall frequency of QTs varies substantially in the subcorpus: QTs are highly frequent in classroom lessons, followed by private and broadcast dialogues, while they are rare or almost absent in legal cross-examinations and parliamentary debates, respectively. In addition, text type also has a significant effect on the distribution of individual QTs. Although most QTs are used in a wide range of text types, there are preferential differences: for example, *hindi/di ba*, *you know*, *OK*, and *'no* occur in (almost) all text types, but *hindi/di ba* and *you know* are preferentially used in private dialogues, whereas there is a preference for *OK* and *'no* over other QT forms in classroom lessons. For forms with a lower frequency, such as *ano*, *ba*, *e*, or *ha*, an analysis of variation across text types is problematic. The descriptive statistics suggest that these four forms seem to be restricted to informal dialogues but due to their low frequency the inferential statistics might be biased. The diverse design of ICE offers the possibility to study the effects of text type in a very detailed way as was done in this study and also by Wilson et al. (2017) and González (2018), but such an approach only works well for high frequency phenomena due to the small size of some individual text types. Despite this methodological problem, the analysis has highlighted that the strong focus on private dialogues in previous research on QTs (e.g. Axelsson 2018; Barron et al. 2015; Columbus 2009, 2010; Kimps 2018) only provides a limited view on QT usage. There is a need of research on QTs and other pragmatic phenomena in a wider range of text types to highlight contextualized patterns of use (see also Barron 2015: 224–225).

Besides text type, the pragmatic function also constrains the QT use. The study has illustrated that there are QTs with a relatively close form-function mapping: for example, speakers mainly use *'no* and *OK* to express facilitative functions and *e*, *ha*, and *you know* for punctuational meanings. While *right* and variant QTs are preferred for informative functions over other QTs, their form-function profile is less closely defined. Despite these preferential differences, most of the QT forms can serve all three functional categories. The form-function relationship has so far only been analyzed for variant QTs (e.g. Barron 2015; Barron et al. 2015; Borlongan 2008; Kimps 2018) or for variant in contrast to all invariant QTs (González 2018). However, this study has shown that invariant QTs are functionally very

diverse and cannot be treated as one homogenous group: they are not chosen randomly but there are fine-grained differences in their functions that influence the selection. The form-function relationship of particular forms in specific varieties of English needs to be addressed more closely. Such work can form the basis for descriptions of the pragmatics of a particular variety and can be used for more detailed definitions of invariant QT forms in grammar books.

The final qualitative analysis has highlighted further complexities of QT use: the exact function of a particular QT also depends on the specific context. This means that text type, function, and form interact. The decisive aspect about text type is not so much the level of formality but the specific roles of speakers they fulfill in a text type (see also González 2018: 138). This role strongly influences the use of QTs and their exact function: for example, a facilitative use of 'no' in a classroom lesson by a lecturer is somewhat different than 'no' in an informal dialogue. Thus, in order to describe the functional profile of QTs—and by extension discourse-pragmatic features in general—corpus-pragmatic research needs to study these pragmatic phenomena in various contexts. Such an approach can only be carried out with corpora that encompass different (spoken) text types. Whereas large online corpora allow studying lower frequency items and drawing more sound conclusions, they only encompass a narrow range of mostly written text types and are thus not suitable for such a context-sensitive corpus-pragmatic approach. The ICE corpora have their limitations due to their small size but have a diverse design and offer the possibility for cross-variety comparisons. Despite the vast body of ICE-based research, a corpus-pragmatic approach that pays close attention to text type and thus uses the diverse design to its full potential is largely unexplored for the ICE corpora.

The study has shown that QTs are a good indicator for the integration of Tagalog into PhiE. QTs encompass both English and Tagalog forms, they are highly frequent, Tagalog QTs even appear in texts where there is no other code-switching, and they are an integral part of most dialogic text types. Thus, the corpus-pragmatic approach taken in this study can provide valuable insights into language contact situations: discourse-pragmatic features from one language might be more easily integratable into another language than syntactically relevant forms. This multilingual variation in the use of QTs does not only characterize PhiE but is also telling of variety-internal heterogeneity as there is substantial variation in the use of QTs with regard to text type. A cross-variety comparison that includes text type variation has the potential to highlight further differences and similarities in QT use across Englishes. A final research desideratum is the combination of this corpus-pragmatic approach with a survey study that has the potential to illustrate the variation in the social meaning of the QT forms. This speaker perspective could prove very valuable to further explain the variation in QT use found in ICE-PHI.

Acknowledgements Open Access funding provided by Projekt DEAL.

Compliance with Ethical Standards

Conflict of interest The 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., & Rühlemann, C. (2015). Corpus pragmatics: Laying the foundations. In K. Aijmer & C. Rühlemann (Eds.), *Corpus pragmatics: A handbook*. Cambridge: Cambridge University Press.
- Anderson, G. (2016). Using the corpus-driven method to chart discourse-pragmatic change. In H. Pichler (Ed.), *Discourse-pragmatic variation and change in English* (pp. 21–40). Cambridge: Cambridge University Press.
- Axelsson, K. (2018). Canonical tag questions in contemporary British English. In V. Brezina, R. Love, & K. Aijmer (Eds.), *Corpus approaches to contemporary British speech: Sociolinguistic studies of the Spoken BNC2014* (pp. 96–119). London: Routledge.
- Barron, A. (2015). And your wedding is the twenty-second <.> of June is it?: Tag questions in Irish English. In C. P. Amador-Moreno, K. McCafferty, & E. Vaughan (Eds.), *Pragmatic markers in Irish English* (pp. 203–228). Amsterdam: Benjamins.
- Barron, A., Pandarova, I., & Muderack, K. (2015). Tag questions across Irish English and British English: A corpus analysis of form and function. *Multilingua*, 34(4), 495–525.
- Bautista, L. (2011a). Some notes on 'no in Philippine English. In L. Bautista (Ed.), *Studies of Philippine English: Exploring the Philippine component of the international corpus of English* (pp. 75–89). Manila: Anvil.
- Bautista, L. (2011b). Introducing the Philippine component of the international corpus of English circa 2010. In L. Bautista (Ed.), *Studies of Philippine English: Exploring the Philippine component of the international corpus of English* (pp. 3–22). Manila: Anvil.
- Biber, D., Johansson, S., Leech, G., Conrad, S., & Finegan, E. (1999). *The Longman grammar of spoken and written English*. London: Longman.
- Borlongan, A. (2008). Tag questions in Philippine English. *Philippine Journal of Linguistics*, 39, 1–34.
- 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). Kenilworth: Rodopi.
- Columbus, G. (2010). A comparative analysis of invariant tags in three varieties of English. *English World-Wide*, 31(3), 288–310.
- Eberhard, D., Simons, G., & Fennig, C. (2019). *Ethnologue: Languages of the world*. Dallas: SIL International.
- Fedriani, C., & Sansò, A. (2017). Pragmatic markers, discourse markers and modal particles: What do we know and where do we go from here? In C. Fedriani & A. Sansò (Eds.), *Pragmatic markers, discourse markers and modal particles: New perspectives* (pp. 1–33). Amsterdam: Benjamins.
- Gold, E., & Tremblay, M. (2006). Eh? and hein?: Discourse particles or national icons. *The Canadian Journal of Linguistics*, 51(2/3), 247–263.
- González, M. D. L. Á. G. (2018). God that came out quick didn't it eh: Variable and invariable tag questions in spoken British English. In M. D. L. Á. G. González & J. L. Mackenzie (Eds.), *The construction of discourse as verbal interaction* (pp. 109–144). Amsterdam: Benjamins.
- Greenbaum, S. (1996). Introducing ICE. In S. Greenbaum (Ed.), *Comparing English worldwide: The international corpus of English* (pp. 3–12). Oxford: Clarendon.

- Gries, S. (2010). Useful statistics for corpus linguistics. In A. Sánchez & M. Almela (Eds.), *A mosaic of corpus linguistics* (pp. 269–291). Frankfurt am Main: Peter Lang.
- Johnson, D. (2009). Getting off the GoldVarb standard: Introducing Rbrul for mixed-effects variable rule analysis. *Language and Linguistics Compass*, 31(1), 359–383.
- Kimps, D. (2018). *Tag questions in conversations: A typology of their interactional and stance meanings*. Amsterdam: Benjamins.
- Kortmann, B., & Schneider, E. W. (Eds.). (2008). *A handbook of varieties of English*. Berlin: de Gruyter.
- Lange, C. (2012). *The syntax of spoken Indian English*. Amsterdam: Benjamins.
- Lim, J., & Borlongan, A. (2011). Tagalog particles in Philippine English: The case of ba, na, 'no, and pa. *Philippine Journal of Linguistics*, 42, 59–74.
- Lim, J., & Borlongan, A. (2012). Corpus-based grammatical studies of Philippine English and language assessment: Issues and perspectives. *The Assessment Handbook*, 8, 51–62.
- Mesthrie, R. (2008). Introduction. In R. Mesthrie (Ed.), *Varieties of English* (Vol. 4, pp. 23–31)., Africa and South and Southeast Asia Berlin: de Gruyter.
- Nelson, G. (1996). The design of the corpus. In S. Greenbaum (Ed.), *Comparing English worldwide: The international corpus of English* (pp. 27–35). Oxford: Clarendon.
- Parviainen, H. (2016). The invariant tag isn't it in Asian Englishes. *World Englishes*, 35(1), 98–117.
- Pichler, H. (2013). *The structure of discourse-pragmatic variation*. Amsterdam: Benjamins.
- Pichler, H. (2016). Uncovering discourse-pragmatic innovations: *innit* in Multicultural London English. In H. Pichler (Ed.), *Discourse-pragmatic variation and change in English* (pp. 59–85). Cambridge: Cambridge University Press.
- Schiffrin, D. (1987). *Discourse markers*. Cambridge: Cambridge University Press.
- Takahashi, M. (2016). A comparative study of tag questions in four Asian Englishes from a corpus-based approach. *Asian Englishes*, 16(2), 101–124.
- Thompson, R. (2003). *Filipino English and Taglish*. Amsterdam: Benjamins.
- Tottie, G., & Hoffmann, S. (2006). Tag questions in British and American English. *Journal of English Linguistics*, 34(4), 283–311.
- Tupas, R. (2004). The global politics of Philippine English: Neocolonialism, global politics, and the problem of postcolonialism. *World Englishes*, 23(1), 47–58.
- Wilson, G., Westphal, M., Hartmann, J., & Deuber, D. (2017). The use of question tags in different text types of Trinidadian English. *World Englishes*, 36(4), 726–743.
- Wong, M. (2007). Tag questions in Hong Kong English: A corpus-based study. *Asian Englishes*, 10(1), 44–61.

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