'They are going tomorrow, isn't it?' On the Use of Tag Questions in Indian English and Hong Kong English

Miriam Criado Peña[*]

Universidad de Málaga

mcriado@uma.es

Abstract

Tag questions in standard British English (BrE) follow a standard pattern consisting of an operator and a subject. This operator generally coincides with the preceding statement while the auxiliary "do" is the choice when the operator is absent. More importantly, a negative tag is generally attached to a positive statement and vice versa (i.e. you know her, don't you?) (Quirk et al. 1985: 810).

The Asian varieties of English are an exception insofar as apparently no standard rule is observed. The present paper investigates the use and distribution of regular and irregular tag questions in Indian English and Hong Kong English with the following objectives: a) to analyze the distribution of the construction of regular and irregular tag questions across these varieties; b) to assess their frequency across speech and writing, text types included; and c) to evaluate the sociolinguistic variation, if any. For the purpose, the Indian and Hong Kong components of the *International Corpus of English* (ICE-Ind and ICE-HK) will be used as sources of analysis.

1 Introduction

In British Standard English (henceforth BrE) a tag question is a subordinate sentence attached to a superordinate sentence following a standard pattern of composition: an operator and a subject. This operator normally coincides with the precedent, and it consists of an auxiliary "do" in its absence (2). The subject, in turn, is a personal pronoun agreeing in number, person, and gender with that of the main sentence (Quirk et al., 1985, p. 810). As shown in examples 1-3 with both main and auxiliary

---

verbs in BrE, reversed polarity tags are more frequent than constant polarity tags, thence, positive anchors are usually accompanied by negative tags and vice versa (Huddleston & Pullum, 2002, p. 892).

    (1)   He has finished, hasn't he?
    (2)   He finished, didn't he?
    (3)   Let's go to the party, shall we?

In World Englishes, however, tag questions sometimes differ from the British English practice[†]. In Australian and New Zealand English tag questions differ from BrE regarding the use of some modal verbs in particular, as in the verbs "should" or "ought": *They ought to go, oughtn't they?* (Trudgill & Hannah, 1985, p. 20). Malaysian English, for its part, includes the invariant phrase *can or not?* in the formation of tag questions: *They must submit the forms tomorrow, can or not?* (Baskaran, 2004, p. 1079). In addition to this, there is evidence of an increased use of the invariant tag questions "isn't it?" and "is it?" in standard and non-standard speakers of varieties such as Welsh, Singapore and Malaysian English, regardless of the typology of the verb or the subject in the main clause (Trudgill & Hannah, 1985, p. 30; Deterding, 2007, p. 56). This phenomenon has been attested in different varieties of English worldwide with a 67% of attestation and 76% of pervasiveness according to *eWAVE* (Kortmann & Lunkenheimer, 2013). In the case of Indian English (henceforth IndE) and Hong Kong English (henceforth HKE) this phenomenon is labeled as pervasive or obligatory, and therefore, its use is very frequent among speakers of these varieties.

In order to provide an accurate classification of World Englishes, Schneider (2007) proposed his Dynamic Model[‡]. According to this model, it is claimed that, despite all surface differences, there is an underlying uniform process driving the formation of many Postcolonical varieties of English. This model accounts for many similarities and is conceived as a progression of five characteristic stages, i.e. formation, exonormative stabilization, nativization, endonormative stabilization and differentiation (Schneider, 2007, pp. 21; 32; Mesthrie & Bhatt, 2008, pp. 32-36). IndE and HKE are in phase 3 (nativization), a stage resulting "in the heaviest effects on the restructuring of the English language itself; it is at the heart of the birth of a new, formally distinct Post-Colonial English" (Schneider, 2007, p. 44). These varieties are, however, well advanced in the process of nativization and already moving towards phase 4, the phase of endonormative stabilization (Setter, Wong & Chan, 2010, p. 116).

Some research has been undertaken on the use of tag questions in different varieties of English worldwide. In Indian English, the topic has been recently addressed by Achiri-Taboh (2015), who has studied the invariant tag question "isn't it?" confirming its increasing emergence in non-standard English varieties around the world. He has also noticed the use of words such as "right" and "okay" as substitutes of this tag. On the other hand, Lange (2012) argues that the relevance must be given to the variation between the invariant tags "isn't?" and "is it?" and the indigenous tags "no" and "na" instead of focusing on the differences between invariant and canonical tags as studied by others scholars. Within this variation she notices a preference for indigenous tags (87.67%) over the rest of the forms, confirming that reversed and constant polarities do not follow any pattern as most of the affirmative

---

[†] In African English the use of the invariant tag questions "is it?" and "isn't it?" is widespread. Speakers of this variety tend to use them without considering the tense, person or auxiliary of the superordinate clause (Trudgill & Hannah, 1985, p. 104). Nevertheless, some distinctions between countries are apparent, as Mbangwana (2004) affirms, in Cameroon English the form "isn't?" is preferable to any other because of its simplicity, while other forms are acceptable as well: "na", "not so", "ein", "is that", "right" and "okay": *I told you she will come, na?* (p. 905). In East African English this form is used together with "not so?": *The price in the display is a very good idea because… Not so?* (Schmied, 2004, p. 935).

[‡] "In pedagogical literature the term 'model' is used in two senses: first, in the sense of acceptability, generally by the native speakers of a language; second, in the sense of fulfilling codified prerequisites according to a given 'standard' or 'norm' at various linguistic levels. In this sense, then, we may say that a model provides a *proficiency scale*" (Kachru, 1983, p. 31).

tags "is it?" are accompanied by positive anchors. She finally asserts that the election of invariant tag questions is not related to politeness but to "speakers' personal styles" (2012, pp. 206-215).

HKE has not received so much attention as compared to other Asian varieties. Bolton (2000, 2008) studied its sociolinguistic background to demonstrate its importance within the English community of speakers reporting that 45% of the population speaks the language. Columbus (2009) analyzed tag questions in five varieties, including IndE and HKE. His results show a larger amount of tag questions in IndE (p. 45) as a result of an outstanding preference for the negative form "isn't it?" over "is it?" On the other hand, HKE speakers prefer the positive invariant tag to formulate questions (25 out of 29), even though the use of other invariant tags is more frequent, being "no" the most utilized in IndE with 237 occurrences and "right" in HKE, amounting to 110 instances. Setter, Wong, and Chan (2010) found out that "Hong Kong speakers use many fewer question tags than do the British speakers" (p. 79) in conversational English with a high percentage of the tags "isn't it?" and "is it?" Wong (2007) studied a larger number of examples supporting the idea of a more frequent use of the tag "is it?" and positive-positive constructions (as cited in Setter, Wong, and Chan, 2010, p. 79).

Apart from these references to the topic, tag questions have been practically neglected in the literature in the recent varieties of English worldwide. The present paper investigates the use and distribution of regular and irregular tag questions in IndE and HKE with the following objectives: a) to analyze the distribution of the construction of regular and irregular tag questions across varieties; b) to assess their frequency across speech and writing, and text types; and c) to evaluate the sociolinguistic variation, if any.

2 Methodology

This section will describe the procedure followed for the analysis, as well as the material and the linguistic tool used during the process. The former consists of a corpus of texts serving as source data, while the latter is the computer software with which the search of tag questions has been conducted.

2.1 International Corpus of English (ICE)

The present study relies on the tagged versions of the *International Corpus of English* (ICE), and for the purpose, the components of IndE and HKE have been surveyed. Huang (2016) defines a corpus as "a systematic collection of naturally occurring spoken or written language or a variety of such language, which can be searchable online" (p. 217). The ICE is "a megacorpus project in nature, which was initiated by Sidney Greenbaum in the late 1980s" (Bernaisch, 2015, p. 58). The aim of this project is "to provide the resources for comparative studies of the Englishes used in countries where it is either a majority first language or an official additional language" (Greenbaum & Nelson, 1996, p. 3). The corpus is designed as a collection of real text material with a total of approximately one million words, being divided into 500 texts of 2000 words each one. At the same time, it consists of spoken texts (60%), and written texts (40%), which vary from informal to highly formal registers (Bernaisch, 2015, p. 58). The following table shows the total number of words forming each component of the corpus. As shown, the HKE corpus is slightly larger than the IndE.

ICE component	Spoken	Written	Total
IndE	694,249	438,691	1,132,940
HKE	975,063	498,893	1,473,956

Table 1: Word-count of the IndE and HKE ICE components

2.2　AntConc

On methodological grounds, the software *Antconc* has been used in order to retrieve the instances automatically. This programme was first released in 2002 (Laurence, 2005, p. 730), and it has been upgraded since then, in this case the version 3.4.3 has been used. Laurence (2005) evaluates the importance of this tool stating that "using a reasonably large corpus, a concordance program can find and display a huge number of examples in varied contexts and situations quickly and more efficiently" (p. 730).

The process, however, is not straightforward as different searches were implemented in order to cope with the variety of tag questions. In this fashion, some of the tags used have been PPH1 for the 3rd person singular neuter pronoun "it"; or PPY for the 2nd person pronoun "you"; those for verbal forms, such as VBDZ for the past tense of the verb to be, "was", or VD0 for the base form of the verb "do"; and those for modal auxiliaries VM, and negative forms, as XX for "not" and "n't". This tagging system allows a very little margin of error of 1.5%, and 3.3% of ambiguity (Leech, Garside, & Bryant, 1994, p. 625).

Taking into account the previously mentioned tags, the first step necessarily consists of the search of the distinct combinations using the appropriate wildcards. For example, in order to obtain a negative tag question with the 2nd person personal pronoun "you", the input will be *_XX@VP, in which the asterisk (*) means "zero or more characters", and the at sign (@) stands for "zero or any word". However, a process of manual disambiguation is also required. For the purpose, an Excel spreadsheet has been used. This process consists in the omission of those hits which are not question tags, and therefore, are beyond the scope of the present study. Next, it is necessary to classify our results into regular and irregular cases, the latter being the most relevant for our study. The disambiguation process has eventually provided us with a total of 387 instances of tag questions, of which 121 belong to the IndE and 266 to the HKE components of ICE.

3　Analysis

The raw figures, percentage and normalized frequencies of the phenomenon are given in Table 2 below, classified not only in terms of the language variety but also in terms of the tag, be it regular or irregular. Frequencies have been normalized to tokens per 100,000 words for the sake of comparison in all the results shown henceforth.

	IndE			HKE		
	Raw	%	n.f	Raw	%	n.f
Regular	53	43.80	4.67	106	38.97	7.19
Irregular	68	56.20	6	166	61.03	11.26

Table 2: Frequencies for regular and irregular tag questions in IndE and HKE

Regular and irregular tag questions occur more frequently in HKE than in IndE, especially when it concerns the use of irregular tags, which outnumbers its occurrences (11.26 over 6). However, in both cases the irregular type is the preferred form for informants (6 against 4.67 in IndE, and 11.26 against 7.19 in HKE). This predilection for the irregularity could be a proof of the development of the language, that is in constant change especially in World Englishes. In this case, the total of tag questions found in ICE-HKE almost double the number of those found in ICE-IND.

	IndE			HKE		
	Raw	%	n.f	Raw	%	n.f
Spoken	66	97.06	5.83	160	96.39	10.86
Written	2	2.94	0.18	6	3.61	0.41

Table 3: Frequencies for irregular tag questions across speech and writing in IndE and HKE

In Table 3 the scope has been narrowed providing the results obtained in the irregular tags, presenting their distribution in the written and spoken domains in IndE and HKE. The preference for these tags in the spoken domain is evident given the large gap between spoken and written language in their number of occurrences and percentage. Therefore, it seems that the use of irregular constructions is related to spontaneous and fluent language. In HKE, for example, the normalized frequency of irregular tags amounts to 10.86 instances in speech compared to the 0.41 examples in the written texts. In the light of this, irregular tags are constrained in written texts since their use is uncommon among speakers of both varieties. Even though the percentage of tags in spoken and written discourse is similar in both varieties, they present important difference in terms of frequency. The use of irregular tag questions in HKE doubles its Indian counterpart; 10.86 tags are encountered every 100,000 words in HKE, while half of them appear in IndE (5.83 per 100,000 words). These results contrast with the ideas argued by Setter, Wong, and Chan (2010), who stated that HKE speakers use invariant tags in the oral discourse to a lesser extent than Indian English speakers "because of the influence of Cantonese, which does not make use of question tags" (p. 79).

As the use of irregular tags is so constrained in the written discourse the focus will be narrowed to their use in the oral discourse. Results have been classified into two categories: dialogues and monologues, and at the same time the former have been subdivided into private dialogues (direct conversations and phone calls) and public dialogues (broadcast discussions and interviews, class lessons, and parliamentary debates among others). The latter has been subdivided into unscripted monologues (commentaries, demonstrations, legal presentations, and unscripted speeches) and scripted monologues (broadcast and non-broadcast talks, and broadcast news).

	Dialogues				Monologues			
	Private		Public		Unscripted		Unscripted	
	Raw	n.f	Raw	n.f	Raw	n.f	Raw	n.f
IndE	45	3.97	17	1.50	4	0.27	0	0
HKE	129	8.75	27	1.83	4	0.35	0	0

Table 4: Irregular tag questions across registers in spoken discourse in IndE and HKE

The results displayed in Table 4 show a clear preference for irregular tags in dialogues suggesting that communication between people favours its use. The corpus of IndE contains a total of 5.47 occurrences of tags in dialogues while only 0.27 occurrences are found in monologues, and similar results are encountered in HKE with 10.58 in dialogues and 0.35 in monologues, thus the use of irregular tag questions in monologues is negligible in both varieties. Nevertheless, there are differences between private and public dialogues with a more widespread use of irregular tags in private conversations, where probably speakers are in a more relaxed mood and do not pay so much attention to their performance. This can also suggest that the use of irregular tag questions is related to more informal contexts. Moreover, the gap of appearances of tags between private and public dialogues is larger in HKE than in IndE (8.75 over 1.83 in HKE, and 3.97 over 1.50 in IndE).

ICE-Ind and ICE-HK corpora have been designed to represent variation in terms of age and gender facilitating a sociolinguistic study. Figures 1 and 2 below show the distribution of irregular tag questions among speakers of IndE and HKE according to their gender and age.
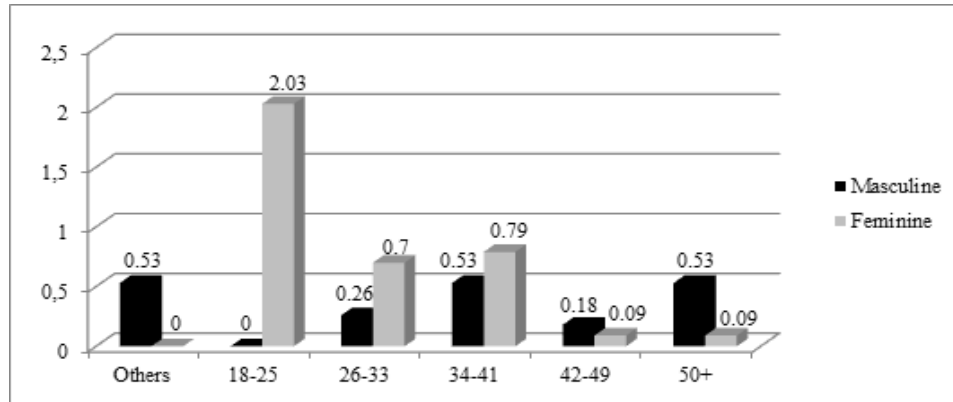

Fig. 1: Irregular tag questions according to gender and age in IndE (n.f)

The data displayed above confirms an outstanding preference for irregular tag questions in women as they represent the largest amount of instances of this construction within the corpus. Within the age group of 18-25 years, females reach 2.03 occurrences while no evidence of the use of tag questions is given among male informants. Thus, taking into account the results, females can be considered to be the gender pioneering this linguistic phenomenon in this variety of English. On the other hand, young speakers of IndE tend to use more irregular tag questions than adults. This preference is more widespread among speakers in-between 18 and 25 years, the group where the highest frequencies of tags are found. These results suggest an increasing dissemination of the use of irregular tag questions promoted by the new generations, a practice that is already being stabilized in the language inasmuch as the following groups of speakers (26-33 and 34-41) gather the next two highest percentages of irregular constructions.
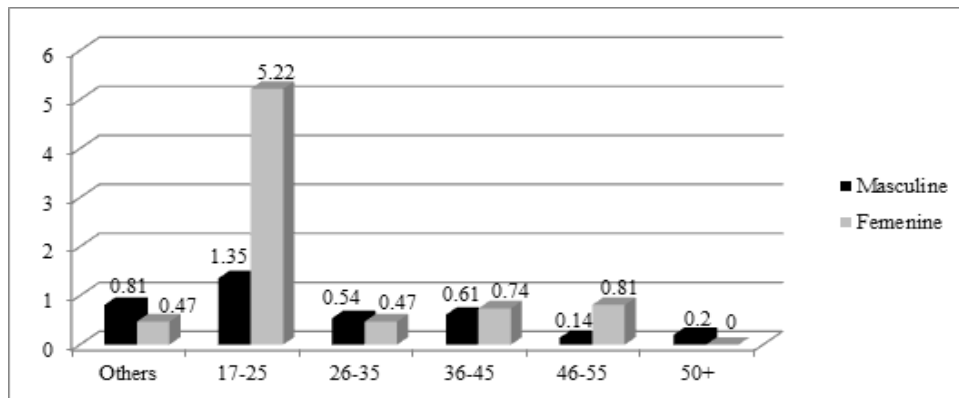

Fig. 2: Irregular tag questions according to gender and age in HKE (n.f)

The distribution of irregular tag questions in ICE-HK is similar to ICE-IndE in the sense that females are again leading the diffusion of irregular tag questions. This difference is more visible

among the group of younger speakers; the number of examples collected in samples of females substantially outnumbers all the other age groups considered. Besides, this use of tags is overwhelmingly favoured among speakers between the ages of 17 and 25 (5.22 and 1.35). Even though the frequencies vary significantly between both genders, they are the highest among all groups, and they are followed by the age ranging 36-45 years, amounting to 0.61 and 0.74. Contrariwise, the use of this irregular construction is constrained among older speakers, to whom the smallest quantity of occurrences correspond (0.2 and 0). The similarities between the results of IndE and HKE corpora are visible given that women are the promoters of the linguistic change in both of them, making use of irregular tag questions with a greater frequency than males.

4 Conclusions

The present paper deals with the use and distribution of tag questions in two varieties of Asian Englishes: IndE and HKE, an analysis carried out in order to examine its distribution across varieties, its diffusion across speech and writing and text types, and the diverse sociological factors affecting the linguistic variation. For the purpose, this study is based on the *International Corpus of English* (ICE), whose dimension and characteristics conforms it as an appropriate input for the study of this linguistic feature. Our study allows us to reach the following conclusions.

First, Wold Englishes differ from the standard practice in terms of tag questions inasmuch as the use of the invariant forms "isn't it?" and "is it?" play an important role since this phenomenon is found to have a 67% of attestation and 76% of pervasiveness according to *eWAVE* (Kortmann & Lunkenheimer 2013), where its use is labeled as obligatory.

Second, it can be gathered that the use of tag questions is more widespread among speakers of HKE, whose results almost double those of IndE. In the employment of this construction there is a clear preference for the irregular forms in both varieties of English, confirming diffusion of the irregular construction in these varieties of English.

Third, the results are examined across text types, showing that this irregularity in the use of tag questions is especially noted in the oral discourse, where the number of occurrences overwhelmingly surpasses that of written texts, suggesting a connection between this construction and the spontaneity of spoken language. There is also an outstanding preference for the irregular construction in dialogues, to the extent that its use is almost non-existent in monologues. However, among dialogues this practice is much more widespread in informal contexts, that is to say, in private dialogues, due to the lack of attention in the performance, being the gap between private and public dialogues more significant.

Finally, this paper also corroborates that this feature responds to sociological factors; in IndE there is an outstanding preference for this construction among female informants, especially in young women in-between 18 and 25 years, and this practice can be considered to be in process of stabilization in this variety since the following age groups accumulate the next highest number of occurrences. In HKE, women are also the leaders of this irregular linguistic construction.

References

Achiri-Taboh, B. (2015). A Generalized Question Tag in English. *English Today, 31*(1), 48-54.

Baskaran, L. (2004). Malaysian English: Morphology and Syntax. In B. Kortmann, E. W. Schneider, K. Burridge, R. Mesthrie & C. Upton (Eds.), *A Handbook of Varieties of English* (pp. 1073-1086). (Vol. 2). Berlin: Mouton de Gruyter.

Bernaisch, T. (2015). *The Lexis and Lexicogrammar of Sri Lankan English*. Amsterdam: John Benjamins Publishing Company.

Bolton, K. (2000). The Sociolinguistics of Hong Kong and the space for Hong Kong English. *World Englishes, 19*(3), 265-285.

Bolton, K. (2008). English in Asia, Asian Englishes, and the issue of proficiency. *English Today, 24*(2), 3-12.

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 Reassessements* (pp. 401-414). Amsterdam: Rodopi.

Deterding, D. (2007). *Singapore English*. Edinburgh: Edinburgh University Press.

Greenbaum, S., & Nelson, G. (1996). The International Corpus of English (ICE) Project. *World Englishes, 15*(1), 3-15.

Huang, Y. (2016). Pragmatics: Language use in context. In K. Allan (Ed.), *The Routledge Handbook of Linguistics* (pp. 205-220). New York, NY: Routledge.

Kachru, B. B. (1983). *The Other Tongue: English Across Cultures*. Oxford: Pergamon Press.

Kortmann, B., & Lunkenheimer, K. (Eds.). (2013). The Electronic World Atlas of Varieties of English. Leipzig: Max Planck Institute for Evolutionary Anthropology. Retrieved from http://ewave-atlas.org, Accessed on 2016-02-10.

Lange, C. (2012). *The Syntax of Spoken Indian English*. Amsterdam: John Benjamins Publishing.

Laurence, A. (Ed.). (2005). *Proceedings of Professional Communication Conference*. Limerick: IEEE.

Leech, G., Garside, R., & Bryant, M. (Eds.). (1994). *Proceedings of the 15th International Conference on Computational Linguistics (COLING 94)*. Kyoto: Association for Computational Linguistics.

Mbangwana, P. (2004). Cameroon English: Morphology and Syntax. In B. Kortmann, E. W. Schneider, K. Burridge, R. Mesthrie & C. Upton (Eds.), *A Handbook of Varieties of English* (pp. 898-908). (Vol. 2). Berlin: Mouton de Gruyter.

Mesthrie, R., & Bhatt, R. M. (2008). *World Englishes: The Study of New Linguistic Varieties*. Cambridge: Cambridge University Press.

Schmied, J. (2004). East African English (Kenya, Uganda, Tanzania): Morphology and Syntax. In Kortmann, B., Schneider, E. W., Burridge, K., Mesthrie, R., & C. Upton (Eds.), *A Handbook of Varieties of English* (pp. 929-947). (Vol. 2). Berlin: Mouton de Gruyter.

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

Setter, J., Wong, C. S. P., & Chan, B. H. S. (2010). *Dialects of English: Hong Kong English*. Edinburgh: Edinburgh University Press.

Quirk, R., Greenbaum, S., Leech, G., & Svartvik, J. (1985). *A Comprehensive Grammar of the English Language*. New York, NY: Longman.

Wong, M. L. Y. (2007). Tag questions in Hong Kong English: a corpus-based English. *Asian Englishes, 10*(1): 44-61.