English Today

cambridge.org/eng

A corpus-based study on the "ungrammatical" *aren't I*

Mingyou Xiang[1] and Xiao Jiang[1,2]

[1]School of international studies, University of International Business and Economics, Beijing, China and [2]School of Foreign Languages, Nanjing Institute of Technology, Nanjing, China

Shorter Article

Cite this article: Xiang M, Jiang X (2024). A corpus-based study on the "ungrammatical" *aren't I*. *English Today* **40**, 226–230. https://doi.org/10.1017/S0266078424000245

First published online: 18 September 2024

Keywords:
aren't I; question tags; contraction form; language phenomenon; corpus

Corresponding author:
Xiao Jiang;
Email: 627780158@qq.comha

Abstract

Concerning the "ungrammatical" interrogative form *aren't I*, many scholars have made their points. However, these scholars' arguments are based on their personal observations and few studies have examined this phenomenon against large corpora. This study aimed at investigating the widespread usage of "ungrammatical" contraction form *aren't I* in question tags from both quantitative and qualitative perspectives. Based on large corpora, this study showed a clear picture of the current frequency of use of the question tags *aren't I* and other alternatives (*amn't I*, *ain't I*, *am I not* and *an't I*) in modern English. From a qualitative perspective, this study found that the reason why *aren't I* has taken hold as a recognized standard form around the globe lies in that the use of *aren't I* appears to be a smart coincidence to imply the potential double roles of "I" as both the addresser and the addressee in a monologue. In addition, the fact of the matter that *amn't I* is difficult to pronounce, *am I not* is bookish, *an't I* is old-fashioned and *ain't I* can only be used in informal situations, increases the popularity of *aren't I*. The findings of this study can justify the usage of "ungrammatical" *aren't I* as a natural norm in both British English and American English. These findings open new research avenues alongside pedagogical and sociolinguistic implications for other similar "ungrammatical" language phenomenon.

Introduction

Question tags are "a very conspicuous phenomenon of spoken language" (Tottie and Hoffman 2006, 284). These short questions (tags), tagged onto a main statement (the anchor), contribute significantly to the flow of spoken English. (Mbakop 2022, 27) Quirk et al. (1985, 810) proposed several general rules that govern the construction of the most common tags, which can be summarized to the anchor-tag predicate match and the polarity rule. The anchor-tag predicate match stipulates that the operator of the anchor should be the same as the operator of the tag. The polarity rule states that if the anchor is positive, the tag is generally negative, and vice versa. According to the above rules, typical cases include (a) *She is hungry, isn't she?* and (b) *They didn't make any mistakes, did they?* Nevertheless, these rules are too simplistic to interpret all the question tag structures found in English. It is noted that there is an obvious verb non-match case (c) *I am brilliant, aren't I?* This sentence violates the anchor-tag predicate match for the operator in the tag *are* is not the same as the operator of the preceding statement *am*. The grammatical alternative seems to be (d) *I am brilliant, amn't I?*

Concerning the above phenomenon, many scholars have made their points. According to Jørgensen (1979, 35), in modern spoken English, the "ungrammatical" interrogative form *aren't I* is becoming increasingly dominant as the normal contraction form like *aren't you, isn't he (she, it)*, etc. The contraction *amn't I*, which might seem obvious superficially, is considered unacceptable in Standard English (Jørgensen 1979, 35), and is mainly used as the colloquial norm in Scottish and Irish English (Quirk et al. 1985; Crystal 2021). The other alternative patterns are obviously falling into disuse. (Jørgensen 1979, 35) As Swan (2019, 767) remarks in Practical English Usage, in questions, *am not* can only be contracted into *aren't*, such as *I'm late, aren't I?* Likewise, Collins COBUILD English Usage expresses a preference for using *aren't I* in questions and tag questions (Hands 2018, 171).

However, these scholars' arguments are based on their personal observations and few studies have examined this phenomenon against large corpora. The present study starts with a quantitative analysis to display the differences in frequency of use between *aren't I* and other alternatives in both British English and American

© The Author(s), 2024. Published by Cambridge University Press

English. Then, qualitative explanations for the widespread usage of *aren't I* in question tags and the limited usage of its alternatives are offered.

Frequency of use: *aren't I* and its alternatives

Corpora

The corpora used to investigate the frequencies of use of *aren't I* and other alternatives include Corpus of Contemporary American English (COCA), British National Corpus (BNC) and the Scottish Corpus of Texts & Speech (SCOTS). COCA is probably the most widely-used corpus of English. It contains more than one billion words from eight genres: blogs, web pages, TV and movies subtitles, spoken, fiction, popular magazines, newspapers, and academic texts. BNC was originally created by Oxford University Press in the 1980s to early 1990s. It contains 100 million words of text from a wide range of genres (e.g., spoken, fiction, magazines, newspapers, and academic). SCOTS, which aims to cover the wide range of Scottish English texts today (1945 to the present day), currently contains over 4.5 million words of written and spoken texts. The reasons for choosing COCA and BNC involve that they are large-scale and representative corpora of American English and British English, and they are publicly available. In addition, SCOTS is included on the account that many scholars claim that *amn't I* is mainly used in Scottish and Irish English (Quirk et al. 1985; Crystal 2021).

Search strategies

In COCA and BNC, the automated search for the question tags *aren't I* and its alternatives (*amn't I*, *ain't I*, *am I not* and *an't I*) was implemented respectively. In SCOTS, the automated search for only *amn't I* was implemented. Considering that the retrieval of question tags is not easy (Tottie and Hoffman 2006), a manual selection was conducted after each automated search in order to avoid the occurrence of which that do not function as tags, and which were counted as irrelevant instances. The exclusion criteria were adopted as follows:

- The exclusion of instances that do not function as tag questions. (e.g., Why *aren't I* ?)
- The exclusion of instances where the question tag does not function as a contraction of *am not*. (e.g., I got to write something, *ain't I* ?)

Frequency of use: aren't I

The search results for *aren't I* in COCA and BNC are demonstrated in Table 1 and Table 2 respectively.

As displayed from the data in the above two tables, *aren't I* was found to occur 0.75 times every million words in COCA and 1.08 times in BNC in the same proportion. A chi-square test was conducted here (chi-square = 12.3836, $p = 0.000 < 0.01$). The result shows that there is a significant difference between 0.75 and 1.08 per million, which means that the frequency of use of *aren't I* is higher in British English than in American English. In COCA, *aren't I* occurred most frequently in the TV/movies section (4.07 times every million words), with the fiction section following after (1.30 times every million words). In comparison, the section with the highest frequency of use of *aren't I* in BNC is the spoken (5.7 times every million words), with the fiction section at the second place (2.83 times every million words). Based on the above data, we can reasonably conclude that *aren't I* has already become a commonly accepted contraction form in both British English and American English.

Frequency of use: alternatives to aren't I

The search results for *amn't I* in COCA, BNC and SCOTS are demonstrated in Table 3.

According to Quirk et al. (1985, 129) and Crystal (2021, 37), *amn't I* is mainly used as the colloquial norm in Scottish and Irish English. That may be the reason why *amn't I* is rarely used in both American English and British English (3 occurrences in COCA and 0 occurrence in BNC). Considering the fact that SCOTS is a much smaller corpus compared with COCA and BNC, the search result for *amn't I* in SCOTS (6 occurrences) suggests that *amn't I* is much more frequently used in Scottish English than in American English and British English.

The search results for *ain't I* and *am I not* in COCA and BNC are exhibited in Table 4.

According to the data above, *ain't I* was found to occur 0.08 times every million words in COCA and 0.06 times in BNC in the same proportion. An interesting comment is found in the Dictionary of Contemporary American Usage: "The difference between the English *aren't* and the American *ain't* is simply the difference we have in the two pronunciations of 'tomato'". (Evans and Evans 1957, 23) It is meant to imply that *ain't* is used more widely in America, as the American equivalent of the British *aren't*. This result that the frequency of use is higher in COCA seems to coincide with many linguists' perspective that *ain't I* is more of a typical American English expression. In addition, compared with *aren't I*, the frequencies of occurrence of *ain't I* in both COCA and BNC are much lower (COCA: 0.08 times every million words for *ain't I* vs. 0.75 times every million words for *aren't I*; BNC: 0.06 times

Table 1. Frequencies of *aren't I* in each section of COCA

SECTION	ALL	TV/M	FIC	SPOK	WEB	BLOG	MAG	NEWS	ACAD
FREQ	748	522	154	22	16	15	10	9	0
WORDS(M)*	993	128.1	118.3	126.1	124.3	128.6	126.1	121.7	119.8
PER MIL*	0.75	4.07	1.30	0.17	0.13	0.12	0.08	0.07	0

https://doi.org/10.1017/S0266078424000245 Published online by Cambridge University Press

Table 2. Frequencies of *aren't I* in each section of BNC

SECTION	ALL	SPOK	FIC	MAG	NON-ACAD	NEWS	ACAD	MISC
FREQ	108	57	45	2	2	1	1	0
WORDS(M)*	100	10.0	15.9	7.3	16.5	10.5	15.3	20.8
PER MIL*	1.08	5.7	2.83	0.27	0.12	0.10	0.07	0

WORDS(M)*: the total word counts for the whole corpus and each section in the unit of million.
PER MIL*: frequency of occurrence every million words.

every million words for *ain't I* vs. 1.08 times every million words for *aren't I*). This contrast points to the fact that the contraction form *aren't I* is used much more frequently than *ain't I* in both British English and American English.

As shown from the data above, *am I not* was found to occur 0.04 times every million words in COCA and 0.07 times in BNC in the same proportion, which are much lower than the frequencies of occurrence of *aren't I* in both COCA and BNC (COCA: 0.04 times every million words for *am I not* vs. 0.75 times every million words for *aren't I*; BNC: 0.07 times every million words for *am I not* vs. 1.08 times every million words for *aren't I*).

Incidentally, the search results for *an't I* in COCA and BNC (both zero occurrence) reveal that this contraction form is out of date.

Motivations behind the popularity of *aren't I*

Why aren't I

Crystal (2021, 37) has a very interesting explanation for the origin of *aren't I*, which came from people mistaking *an't* (a substitute people created for *amn't*) for *aren't* since [r] after the vowel is silent in the newly emerging Received Pronunciation around 1800. This explanation seems to be intriguing, however, *aren't I* is first attested in the Literary World (in Yang et al. 2017, 207):

> The inventor of an abbreviated form of "am not I?" will do an important service to the language. It is hard to speak these words distinctly in rapid utterance. "ain't I" is much easier, but is undeniably inelegant. "Aren't I?" seems to be thought the correct thing; but why should we say "Aren't I" any more than "I are not"?

Given that writing is more conservative than spoken usage, it is probable that *aren't I* appeared in speech much earlier. In the earliest attestation from 1872, *aren't I* was already perceived as a correct and standard form. Due to the fact that it has been standard for more than 150 years, some scholars approve of the usage of *aren't I*. Thomson and Martinet (1986, 109) consider *aren't I* as an irregular contraction of *am I not*. Fowler and Butterfield (2015, 96) maintain that whatever its origin, *aren't I* is used in regular and natural tag questions in standard Britain English. Crystal (2021, 37) holds that *aren't I* is the standard form in British English, and it is widely used in American English too. However, there are those who put rules above reasonability and consider *aren't I* illogical for the reason that "I" and "are" are not consistent grammatically. (Phillipps 1984; Partridge 1973)

Table 3. Frequencies of *amn't I* in each corpus

Corpus	COCA	BNC	SCOTS
FREQ	3	0	6

Nevertheless, from our point of view, the use of *aren't I* is an ideal method to deal with the real difficulty faced in question tags: there is no appropriate contraction form for am not (*amn't* is considered unacceptable by most English-speaking people). We will find its wide use reasonable if we analyze it from another perspective. As is known, two interlocutors are needed at most to start and propel a conversation: the addresser and the addressee. In speech communication, the addresser and the addressee usually ask and answer questions in turn. There is also the case of asking a question and answering it by oneself. An example is shown in (1):

(1) Person A: *Am I a student?*
Person B_1:*Yes, you are. (No, you are not.)*
Person B_2:*Yes, I am. (No, I am not.)*

In this example, Question A has two possible answers: B_1 and B_2. B_1 is the addressee's answer to the question (dialogue), while B_2 is the addresser's answer to the question he or she poses (monologue). In the latter case, the role "I" is both the addresser and the addressee. An instance is shown in (2):

(2) Person A: *I am a student, aren't I?*
Person B_1:*Yes, you are. (No, you are not.)*
Person B_2:*Yes, I am. (No, I am not.)*

In this instance, similar to Example One, Question A also has two possible answers: B_1 and B_2. B_1 is the addressee's answer to the addresser's question (dialogue), whereas B_2 is the addresser's answer to the question he or she poses (monologue). Therefore, "I" in this question tag may play double roles of both the addresser and the addressee. Given the fact that people refuse to use *amn't*, the use of *aren't I* here appears to be a smart coincidence: "I" in the question tag still represents the addresser, while "are" in the same question tag marks its potential role as an addressee, for "are" can be collocated with the singular addressee "you".

https://doi.org/10.1017/S0266078424000245 Published online by Cambridge University Press

Table 4. Frequencies of *ain't I* and *am I not* in COCA and BNC

ain't I			*am I not*		
Corpus	COCA	BNC	Corpus	COCA	BNC
FREQ	79	6	**FREQ**	39	7
WORDS(M)	993	100	**WORDS(M)**	993	100
PER MIL	0.08	0.06	**PER MIL**	0.04	0.07

Why not amn't I

The fact that most English-speaking people decline to use *amn't I* may result from its pronunciation problem. In terms of articulation rules in modern English, the consonant cluster [mn] is awkward. Although [m] and [n] are both nasal consonants, their places of articulation are distinct. The bilabial [m] is produced using both lips, while the alveolar [n] is produced with the front part of the tongue on the alveolar ridge. Therefore, the related articulators cannot transform from [m] to [n] instantly. Moreover, the power in chest cavity used to produce these two consonants in succession is not sufficient. Hudson (2000, 302) remarks on the *amn't* form like this: "we don't know how to pronounce (or write) the word, and we can't use it." He is not the only linguist to make such kind of comment on *amn't*. Rothstein and Rothstein (2009, 50–51) also hold that we cannot say *amn't I* for the reason that *m* and *n* adjoin, and we are forced to pronounce each word in full.

However, despite the fact that the consonant cluster [mn] is difficult to pronounce, there are some English words with *mn* combination in spelling, the common instances include *mnemonic*, *column*, *condemn*, *damn*, *hymn* and *solemn*. When pronouncing these words, either [m] or [n] is not voiced. For example, *mnemonic* is pronounced as [nɪˈmɒnɪk], and *column* is pronounced as [ˈkɒləm]. In that case, why can't this elision rule apply to the pronunciation of *amn't* as it does to the above-mentioned words? The reason lies in that when *am* and *not* is contracted to *amn't*, this negative contraction should be stressed (Swan 2019, 680), and thus the elision rule cannot apply. In addition, the consonant cluster [mn] in *amn't* is followed by a consonant rather than a vowel in spelling, then [m] and [n] cannot form syllables with their neighboring vowels respectively like in the case of the word *omniscient* whose pronunciation is [ɒmˈnɪsɪənt].

Other alternatives

Apart from *amn't I*, there are some other alternatives that are available for *aren't I*: *am I not, an't I* and *ain't I*. The "bookish" *am I not* is often used in intensely formal situations, such as in literary language. For example, *Mrs. Hunnard gave delighted little cries. "If only I had a pretty voice -- I'm quite without tune, am I not?"* (From BNC, fiction: The diamond waterfall) However, this form is rarely used in colloquial speech, or it will sound completely unnatural and stuffy.

As for the obsolete *an't I*, there are two different explanations for its origin. According to Jespersen (1917, 118), in earlier times, *am not* became *an't* in a similar way that *cannot* and *shall not* became *can't* and *shan't*, with a change of the vowel from [æ] to [a:]. Crystal (2021, 37) provided another explanation for the origin of *an't*: The form *amn't*, which owns an awkward consonant cluster, was naturally simplified as *an't* for the reason of pronunciation convenience that [n] and [t] are both articulated on the alveolar ridge behind the top teeth. Nevertheless, whatever the origin of *an't*, this form is obviously outdated nowadays. Despite the fact that *an't I* was very popular a little earlier in modern English, as in Smollett and Dickens (Curme 1931, 137), we can find little support for this contraction form in actual modern usage.

With respect to the form *ain't I*, which is generally denounced as illiterate or vulgar, people seem to hold a more tolerant attitude. Regarding the origin of this form, my speculation is that it may have something to do with the Great Vowel Shift, through which the pronunciations of all Middle English long vowels were changed. Among them, the pronunciation of the long vowel [a:] experienced several changes ([æ:], [ɛ:] and [e:]), and finally shifted to [ei]. (Stockwell 2002) Given this fact, it is probable that the form *ain't I* came from the form *an't I* whose pronunciation was changed in some dialects during the Great Vowel Shift, and then when recording this form in writing after the Great Vowel Shift, it became *ain't I*. As Random House Webster's College Dictionary (2016, 36) describes, *ain't* is nonstandard except in some dialects, and it can not only replace *am not*, *are not* and *is not*, but also substitute *have not*, *has not*, *do not*, *does not* and *did not*. As regards the usage of *ain't*, this Dictionary (2016, 36) further claims that *ain't* is more common in uneducated speech than in educated, but it occurs occasionally in the informal speech of the educated, especially in the interrogative *ain't I* used as a substitute for the formal *am I not* or for the ungrammatical *aren't I* or for the awkward *amn't I*. However, from my standpoint, the fact that *ain't* can replace *am not*, *are not* and *is not* and even *have not*, *has not*, *do not*, *does not* and *did not* in speech will lead to greater chaos and complexity in actual usage. Superficially, using *ain't* in speech will reduce addressers' burden because it can express a variety of meanings in communication, and addressers can thus simplify their utterance. However, the consequence of addressers' "least effort" may be that addressees have to spend more time determining the appropriate meaning of *ain't* from a range of possibilities.

Conclusion

The present study aimed at investigating the widespread usage of "ungrammatical" contraction form *aren't I* in question tags from both quantitative and qualitative perspectives.

https://doi.org/10.1017/S0266078424000245 Published online by Cambridge University Press

According to large corpora, we can get a clear picture of the current frequency of use of the question tags *aren't I* and other alternatives (*amn't I*, *ain't I* and *an't I*) in modern English. *Aren't I* has already become a commonly accepted contraction form in both British English and American English. *Amn't I*, which is considered as the colloquial norm in Scottish and Irish English (Quirk et al. 1985; Crystal 2021), is rarely used in both American English and British English. *Ain't I*, which is mainly used in informal speech as a competitive alternative, has a much lower frequency of use than *aren't I* in both American English and British English. *An't I* is obsolete nowadays as we can find little support for this contraction form in actual modern usage. All these point to the fact that *aren't I* has already gained its dominance universally as a recognized standard form.

In our opinion, the reason why *aren't I* has taken hold as a recognized standard form around the globe lies in that the use of *aren't I* appears to be a smart coincidence to imply the potential double roles of "I" as both the addresser and the addressee in a monologue. Its "grammatical" alternative *amn't I*, however, is considered unacceptable in Standard English. The fact that most English-speaking people feel reluctant to use *amn't I* may result from its pronunciation problem: the awkward consonant cluster [mn] makes people difficult to articulate this contraction. Among other alternatives, the "bookish" *am I not* is rarely used in colloquial speech, whereas the outdated *an't I*, which was very popular a little earlier in modern English (Curme 1931, 137), is obviously falling into disuse nowadays. The relatively competitive alternative *ain't I*, which is generally viewed as illiterate or vulgar, is mainly used in informal speech. In addition, from my standpoint, the chaotic and complicated usage of *ain't* in American English may force listeners to spend more time determining the appropriate meaning of *ain't* from a range of choices. To summarize, the fact of the matter that *amn't I* is difficult to pronounce, *an't I* is old-fashioned and *ain't I* can only be used in informal situations, increases the popularity of *aren't I*.

The present findings can justify the usage of "ungrammatical" *aren't I* as a natural norm in both British English and American English. These findings open new research avenues alongside pedagogical and sociolinguistic implications for other similar "ungrammatical" language phenomena.

References

Crystal, D. 2021. *David Crystal's 50 Questions about English Usage*. Cambridge: Cambridge University Press.

Curme, G. O. 1931. *A Grammar of the English Language*. Vol. III. Essex: Verbatim.

Evans, B., and Evans, C. 1957. *Dictionary of Contemporary American Usage*. New York: Random House.

Fowler, H. W., and Butterfield, J. 2015. *Fowler's Dictionary of Modern English Usage*. Oxford: Oxford University Press.

Hands, P. 2018. *Collins COBUILD English Usage*. Beijing: The Commercial Press.

Hudson, R. 2000. "I Amn't." *Language* 76 (2): 297–323.

Jespersen, O. 1917. *A Modern English Grammar on Historical Principles*. Vol. II. Copenhagen: Munksgaard.

Jørgensen, E. 1979. "'Aren't I?' and Alternative Patterns in Modern English." *English Studies* 60 (1): 35–41.

Mbakop, A.W. N. 2022. "Question Tags in Cameroon English." *English Today* 38 (1): 27–37.

Partridge, E. 1973. *Usage and Abusage: A Guide to Good English*. Penguin.

Phillipps, K. C. 1984. *Language and Class in Victorian England*. Oxford: Basil Blackwood.

Quirk, R. *et al*. 1985. *A Comprehensive Grammar of the English Language*. New York: Longman.

Random House. 2016. *Random House Webster's College Dictionary*. Beijing: The Commercial Press.

Rothstein, E., and Rothstein, A. S. 2009. *English Grammar Instruction that Works! Developing Language Skills for all Learners*. California: Corwin Press.

Stockwell, R. 2002. "How Much Shifting Actually Occurred in the Historical English Vowel Shift?" In *Studies in the History of the English Language: A Millennial Perspective*, edited by D. Minkova, and R. Stockwell, 267–282. Berlin: Mouton de Gruyter.

Swan, M. 2019. *Practical English Usage*. Oxford: Oxford University Press.

Thomson, A. J., and Martinet, A. V. 1986. *A Practical English Grammar*. Oxford: Oxford University Press.

Tottie, G., and Hoffman, S. 2006. "Tag Questions in British and American English: The First Century." *Journal of English Linguistics* 34 (4): 283–311.

Yang, S. Z. *et al*. 2017. "附加疑问句 Aren't I [The Question Tag *Aren't I*]." *Overseas English* 4: 206–208.



MINGYOU XIANG is Professor at the University of International Business and Economics (School of International Studies). His research interests include pragmatics, functional linguistics and English grammar. Email: xiangmingyou@163.com



XIAO JIANG is currently a Ph.D. student at the University of International Business and Economics (School of International Studies). She is also a Lecturer at Nanjing Institute of Technology (School of Foreign Languages). Her research interest is in pragmatics. Email: 627780158@qq.com

https://doi.org/10.1017/S0266078424000245 Published online by Cambridge University Press