# 11 *'You Still Talking to Me?'* The Zero Auxiliary Progressive in Spoken British English Twenty Years On

*Andrew Caines, Michael McCarthy and Paula Buttery*

## 11.1 Introduction

The auxiliary verb is generally thought to be obligatory in progressive aspect constructions in English. However, we previously demonstrated that this is not always the case in spontaneous conversation, especially in certain lexico-syntactic contexts and for certain demographic groups (Caines, 2010; Caines & Buttery, 2010; Caines, McCarthy, & O'Keeffe, 2016). For instance, within the ten-million-word spoken section of the British National Corpus (Spoken BNC1994) there is a four-million-word conversational subcorpus—the so-called 'demographic' subsection—in which the auxiliary is absent for 34% of progressive aspect interrogatives with second person subjects (1).



(1) How you feeling now? [BNC1994 KBK 3474][1]

We coined the term 'zero auxiliary' to refer to such constructions, and found that it was being used in diffuse lexico-syntactic contexts[2]: not just open interrogatives (featuring a *wh*-word as in (1)) with second person subjects, but also closed interrogatives (2), other person-number subjects (3)–(4) and 'zero subject' declaratives (5).

(2) You been waiting long? [BNC1994 KDK 510]
(3) We opening them now? [BNC1994 KD0 5133]
(4) What they charging him with? [BNC1994 KDP 556]
(5) Yeah hold on just looking at something. [BNC1994 KD1 920]

We proposed that the zero auxiliary is not so much an 'error' or 'omission' but rather a variant of the progressive aspect construction in British English, deliberately selected as an alternative to the progressive with auxiliary verb (Caines, 2010). The zero auxiliary may be seen as a feature of spoken grammar, which is itself a fundamental part of people's linguistic competence (Biber, Johansson, Leech, Conrad, & Finegan, 1999; Carter & McCarthy, 2017).

These findings were based on the original British National Corpus (BNC World, 2001; 'BNC1994', henceforth), specifically the demographic section of spontaneous conversation, which features recordings collected in the early 1990s (the Spoken BNC1994DS). In this chapter, we investigate zero auxiliary use in the early-access Sample of the its successor ('the Spoken BNC2014S', henceforth)—more information about this corpus can be found in Chapter 1. Its comparable design and similar size meant we could use the Spoken BNC2014S to seek out any change in zero auxiliary use in the intervening two decades.

However, we found that the zero auxiliary has decreased in frequency in progressive interrogatives with pronominal subjects, according to our survey of the Spoken BNC2014S. Whereas in the Spoken BNC1994DS the zero auxiliary occurred in 25% of such progressive constructions, including a 34% occurrence rate for second person interrogatives, in the BNC2014 those rates drop to 6% and 9% respectively. Such a decrease in use is discussed in the context of a concurrent increase in zero auxiliary occurrence in written English, thanks to the emergence of the web domain, and the uncertainties of language sampling. Nevertheless the decrease is a pronounced one, and the sociolinguistic implication is that the age and class grading seen in the BNC1994 has flattened out in the BNC2014, such that there is no longer a strong association of younger working-class speakers with the zero auxiliary interrogative.



## 11.2 The Progressive Aspect in English

Cross-linguistically, auxiliary verbs "denote a closed class of verbs that are characteristically used as markers of tense, aspect, mood and voice" (Huddleston & Pullum, 2002, p. 102). The auxiliary verbs of English are distinguished from lexical verbs on account of various syntactic properties which lexical verbs do not possess. Foremost among them are the so-called NICE properties: negation, inversion, code and emphasis (for a discussion of these see Quirk, Greenbaum, Leech, & Svartvik, 1985, pp. 121–125; Huddleston & Pullum, 2002, pp. 92–102).

Within the set of auxiliary verbs, a further distinction is made between modal verbs and primary verbs (Quirk et al., 1985). Modal verbs fundamentally contribute meaning relating to concepts of mood such as "volition, probability and obligation" (Quirk et al., 1985, p. 120). Primary verbs have a range of uses including aspect (*be, have*), voice (*be*), and dummy support (*do*). We restrict our focus to the primary verbs, specifically *be* and *have*, as we study the use of progressive aspect auxiliary verbs, the paradigm for which is given in Table 11.1.

As shown in Table 11.1, the progressive aspect construction is standardly formed by periphrastic combination of auxiliary *be*, plus *have* if in the perfect aspect, and the *-ing* form of a lexical verb. Table 11.1 shows the full form auxiliary verb in declarative progressives; alternatively, the

*Table 11.1* The Progressive Aspect Construction in English

| | *Present* | *Past* | *Present Perfect* | *Past Perfect* |
|---|---|---|---|---|
| 1st singular | I am being | I was being | I have been being | I had been being |
| 2nd | You are being | You were being | You have been being | You had been being |
| 3rd singular | She is being | She was being | She has been being | She had been being |
| 1st plural | We are being | We were being | We have been being | We had been being |
| 3rd plural | They are being | They were being | They have been being | They had been being |

first auxiliary in all the cells may be contracted, except for the past tense column as there are typically no contracted forms for *be* in the past tense—e.g., *I'm being, you've been being, she'd been being*, etc. The progressive appears in interrogative form through inversion of the subject noun and auxiliary verb—e.g., *were we being, have they been being*. It may be negated through insertion of a negative adverbial between the auxiliary and *-ing* form—e.g., *he's not being, wasn't it being, had they not been being*.



The English progressive has been the topic of many recent publications reporting ongoing changes in its use. These relate to three broad research questions:

- The origins question: for the progressive in English, the precise answer to this issue remains a matter of debate (Denison, 1998).
- The function question: beyond a core meaning of continuousness, the progressive has developed a more complex set of meanings compared to the progressive in other languages (Dahl, 1985; Bybee, Perkins, & Pagliuca, 1994; Lee, 2007).
- The frequency question: this revolves around the progressive's "meteoric increase in frequency in the Modern English period" (Leech, Hundt, Mair, & Smith, 2009, p. 118; Hundt, 2004).

Putting aside the origins issue as a historical linguistic question beyond the scope of this study, we focus on the second and third questions, which relate to semantics and corpus linguistics. At its core, the progressive aspect refers to a happening with a limited duration which is not necessarily complete (Quirk et al., 1985). Beyond this central use the progressive has developed functions extending to future time (6)–(7) and states rather than events (8)–(10).

(6) What are you going to do? [BNC1994 KB3 491]
(7) You gonna let me have a go? [BNC1994 KBK 2676]

(8) He's being a bit sarcastic. [BNC1994 J90 109]
(9) A receptor, where G A B A is standing for gamma amino butyric acid. [BNC1994 J8K 197]
(10) We are living in a very sophisticated time. [BNC1994 KRT 807]

It has been proposed that the rapid increase in frequency of the progressive may also have facilitated its several functional extensions, especially the development of stative progressives in an aspect which was canonically reserved for non-stative use (Comrie, 1976). In addition to the increasing versatility of the progressive in terms of verb classes, the rapid change in progressive usage has allowed for experimentation with its spoken form. Notably there has been reduction of the velar nasal [ŋ] in *-ing* to variants closer to alveolar [n]. This reduction has been studied extensively in the sociolinguistic field, since it has been found to be a social marker (Campbell-Kibler, 2008). Meanwhile, contraction of the auxiliary verb has long been noted as the unmarked variant in spoken language (McElhinny, 1993). We also note that the pronoun and *-ing* form may regularly occur without auxiliary verb as a clausal complement or clausal subject (11), or may occur as an interrogative followed by a tag question which features an auxiliary verb (12).



(11) she was doing her main lessons and you maybe listening to it you'd pick up stuff [BNC2014 S32W 585]
(12) —still peeling
—yeah
—is she? [Spoken BNC2014 STN8 93]

We are not in a position to state which came first, but certainly the legitimate use of pronoun and *-ing* without auxiliary in other contexts may help habituate speakers to the zero auxiliary (and vice versa). Complete ellipsis may be a further consequence of the progressive being "in flux" (Comrie, 1976) and at least initially came with social marking.

## 11.3 The Zero Auxiliary Progressive

Here we discuss how the English progressive described in Section 11.2 has developed a zero auxiliary variant. In our previous study we found that the zero auxiliary is found throughout the progressive paradigm (Table 11.1) but occurs more frequently in certain lexico-syntactic contexts and is socially conditioned by speaker demographics (Caines, 2010).

How is this allowed to happen? Without the information carried by the auxiliary, how is it that communication is not impaired by its omission? In fact, in English, relatively little semantic information is carried by the auxiliary verb compared to the *-ing* participle in progressive constructions. The lexical content of the verb group is contained in the participle,

while aspect marking is shared between the participle and auxiliary or auxiliary verb group. So if that first auxiliary is omitted, aspectual and lexical information is still borne by the participle verb. What the (first) auxiliary does do is carry tense marking (recall Table 11.1), and for that reason past tense zero auxiliaries are rare except where context unambiguously situates time reference in the past (13)–(15), while zero auxiliaries are typically interpreted as present tense.

(13) They all doing it the other day. [BNC1994 K6N 779]
(14) When we talking earlier. [BNC1994 JAD 451]
(15) That wasn't what you saying. [BNC1994 F7U 806]

In speech production the auxiliary is usually unstressed and therefore "comparatively insignificant" (Jespersen, 1933, p. 100). The manner in which it is so frequently reduced and affixed to preceding phonological material as an enclitic is one outcome of this insignificance. According to the 'principle of least effort' in speech, "if a simple articulatory gesture works just as well as a complex one, there is a natural tendency to prefer it, thus rendering the articulatory movements in speech simpler" (Wells, 1982, p. 94). Thus the principle "leads us to tend to pronounce words and sentences in a way which involves the minimum of articulatory effort consistent with the need to maintain intelligibility" (Wells, 1982, p. 94). One instantiation of this principle is auxiliary contraction, in which the reduced auxiliary clitic attaches to the preceding item—often a subject pronoun. We may think of the zero auxiliary as an extreme outcome of this principle.



Bybee observes that frequency of use often goes hand in hand with the phonetic reduction described above, terming it the 'Reducing Effect' (2007). "Oft-repeated phrases . . . tend to reduce phonetically" since "repetition of neuromotor sequences leads to greater overlap and reduction of the component articulatory gestures", and as a result, "general reductive sound change occurs earlier in high-frequency words and special reduction occurs in very high-frequency words and phrases" and "thus frequency of use is one factor in explaining sound change" (Bybee, 2007, p. 11).

It should not come as a surprise that repeated use results in physical reduction, given previously made observations in the literature. Case studies have included an investigation by Bybee (1999) into *don't*, a chunk which was found to reduce more in highly frequent contexts such as post-*I* and with verbs such as *know* (it is at its most reduced in the sequence *I don't know*). Jurafsky and colleagues (2001) show that deletion of final /t/ and /d/ strongly correlates with the relative frequency of the word in question: high-frequency *want* and *good* versus low-frequency *let* and *heard*. Since the auxiliary verbs *be* and *have* are themselves high-frequency items in English, as is the progressive construction generally,

the obvious assumption to make—and it must remain an assumption since the appropriate historical speech data do not exist—is that zero auxiliary constructions emerged as a fast speech variant to the contracted or full forms. But of course frequency is not the only factor behind sound change: predictability of the word in context (Sanford, 2008) and the speaker's age and gender have also been proposed as influencing factors (Pluymaekers, Ernestus, & Baayen, 2005).

Causes of language change are difficult to identify, and very often there are multiple interacting factors. Based on the results of our previous study we proposed that the zero auxiliary emerged because of low semantic content, phonetic reduction of a high frequency form and a sociolinguistic distribution typical of patterns indicating 'covert prestige' (Trudgill, 1995). Moreover, we proposed that its frequency, lexical and sociolinguistic usage patterns and psycholinguistic processing indicated that it is a valid construction in its own right—an alternative to equivalents with a full or contracted auxiliary, which speakers select in appropriate contexts, contexts which we were able to model stochastically (Caines, 2010; Caines & Buttery, 2010; Caines, 2012).

In this way we move away from the tradition in Generative Grammar that there is an underlying logical form which may become degraded in performance in its surface realisation (Chomsky, 1965). Instead we align with the idea that speakers have a 'constructicon', or a "registry of constructions" (Fillmore, Lee-Goldman, & Rhodes, 2012), from which they make their selections according to linguistic and extra-linguistic variables (Lee, 2007; Sag, 2012). Two findings support this claim: firstly, that in a psycholinguistic study, subjects' performance measures for zero auxiliary stimuli were similar to those for other non-standard constructions often used in speech (e.g., *I wanna go*) but 'better' (faster, less errorful) than those for ungrammatical stimuli such as subject-verb agreement errors and word order transpositions (Caines, 2012). Secondly, frequency of use was conditioned by demographic variables—namely, social class and age group (not gender), such that younger working-class groups were found to use the zero auxiliary most often (Caines, 2010). This indicates to us that there is a social value attached to the zero auxiliary, which was being used in alternation with the full auxiliary form favoured by older middle-class groups. In this new study of the Spoken BNC2014S, we investigate whether the zero auxiliary has progressed, regressed or held steady in its frequency of use in spoken British English.



## 11.4 Corpus Studies Old and New

In this section we firstly recall our original study of auxiliary realisation in progressive constructions in the Spoken BNC1994 (Section 11.4.1), then report on our comparative study of the Spoken BNC2014S (Section 11.4.2). We found that the rate of zero auxiliary use has dramatically

fallen in the two decades between the two corpora, and that the sociolinguistic contours found in the former have flattened out in the latter.

### *11.4.1 Corpus Study 1: The British National Corpus 1994*

The BNC1994 was originally chosen for a study of progressive constructions for three reasons: (i) because it features spoken language, (ii) at the time it was the most recent large corpus of British English in existence, and (iii) because of its emphasis on balanced speaker recruitment with respect to a number of demographic variables including age, gender and social class (Crowdy, 1993).

Here we recap the results of our previous survey of auxiliary realisation in progressive aspect interrogatives in conversational British English (Caines, 2010). We comprehensively surveyed all subsections of the BNC1994 and found that the zero auxiliary occurs ubiquitously, even in writing, albeit to a lesser extent and only in transcriptions of speech and representations of dialogue in works of fiction. However, relative zero auxiliary frequencies were higher in the spoken section and so for the sake of a richer dataset we focused on that subcorpus ('the Spoken BNC1994').

In total we retrieved 93,253 progressive aspect constructions from the Spoken BNC1994 and annotated each one for a number of morpho-syntactic variables—namely, person and number of subject, nominal or pronominal, clause type, *wh*-word in subject noun phrase (or not), polarity, tense, perfectivity and auxiliary form (Caines & Buttery, 2012). Our annotation variables and values are more fully presented in Table 11.2.



We found that the zero auxiliary occurred across all speech genres contained in the Spoken BNC1994, even including the more formal settings such as meetings, broadcasts and lectures. Nevertheless, the locus of the zero auxiliary was found to be the four-million-word spontaneous conversation section, and indeed, that section offers the closest comparison to the Spoken BNC2014S in size and in terms of the way it was collected and the situational contexts of the recordings.

In Table 11.3 we present the results of our previous study of progressive constructions in the conversation section of the Spoken BNC1994, and furthermore have restricted our attention to interrogatives with pronominal subjects so as to keep the new study of a manageable size. Having been the basis of a doctoral thesis, the previous study also covered declaratives and all nominal subjects—pronouns and other nouns.

As can be seen in Table 11.3, the full auxiliary form was the predominant one for interrogatives in the Spoken BNC1994. The contracted auxiliary was the minority form (partly as it is restricted to *wh*-questions only, in which the auxiliary enclitic has a token to 'attach' to) and the zero auxiliary was found throughout the person and number paradigm—above all with second person, first person plural and third person plural

*Table 11.2* Progressive Aspect Construction Annotation Scheme

| *Variable* | *Value* | *Example* |
|---|---|---|
| Subject type | nominal | The computer is working |
| | pronominal | She is working |
| Subject person & number | 1st singular | I am working |
| | 2nd | You are working |
| | 3rd singular | It is working |
| | 1st plural | We are working |
| | 3rd plural | They are working |
| | zero | Just working today |
| Clause type | declarative | You are working |
| | interrogative | Are you working? |
| *wh*-word | true | When is he working? |
| | false | Is he working? |
| Tense | present | I am working |
| | past | I was working |
| Perfect aspect | present perfect | They have been working |
| | past perfect | They had been working |
| Polarity | positive | It is working |
| | negative | It is not working |
| Auxiliary | full | You are working |
| | contracted | You're working |
| | zero | You working |



*Table 11.3* Auxiliary Realisation for Progressive Aspect Interrogatives With Pronominal Subjects in the Conversational Section of the Spoken BNC1994

| | *Count* | *Full Aux (%)* | *Contracted (%)* | *Zero Aux (%)* |
|---|---|---|---|---|
| 1st person singular | 263 | 258 (98.1) | 2 (0.8) | 3 (1.1) |
| 2nd person sg/pl | 3,553 | 2,285 (64.3) | 55 (1.6) | 1,213 (34.1) |
| 3rd person singular | 1,316 | 613 (46.6) | 645 (49.0) | 58 (4.4) |
| 1st person plural | 483 | 363 (75.2) | 6 (1.2) | 114 (23.6) |
| 3rd person plural | 287 | 224 (78.1) | 5 (1.7) | 58 (20.2) |
| Progressive interrogatives | 5,902 | 3,743 (63.4) | 713 (12.1) | 1,446 (24.5) |

pronouns. Figure 11.1 illustrates the proportional auxiliary realisations given in parentheses in Table 11.3.

Thanks to the demographic metadata collected for participants in the conversational section of the Spoken BNC1994, we could associate each progressive interrogative in the corpus with the gender of its speaker

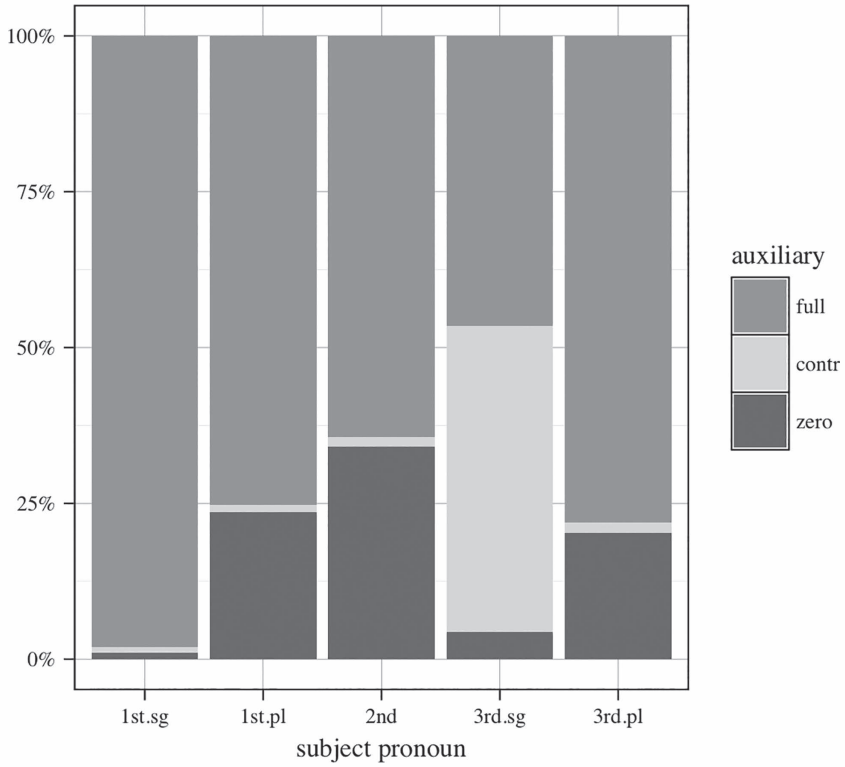


*Figure 11.1* Auxiliary Realisation for Progressive Aspect Interrogatives With Pronominal Subjects in the Conversational Section of the Spoken BNC1994

*Table 11.4* Gender of Speaker for Second Person Progressive Interrogatives in the Conversational Section of the Spoken BNC1994

| | *Count* | *Full Aux (%)* | *Contracted (%)* | *Zero Aux (%)* |
|---|---|---|---|---|
| Female | 2,072 | 1,346 | 30 | 696 |
| (n = 560) | | (65.0) | (1.4) | (33.6) |
| Male | 1,019 | 674 | 17 | 328 |
| (n = 511) | | (66.1) | (1.7) | (32.2) |
| Unclassified | - | 265 | 8 | 189 |
| 2nd person sg/pl | 3,553 | 2,285 | 55 | 1,213 |
| | | (64.3) | (1.6) | (34.1) |

(Table 11.4). We then ascertained that gender of speaker did not significantly affect zero auxiliary use in the second person progressive interrogative set, using mixed effects logistic regression[3] (Winter & Wieling, 2016) with auxiliary realisation transformed into a binary variable (zero auxiliary or not zero auxiliary), gender as fixed effect, speaker and verb

as random effects. No significant improvement came from adding gender as a fixed effect in comparison with a baseline model featuring speaker and verb random effects only (AIC = –1.13; $\chi2(1) = 0.87$, $p = 0.351$).[4]

Age and social class of speaker, on the other hand, did affect zero auxiliary use. We show auxiliary realisation counts in Tables 11.5 and 11.6, and illustrate these demographic variables in Figures 11.2 and 11.3. Mixed effects logistic regression indicated that age as a fixed effect (centred values) brought significant improvement over the baseline model (random effects only): AIC = 3.26; $\chi2(1) = 5.26$, $p = 0.022$. Likewise, social class as a fixed effect constructed a significantly better model for the zero auxiliary than the baseline: AIC = 9.77; $\chi2(3) = 17.8$, $p = 0.0014$.

Social class information was provided for the Spoken BNC1994 conversation section according to the Social Grade classification, developed for British society in the 1950s by the National Readership Survey, a market research organisation. Classifications are made on a six-point scale (A, B, C1, C2, D, E) on the basis of the occupation of the chief income earner in a household, with A and B representing higher managerial or professional roles, C1 junior managerial or professional roles, C2 skilled manual workers, D semi- and unskilled workers, and E casual or lowest grade workers. The social grades in Spoken BNC1994 were actually given on a four-point scale, following the convention to merge the top two and bottom two grades, AB and DE, such that AB is taken as a proxy for middle class, C1 for lower middle class, C2 for upper working class, and DE for working class.



*Table 11.5* Age Group of Speaker for Second Person Progressive Interrogatives in the Conversational Section of the Spoken BNC1994

| | *Count* | *Full Aux (%)* | *Contracted (%)* | *Zero Aux (%)* |
|---|---|---|---|---|
| 0–9 years | 93 | 63 | 2 | 28 |
| (n = 78) | | (67.7) | (2.2) | (30.1) |
| 10–19 years | 684 | 436 | 9 | 239 |
| (n = 251) | | (63.8) | (1.3) | (34.9) |
| 20–29 years | 411 | 238 | 5 | 168 |
| (n = 148) | | (57.9) | (1.2) | (40.9) |
| 30–39 years | 688 | 426 | 12 | 250 |
| (n = 139) | | (61.9) | (1.8) | (36.3) |
| 40–49 years | 432 | 306 | 4 | 122 |
| (n = 128) | | (70.8) | (0.9) | (28.3) |
| 50–59 years | 380 | 259 | 8 | 113 |
| (n = 85) | | (68.2) | (2.1) | (29.7) |
| 60–69 years | 136 | 105 | 2 | 29 |
| (n = 69) | | (77.2) | (1.5) | (21.3) |
| 70–79 years | 181 | 134 | 3 | 44 |
| (n = 47) | | (74.0) | (1.7) | (24.3) |
| Unclassified | 522 | 299 | 10 | 213 |
| 2nd person sg/pl | 3,553 | 2,285 | 55 | 1,213 |
| | | (64.3) | (1.6) | (34.1) |

Though Table 11.5 and Figure 11.2 indicate a U-shaped pattern of age group use of the zero auxiliary in Spoken BNC1994, if we group the first four age groups against the last four age groups it is apparent that the zero auxiliary is associated more with the younger speakers in the corpus than the older ones. The fact that its use is driven by speakers in their twenties and thirties above all will become important in our subsequent study of the Spoken BNC2014, as corpus statistics from the newer corpus indicate that the age group pattern we see in Figure 11.2 does not continue into the 21st century.

Similarly, Table 11.6 and Figure 11.3 indicate that the zero auxiliary is favoured by working-class rather than middle-class speakers. These are patterns typical of linguistic features with covert prestige, according to sociolinguistic theory (Trudgill, 1995). In other words, those speaker groups with less reason to identify with standard forms—generally set by older middle-class speaker groups—are those who tend to use the zero auxiliary progressive more often.

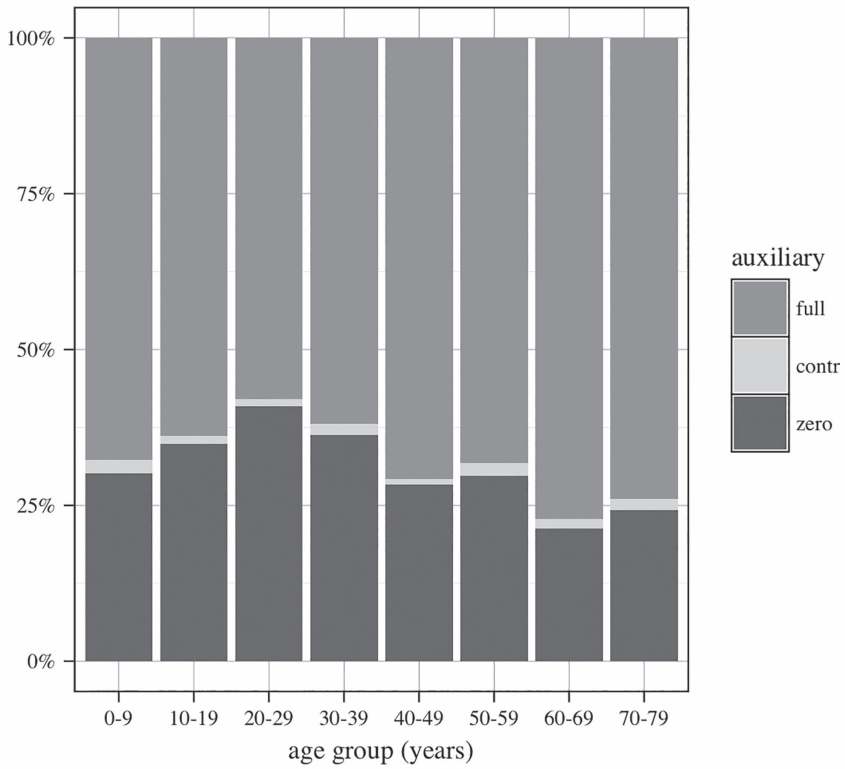


*Figure 11.2* Age Group of Speaker for Second Person Progressive Interrogatives in the Conversational Section of the Spoken BNC1994

*Table 11.6* Social Class of Speaker for Second Person Progressive Interrogatives in the Conversational Section of the Spoken BNC1994

| | *Count* | *Full Aux (%)* | *Contracted (%)* | *Zero Aux (%)* |
|---|---|---|---|---|
| Middle | 612 | 458 | 8 | 146 |
| (AB, n = 88) | | (74.8) | (1.3) | (23.9) |
| Lower middle | 697 | 462 | 11 | 224 |
| (C1, n = 116) | | (66.3) | (1.6) | (32.1) |
| Upper working | 665 | 401 | 13 | 251 |
| (C2, n = 99) | | (60.3) | (2.0) | (37.7) |
| Working | 316 | 166 | 6 | 144 |
| (DE, n = 60) | | (52.5) | (1.9) | (45.6) |
| Unclassified | 1,263 | 798 | 17 | 448 |
| 2nd person sg/pl | 3,553 | 2,285 | 55 | 1,213 |
| | | (64.3) | (1.6) | (34.1) |

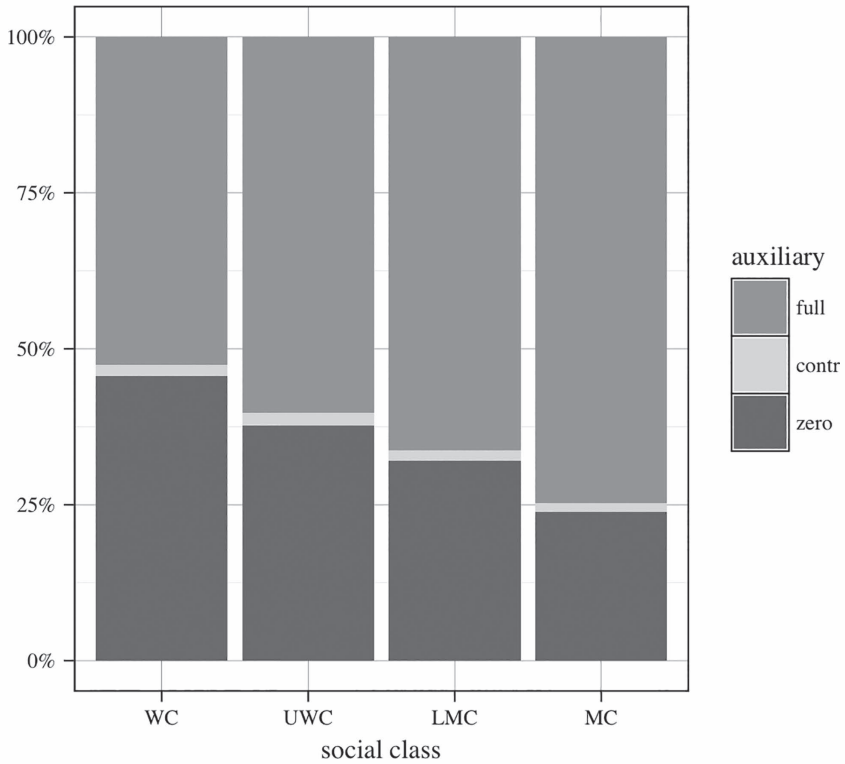


*Figure 11.3* Social Class of Speaker for Second Person Progressive Interrogatives in the Conversational Section of the Spoken BNC1994

As an investigation of lexical effects on zero auxiliary use, we conducted a 'collostructional analysis' (Stefanowitsch & Gries, 2003; Gries, Hampe, & Schönefeld, 2005) to confirm whether the zero auxiliary associated strongly with any specific verbs. Specifically, we employed the collostructional technique known as 'collexeme analysis'—"the study of collocation with association measures" (Gries, 2015)—with an R program which has been made publicly available for this purpose (Gries, 2014)[5]. The program takes corpus, verb and construction frequencies as the input, and outputs a table of association strength for each verb based on the observed and expected frequencies of that verb inside and outside the target construction. Thus we are presented with the verbs ranked on a scale of attraction to the zero auxiliary progressive. The benefit of collexeme analysis is that it does not assume a normal distribution or homogeneity of variance, properties rarely encountered in language (Mandelbrot, 1966). The analysis is underpinned by the Fisher-Yates exact test and results are presented in the form, negative log to the base of ten of the one-tailed p-value computed by the test ($p_{-\log,10}$).

As shown in Table 11.7, several verbs are highly attracted, and repulsed, by the second person zero auxiliary progressive interrogative. *Doing* and *going/gonna* are chief among these on the attraction side, whilst *saying, taking* and *working* are very frequent in progressive constructions but have a negative association with the zero auxiliary variant. Furthermore, if we view the zero auxiliary set on a constructional level, we find that



*Table 11.7* Collexeme Analysis of Selected Verbs and Zero Auxiliary Second Person Progressive Interrogatives in the Spoken BNC1994 (* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$)

| *Verb* | *In Progressive Constructions in the Spoken BNC1994* | *Zero Auxiliary 2nd Person Interrogative* | | *Relation* | $p_{-log,10}$ |
|---|---|---|---|---|---|
| | | *Expected Frequency* | *Observed Frequency* | | |
| doing | 6,243 | 90 | 295 | attraction | 74.6*** |
| going/gonna | 29,161 | 419 | 553 | attraction | 14.7*** |
| laughing | 143 | 2 | 9 | attraction | 3.62*** |
| calling | 146 | 2 | 7 | attraction | 2.27** |
| having | 1,691 | 24 | 34 | attraction | 1.47* |
| looking | 2,424 | 35 | 41 | attraction | 0.79 |
| getting | 3,031 | 44 | 45 | attraction | 0.37 |
| coming | 2,637 | 38 | 32 | repulsion | 0.73 |
| talking | 3,117 | 45 | 33 | repulsion | 1.43* |
| thinking | 1,257 | 18 | 8 | repulsion | 2.20** |
| trying | 2,192 | 32 | 11 | repulsion | 4.74*** |
| working | 1,598 | 23 | 5 | repulsion | 5.24*** |
| taking | 1,089 | 16 | 1 | repulsion | 5.65*** |
| saying | 3,814 | 55 | 19 | repulsion | 7.97*** |

nine constructional schemas account for more than half the zero auxiliary second person progressive interrogatives in the Spoken BNC1994: namely, *what you doing, how you doing, what you laughing for/at, what you looking for/at, what you talking about, what you having, where you going, you going/gonna* V, WH *you going/gonna* V. Such a distribution—many contributed by the few—is not surprising given the Zipfian nature of human languages (Zipf, 1965 [1935]), but it does remind us that the zero auxiliary has not spread evenly through the lexicon nor the constructicon. Instead there are prototypical high-frequency contexts for the zero auxiliary, which seemingly spreads by analogy to other contexts, albeit at lower frequencies for now. In the next section we verify whether the zero auxiliary has diffused further across the lexicon and construction of English.

### *11.4.2 Corpus Study 2: The Early Access Sample of the Spoken BNC2014*

In order to carry out a comparative study of contemporary zero auxiliary use with our previous study, we were able to query the early access Sample of the Spoken BNC2014 ('the Spoken BNC2014S'). Access to the Sample was granted via Lancaster University's CQPweb server (Hardie, 2012) and therefore we retrieved utterances of interest using part-of-speech searches in CQP ('corpus query processor') syntax.



Since the zero auxiliary was more likely to be found in interrogative rather than declarative clauses (Caines & Buttery, 2012), we opted to retrieve only interrogatives from the Spoken BNC2014S for our comparative study with its predecessor. As an additional constraint we focused on pronominal subjects only, again on the basis of our finding that the zero auxiliary occurred more often with pronouns as subject than with other noun types (Caines & Buttery, 2012). Every CQP search therefore centred around the juxtaposition of a pronoun and an *-ing* participial form, optionally with intervening negative and adverbial items. These searches were designed to capture progressive interrogatives with full auxiliaries (16), contracted auxiliaries (17), and zero auxiliaries (18), negated (19), adverbials (20), the past tense (21) and the perfect aspect (22).

(16) what time are you going back on Saturday?[6] [BNC2014 STXT 391]
(17) what's she doing? [BNC2014 S9MK 460]
(18) what time you going home? [BNC2014 SXKQ 63]
(19) why aren't I counting? [BNC2014 SHTW 623]
(20) Are you just googling it? [BNC2014 SDR9 620]
(21) why were you watching Home and Awa-? oh [BNC2014 S37E 447]
(22) has she been snorkelling? [BNC2014 SCA5 512]

With our set of CQP queries[7] we retrieved a corpus of 5,674 text strings which were potentially progressive aspect interrogatives. We ran

a supervised annotation procedure in R (R. Core Team, 2016) to confirm whether the *-ing* form indeed formed part of a progressive construction. Inevitably, our searches accumulated a bit of noise due to the syncretism of *-ing* forms in adjectival (23) and nominal (24) functions, alongside the verbal participle which features in the progressive. Plus, the *-ing* form often occurs in non-finite clause complements (25).

(23) —they're really good thank you
—cracking
—mm [BNC2014 SEKZ 636]
(24) is it knitting tonight? [BNC2014 S48K 926]
(25) someone else is gonna have to listen to me swearing [BNC2014 SHSL 518]

For the 3,873 hits confirmed as progressive interrogatives, we proceeded to code subject and clause properties of the construction in question using the same variables and values shown in Table 11.1. In this way our corpus was reduced to 3,674 progressive aspect interrogatives with pronominal subjects. In Table 11.8 we show auxiliary realisation patterns for these progressive interrogatives by subject person and number.

Firstly, we note that we retrieved many fewer progressive interrogatives with pronoun subjects from the Spoken BNC2014S (3,674) than from the conversation section of its predecessor (5,902), even though both corpora contain four to five million words and both feature approximately 67,000 verbal *-ing* forms. So the disparity is either due to our undertaking a less-than-comprehensive search of the Spoken BNC2014S, or because many fewer progressive interrogatives were uttered—especially with second person, third person singular and first person plural pronominal subjects (cf. Table 11.3). We cannot establish which is the



*Table 11.8* Auxiliary Realisation for Progressive Aspect Interrogatives With Pronominal Subjects in the Spoken BNC2014S

| | *Count* | *Full Aux (%)* | *Contracted (%)* | *Zero Aux (%)* |
|---|---|---|---|---|
| 1st person singular | 268 | 265 (98.8) | 2 (0.8) | 1 (0.4) |
| 2nd person sg/pl | 2,091 | 1,896 (90.7) | 8 (0.4) | 187 (8.9) |
| 3rd person singular | 668 | 477 (71.4) | 181 (27.1) | 10 (1.5) |
| 1st person plural | 368 | 352 (95.6) | 1 (0.3) | 15 (4.1) |
| 3rd person plural | 279 | 273 (97.8) | 0 (0) | 6 (2.2) |
| Progressive interrogatives | 3,674 | 3,263 (88.8) | 192 (5.2) | 219 (6.0) |

case without undertaking an analysis of every utterance in the Spoken BNC2014S containing a verbal *-ing* form, something which time restrictions do not allow for.

However, if we accept this as a faithful sample of progressive interrogatives in the Spoken BNC2014S, the second noticeable difference with the BNC1994 is the much lower rate of zero auxiliary occurrence: 6.0% overall, compared to 24.5% (cf. Table 11.3). Zero auxiliary frequencies are down across all persons and numbers compared to the BNC1994, but especially the second person, first person plural and third person plural pronouns. We illustrate these differences in Figure 11.4, repeating BNC1994 auxiliary realisations (Figure 11.1) alongside those for the Spoken BNC2014S, for convenience.

We now turn to our demographic variables of interest—gender, age and social class—repeating our survey of the BNC1994 in which we investigated second person progressive interrogatives only. Table 11.9 shows the gender of speaker by auxiliary realisation for second person progressive interrogatives in the Spoken BNC2014S. As in the Spoken BNC1994, gender does not affect rates of auxiliary use (cf. Table 11.4; AIC = –1.97; $\chi 2(1) = 0.026$, $p = 0.871$).

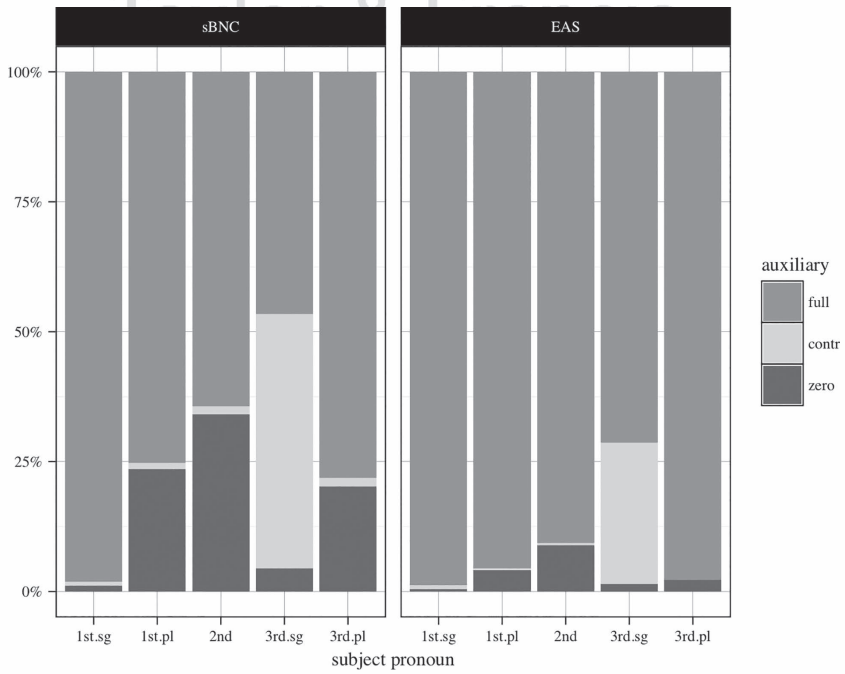


*Figure 11.4* Auxiliary Realisation for Progressive Aspect Interrogatives With Pronominal Subjects in the Conversational Section of Spoken BNC2014DS (left) and the Spoken BNC2014S (right).

*Table 11.9* Gender of Speaker for Second Person Progressive Interrogatives in the Spoken BNC2014S

| | *Count* | *Full Aux (%)* | *Contracted (%)* | *Zero Aux (%)* |
|---|---|---|---|---|
| Female (n = 207) | 1,400 | 1,271 (90.8) | 5 (0.4) | 124 (8.8) |
| Male (n = 171) | 691 | 625 (90.5) | 3 (0.4) | 59 (9.1) |
| 2nd person sg/pl | 2,091 | 1,896 (90.7) | 8 (0.4) | 187 (8.9) |

Next, we inspect how the age group and social class[8] of speaker are found to affect zero auxiliary use in the Spoken BNC2014S, in Tables 11.10 and 11.11, starting with age group.

As we see in Table 11.10, age group does not seem to have the same effect on zero auxiliary occurrence in the Spoken BNC2014S as it did in the BNC1994 (AIC = –1.92; $\chi 2(1) = 0.08$, $p = 0.777$). Whereas before the younger age groups tended to use the zero auxiliary more than the older age groups, in the Spoken BNC2014S we instead find that zero auxiliary use is fairly constant across the ages of 10 to 79 years (Figure 11.5). We note also that zero auxiliary rates are lower across the board, and that while there are older age groups in the Spoken BNC2014S than there were in BNC1994 (max. 90–99 rather than 70–79), there are also far fewer speakers aged 19 and under (cf. Table 11.5). Indeed, the original Spoken BNC1994 contains a large subset of teenage speech because it features the Bergen Corpus of London Teenage Language, compiled in 1993 (COLT; Stenström, Andersen, & Kristine Hasund, 2002). COLT contains 445,000 words from thirty-one contributors aged 10–19 (with coincidental utterances by pre- and post-adolescent interlocutors) and the recordings by contributors aged 16 or under were included in the BNC1994, as recruitment for the demographic component had included speakers aged 15 upwards (Burnard, 2007).



In Table 11.11 we see less of an effect of the speaker's social class on zero auxiliary use in the Spoken BNC2014S compared to the BNC1994 (cf. Table 11.6; AIC = –1.92; $\chi 2(3) = 1.95$, $p = 0.58$). Figure 11.6 confirms that, whereas in the BNC1994 there was a pronounced gradient for social class, the rate of zero auxiliary use in the Spoken BNC2014S has flattened across classes. This is symptomatic of the general reduction in zero auxiliary use in the Spoken BNC2014S, and certainly suggests that the variant's covert prestige has reduced in the intervening years since the BNC1994 was collated.

As for the distribution of speakers across the social classes, it is somewhat problematic for the representativeness of the corpus that there were so few contributors to the Spoken BNC2014S from the C2 band. However, it is doubtful that this has affected our results, as the line from C1 to DE is fairly flat, and we can assume from this that even with more C2 speakers the zero auxiliary rates would be fairly similar, give or take

*Table 11.10* Age Group of Speaker for Second Person Progressive Interrogatives in the Spoken BNC2014S[9]

| | *Count* | *Full Aux (%)* | *Contracted (%)* | *Zero Aux (%)* |
|---|---|---|---|---|
| 0–9 years (n = 3) | 1 | 1 (100) | 0 (0) | 0 (0) |
| 10–19 years (n = 21) | 100 | 91 (91.0) | 0 (0) | 9 (9.0) |
| 20–29 years (n = 142) | 985 | 888 (90.2) | 5 (0.5) | 92 (9.3) |
| 30–39 years (n = 54) | 413 | 384 (93.0) | 1 (0.2) | 28 (6.8) |
| 40–49 years (n = 44) | 149 | 132 (88.6) | 0 (0) | 17 (11.4) |
| 50–59 years (n = 41) | 166 | 144 (86.7) | 0 (0) | 22 (13.3) |
| 60–69 years (n = 48) | 179 | 168 (93.9) | 1 (0.6) | 10 (5.5) |
| 70–79 years (n = 13) | 64 | 57 (89.1) | 0 (0) | 7 (10.9) |
| 80–89 years (n = 7) | 6 | 6 (100) | 0 (0) | 0 (0) |
| 90–99 years (n = 2) | 3 | 3 (100) | 0 (0) | 0 (0) |
| Unclassified | 25 | 22 | 1 | 2 |
| 2nd person sg/pl | 2,091 | 1,896 (90.7) | 8 (0.4) | 187 (8.9) |



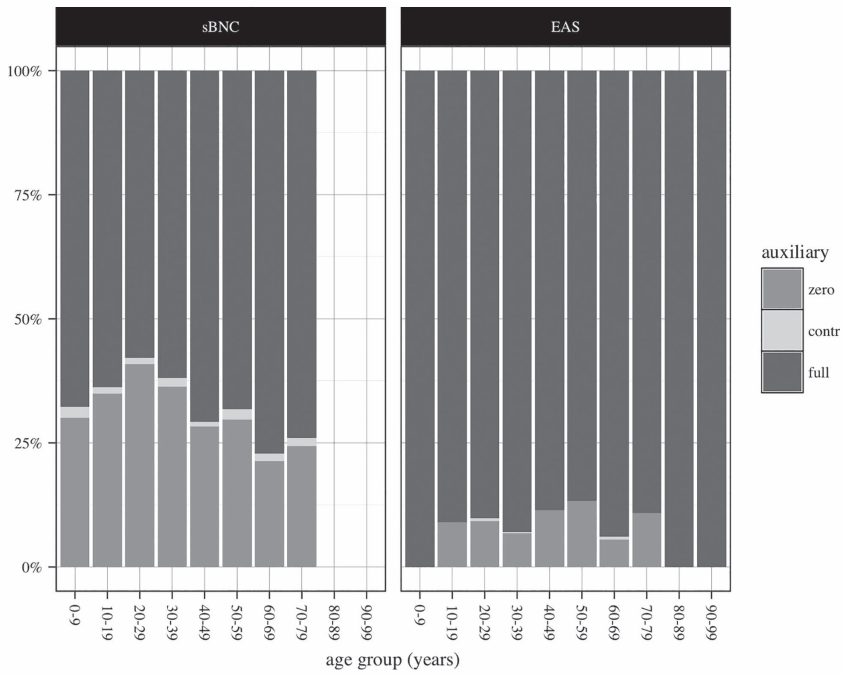


*Figure 11.5* Age Group of Speaker for Second Person Progressive Interrogatives in the Conversational Section of Spoken BNC1994DS (Left) and the Spoken BNC2014S (Right)

*Table 11.11* Social Class of Speaker for Second Person Progressive Interrogatives in the Spoken BNC2014S

| | *Count* | *Full Aux (%)* | *Contracted (%)* | *Zero Aux (%)* |
|---|---|---|---|---|
| Middle | 1,188 | 1,078 | 6 | 105 |
| (AB, n = 145) | | (90.7) | (0.5) | (8.8) |
| Lower middle | 248 | 228 | 0 | 20 |
| (C1, n = 53) | | (91.9) | (0) | (8.1) |
| Upper working | 51 | 49 | 0 | 2 |
| (C2, n = 11) | | (96.1) | (0) | (3.9) |
| Working | 580 | 520 | 1 | 59 |
| (DE, n = 167) | | (89.6) | (0.2) | (10.2) |
| Unclassified | 23 | 21 | 1 | 1 |
| 2nd person sg/pl | 2,090 | 1,896 | 8 | 187 |
| | | (90.7) | (0.4) | (8.9) |

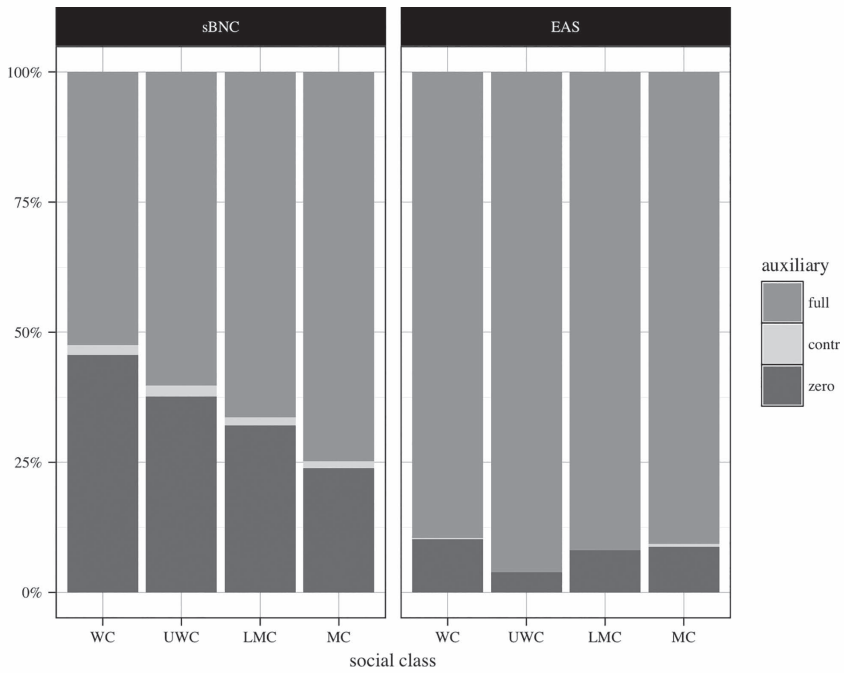


*Figure 11.6* Social Class of Speaker for Second Person Progressive Interrogatives in The Conversational Section of the Spoken BNC1994 (Left) and the Spoken BNC2014S (Right)

some idiosyncrasies and other group dynamics. Moreover, the inference of social class from occupation alone is a questionable step—indeed the assignment of social class to an individual is highly fraught, full stop (Savage, 2010)—and we can be sure that a balance of genders, regions and age groups is more important to corpus representativeness than what

is essentially a range of occupations. But that is the convention and data available, both in the BNC1994 and Spoken BNC2014S, and even though we are comparing 'social class' via proxy from occupation, the comparison is a fair one and certainly the BNC2014 results show that the zero auxiliary gradient previously seen has almost completely flattened out.

Now, if we introduce *corpus* as a fixed effect through mixed effects logistic regression—Spoken BNC1994 versus Spoken BNC2014S—and include progressive interrogatives from both corpora, we find that it significantly improves on a baseline model of random effects only (speaker and verb): AIC = 568; $\chi2(1) = 570$, $p < 0.001$. Finally, additional small improvement is brought by introducing age and social class (with interaction) to the corpus model: AIC = 2.47; $\chi2(9) = 20.5$, $p = 0.015$. Our regression analyses confirm that 'time' (taking the two corpora as temporal proxy) is the strongest predictor of our dependent variable, use of the zero auxiliary in progressive interrogatives.

We repeat the collexeme analysis conducted on zero auxiliary second person progressive interrogatives in the BNC1994 (Stefanowitsch & Gries, 2003; Gries, 2014). Though there are many fewer zero auxiliaries in the Spoken BNC2014S, the method is robust as it measures association based on observed and expected frequencies calculated from the total number of zero auxiliaries and the frequency of any given verb in all progressive constructions in the corpus. Recall that in the BNC1994, *doing* and *going/gonna* were the most strongly attracted verbs to the zero auxiliary construction (Table 11.7). Here we again see that the same two verbs are top-ranked with an attraction to the zero auxiliary, albeit with their order reversed (Table 11.12). In contrast, *working*, *looking*, *getting*



*Table 11.12* Collexeme Analysis of Selected Verbs and Zero Auxiliary Second Person Progressive Interrogative in the Spoken BNC2014S (* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$)

| *Verb* | *In Progressive Constructions in the Spoken BNC2014S* | *Zero Auxiliary 2nd Person Interrogative* | | *Relation* | $p_{-log,10}$ |
|---|---|---|---|---|---|
| | | *Expected Frequency* | *Observed Frequency* | | |
| going/gonna | 14,386 | 39 | 77 | attraction | 9.65*** |
| doing | 5,089 | 14 | 36 | attraction | 6.96*** |
| turning | 111 | 0.3 | 2 | attraction | 1.43* |
| running | 440 | 1 | 3 | attraction | 0.92 |
| using | 511 | 1.4 | 2 | attraction | 0.39 |
| talking | 1,958 | 5 | 6 | attraction | 0.35 |
| coming | 1,829 | 4.9 | 5 | attraction | 0.25 |
| having | 2,229 | 6 | 5 | repulsion | 0.36 |
| getting | 2,706 | 3.4 | 6 | repulsion | 0.41 |
| working | 1,266 | 3.4 | 2 | repulsion | 0.48 |
| looking | 1,828 | 5 | 3 | repulsion | 0.58 |

and *having* occur frequently in progressive constructions in the Spoken BNC2014S but are found to hold a negative association with the second person zero auxiliary. The association of the latter three has switched from 'attraction' to 'repulsion', from the Spoken BNC1994 to the Spoken BNC2014S, indicative of the general decrease in frequency of the zero auxiliary variant.

## 11.5 Discussion

Our comparison of auxiliary realisations in progressive constructions in the Spoken BNC2014S and Spoken BNC1994 clearly indicates a sharp fall in use of the zero auxiliary variant in the intervening two decades. Both corpora were designed and collated in a very careful manner; we take them as faithful snapshots of spoken British English in the 1990s and 2010s, and consequently conclude that the zero auxiliary has declined in general use, in progressive interrogatives with pronominal subjects at least.

Nevertheless, we must keep in mind that all corpora are samples of language as it is actually and generally used. If we consider our studies of auxiliary realisation in progressive constructions as measurements from these samples, we can consider the possibility that either reading—from the Spoken BNC1994 or the Spoken BNC2014S—might be an anomaly, since language is not normally distributed and highly idiosyncratic. For instance, just eighty-two of the 288 speakers who uttered a progressive interrogative in the Spoken BNC2014S produced the 226 zero auxiliaries we found there. Furthermore, 50% of those 226 zero auxiliaries are produced by only eleven speakers, the other seventy-one speakers producing no more than five zero auxiliaries each. A similar Zipfian distribution was found for zero auxiliaries and the speakers who uttered them in the Spoken BNC1994. Thus, even such large corpora as the conversational Spoken BNC1994 and Spoken BNC2014S are prone to the idiosyncrasies of their contributors. But still, we view them each as representative of their time, and note that zero auxiliary frequencies in the Spoken BNC1994 may have been extraordinarily high as much as those in the Spoken BNC2014S may be extraordinarily low. One further caveat is that the Spoken BNC1994 contained a large subset of teenage speech (the COLT corpus), and the Spoken BNC2014S is noticeably short of C2 speakers—not that the outdated Social Grade scale is a wholly satisfactory way to represent the complex British class system.

Nevertheless, accepting the corpora as representative and the results as a fair comparison, it is apparent that the zero auxiliary use has markedly decreased in frequency since the early 1990s. We find this outcome a little surprising as the variant has, for instance, found its way into informal genres of writing. Anecdotal examples include zero auxiliaries from Twitter (26)–(28).

(26) How you finding our new #BBCIntroducing time slot? Which show are you listening to?
"BBC Introducing" (@bbc_introducing) 2013-01-05 20:47.[10]
(27) When we gonna have cameras behind phone screens?
"Jack Garratt" (@jackgarratt) 2016-10-30 12:05.[11]
(28) What you doing?
"Matty." (@matty_selley) 2016-10-30 09:44.[12]

Indeed, a brief survey of 5.8 million Twitter conversations—series of tweets between 'interlocutors' collected using the unsupervised method described by Ritter and colleagues (2010)—containing 59 million tokens confirms that zero auxiliary rates are higher than those found in the Spoken BNC2014S and indeed are close to what they were in the Spoken BNC1994. We searched for fourteen open interrogatives—featuring a *wh*-word—with verbs both attracted to and repulsed by the zero auxiliary[13] (Table 11.12) and retrieved 12,362 progressive interrogatives in which 62.7% had the full auxiliary form, 1.9% had a contracted form and 35.4% were zero auxiliaries. It may be, then, that the zero auxiliary, having originated in the spoken domain—as suggested by a comparison of its frequency in the Spoken BNC1994 and its written counterpart (Caines, 2010)—has more recently transitioned away from speech to the more informal genres of the written medium, primarily the digital domains in which the pressure to communicate efficiently is not only physical, as it is in speech, but also at times explicitly set. For instance, SMS texts have a per-message limit of 160 characters. In many parts of the world this matters, since 'pay as you go' usage of mobile phones remains the majority use case, even if contracted usage with limitless SMS messaging is the norm in the United Kingdom.[14] Similarly, though free of charge, Twitter users are restricted to 140 character posts at a time. In these domains, omission of semantically-light material is beneficial and potentially brings both time and cost savings.



In sociolinguistic terms, the zero auxiliary has become less of a marker of age group and social class, instead being used more equally across speaker demographics. These findings are important not only for the sake of understanding language use, language change and spoken grammar, but they also have pedagogical implications. Namely, whereas before we might have advised introducing the zero auxiliary to learners of English as an informal speech variant (Caines et al., 2016), now we would say the same while not worrying about sociolinguistic consequences for the learner. We would also encourage its use in less formal written genres such as Internet communication, particularly for collostructions with high attraction values.

## Acknowledgements

This chapter reports on research supported by Cambridge English, University of Cambridge. We gratefully acknowledge the help and advice of

Jim Blevins, Anne O'Keeffe, Steve Walsh and Martijn Wieling. We thank the editors of this volume and two anonymous reviewers for their feedback which greatly improved the chapter.

## Notes

1 All rights in the texts cited from the British National Corpus are reserved (Oxford University Computing Services on behalf of the BNC Consortium). Each extract is followed by 'BNC1994', a unique text identifier and a sentence number.
2 We wish to state at the outset that we accept the judgements of the trained transcribers of both corpora used here in terms of the presence or absence of auxiliary verbs. By 'zero' auxiliary we mean 'zero' or 'near-zero' presence of an auxiliary verb form. We have listened to a sample of BNC1994 recordings through the Audio BNC project (www.phon.ox.ac.uk/AudioBNC), and concurred in all cases with the transcribers that the auxiliary is not there—or at least, we cannot perceive it. We are sometimes asked whether the zero auxiliary could not in fact be an intended but barely pronounced form, especially for the English form *are*. This could well be the case, but even so would speak of the option to whittle away at auxiliary forms, and does not alter what was actually produced and perceived, whatever the speaker's intentions (the zero auxiliary and speaker intention is an issue addressed in Caines, 2012).
3 All regression analyses reported in this chapter use R (R. Core Team, 2016) and the lme4 package (Bates et al., 2015).
4 Where AIC is 'Akaike information criterion'—an information-theoretic measure of the relative quality of statistical models on the same dataset—and $\chi 2(1)$ represents a chi-square test with one degree of freedom in an analysis of variance (ANOVA) between the baseline and gender regression models.
5 Even though we originally ran a collexeme analysis with a previous version of the Coll.analysis program (Coll.analysis 3.2a. A program for R for Windows 2.x.) we re-ran the analyses with the latest version of the program, as cited (www.linguistics.ucsb.edu/faculty/stgries/teaching/groningen; accessed 2017–2002–2016).
6 All rights in the texts cited from the Spoken BNC2014 Early Access Subset are reserved (Cambridge University Press). Each extract is followed by 'BNC2014', a unique text identifier and a line number.
7 Queries took one of three base forms—(1) [pos = "PP.*"] [pos = "V.G"], (2) [pos = "PP.*"] [pos = "VBN"] [pos = "V.G"], (3) [pos = "PP.*"] [word = "gon"] [word = "na"]—with variants to include all combinations of negative [pos = "XX"] and adverb tokens [pos = "R.*"].
8 Social class was coded in the Spoken BNC2014S according to the National Statistics Socio-Economic Classification (NS-SEC), used for the UK population census since 2001. NS-SEC labels were mapped back to the Social Grade classes used in the BNC1994 (Love, Dembry, Hardie, Brezina & McEnery, 2017).
9 Early contributors to the Spoken BNC2014S were asked to place themselves in age ranges which slightly differ from the ones we used for our BNC1994 study: 0–10 years, 11–18 and 19–29 rather than 0–9, 10–19 and 20–29. Later, speakers were instructed to declare their exact age in years; where they did so and were found to be aged 10 or 19 we adjusted the age group counts accordingly.
10 http://twitter.com/bbc_introducing/status/287661510598209536 (accessed 2016-11-01).

11 https://twitter.com/JackGarratt/status/792683772013600768 (accessed 2016-11-01).
12 https://twitter.com/matty_selley/status/792648472130297857 (accessed 2016-11-01).
13 Namely, *what (are) you* followed by *going, gonna, doing, talking, saying, having, getting, looking* and *taking*, plus *how (are) you doing, where (are) you going, when (are) you coming, why (are) you looking* and *where (are) you taking*.
14 Source: YouGov https://yougov.co.uk/news/2015/08/19/sim-only-march-consumers-hold-handsets-longer (accessed 2016–2010–2031).

## References

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1–48.

Biber, D., Johansson, S., Leech, G., Conrad, S., & Finegan, E. (1999). *Longman grammar of spoken and written English*. London: Longman.

The British National Corpus, version 2 (BNC World). (2001). Distributed by Oxford University Computing Services on behalf of the BNC Consortium. Retrieved from www.natcorp.ox.ac.uk

Burnard, L. (Ed.). (2007). *The British national corpus users reference guide*. Retrieved from www.natcorp.ox.ac.uk/docs/userManual

Bybee, J. (1999). The effect of usage on degrees of constituency: The reduction of *don't* in English. *Linguistics*, *37*, 575–596.

Bybee, J. (2007). *Frequency of use and the organisation of language*. Oxford: Oxford University Press.

Bybee, J., Perkins, R. D., & Pagliuca, W. (1994). *The evolution of grammar: Tense, aspect and modality in the languages of the world*. Chicago: University of Chicago Press.

Caines, A. (2010). *You talking to me? Zero auxiliary constructions in British English*. Ph.D thesis. Cambridge: University of Cambridge.

Caines, A. (2012). You talking to me? Testing corpus data with a shadowing experiment. In S. Th. Gries & D. Divjak (Eds.), *Frequency effects in language learning and processing*. Berlin: Mouton de Gruyter.

Caines, A., & Buttery, P. (2010). You talking to me? A predictive model for zero auxiliary constructions. In *Proceedings of the 2010 Workshop on NLP and Linguistics: Finding the Common Ground, ACL 2010*. Retrieved from www.aclweb.org/anthology/W10-2107

Caines, A., & Buttery, P. (2012). Annotating progressive aspect constructions in the spoken section of the British National Corpus. In *Proceedings of the Eighth International Conference on Language Resources and Evaluation (LREC'12)*. Retrieved from www.lrec-conf.org/proceedings/lrec2012/pdf/1087_Paper.pdf

Caines, A., McCarthy, M., & O'Keeffe, A. (2016). Spoken language corpora and pedagogical applications. In F. Farr & L. Murray (Eds.), *The Routledge handbook of language learning and technology*. London: Routledge.

Campbell-Kibler, K. (2008). I'll be the judge of that: Diversity in social perceptions of (ING). *Language in Society*, *37*, 637–659.

Carter, R., & McCarthy, M. (2017). Spoken grammar: Where are we and where are we going? *Applied Linguistics*, *38*(1), 1–20.

Chomsky, N. (1965). *Aspects of the theory of syntax*. Cambridge, MA: MIT Press.

Comrie, B. (1976). *Aspect*. Cambridge: Cambridge University Press.

Crowdy, S. (1993). Spoken corpus design. *Literary and Linguistic Computing*, *8*, 259–265.

Dahl, Ö. (1985). *Tense and aspect systems*. Oxford: Wiley-Blackwell.

Denison, D. (1998). Syntax. In S. Romaine (Ed.), *The Cambridge history of the english language: Volume IV, 1776–1997*. Cambridge: Cambridge University Press.

Fillmore, C. J., Lee-Goldman, R. R., & Rhodes, R. (2012). The frameNet constructicon. In H. C. Boas & I. A. Sag (Eds.), *Sign-based construction grammar*. Stanford, CA: CSLI Publications.

Gries, S. Th. (2014). *Coll analysis 3.5: A script for R to compute perform collostructional analyses*. Retrieved from www.linguistics.ucsb.edu/faculty/stgries/teaching/groningen/index.html

Gries, S. Th. (2015). More (old and new) misunderstanding of collostructional analysis: On Schmid and Küchenhoff 2012. *Cognitive Linguistics*, *26*(3), 505–536.

Gries, S. Th., Hampe, B., & Schönefeld, D. (2005). Converging evidence: Bringing together experimental and corpus data on the association of verbs and constructions. *Cognitive Linguistics*, *16*, 635–676.

Hardie, A. (2012). CQPweb—combining power, flexibility and usability in a corpus analysis tool. *International Journal of Corpus Linguistics*, *17*(3), 380–409.

Huddleston, R., & Pullum, G. (2002). *The Cambridge grammar of the English language*. Cambridge: Cambridge University Press.

Hundt, M. (2004). Animacy, agentivity and the spread of the progressive in Modern English. *English Language and Linguistics*, *8*, 47–69.

Jespersen, O. (1933). *Essentials of English grammar*. London: George Allen and Unwin.

Jurafsky, D., Bell, A., Gregory, M., & Raymond, W. D. (2001). Probabilistic relations between words: Evidence from reduction in lexical production. In J. L. Bybee & P. Hopper (Eds.), *Frequency and the emergence of linguistic structure*. Amsterdam: John Benjamins.

Lee, S-A. (2007). Ing forms and the progressive puzzle: A construction-based approach to English progressives. *Journal of Linguistics*, *43*(1), 153–195.

Leech, G., Hundt, M., Mair, C., & Smith, N. (2009). *Change in contemporary English: A grammatical study*. Cambridge: Cambridge University Press.

Love, R., Dembry, C., Hardie, A., Brezina, V., & McEnery, T. (2017). The spoken BNC2014—designing and building a spoken corpus of everyday conversations. *International Journal of Corpus Linguistics*, *22*(3).

Mandelbrot, B. (1966). Information theory and psycholinguistics: A theory of word frequencies. In P. Lazarsfeld & N. Henry (Eds.), *Readings in mathematical social science*. Cambridge, MA: MIT Press.

McElhinny, B. S. (1993). Copula and auxiliary contraction in the speech of White Americans. *American Speech, 68*, 371–399.

Pluymaekers, M., Ernestus, M., & Baayen, R. H. (2005). Lexical frequency and acoustic reduction in spoken Dutch. *Journal of the Acoustical Society of America*, *118*, 2561–2569.

Quirk, R., Greenbaum, S., Leech, G., & Svartvik, J. (1985). *A comprehensive grammar of the English language*. London: Longman.

R. Core Team. (2016). *R: A language and environment for statistical computing*. Vienna: R Foundation for Statistical Computing. Retrieved from www.R-project.org

Ritter, A., Cherry, C., & Dolan, B. (2010). Unsupervised modeling of Twitter conversations. In R. M. Kaplan (Ed.), *Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics*. Los Angeles: Association for Computational Linguistics.

Sag, I. A. (2012). Sign-based Construction Grammar: An informal synopsis. In H. C. Boas & I. A. Sag (Eds.), *Sign-based construction grammar*. Stanford, CA: CSLI Publications.

Sanford, D. (2008). Metaphor and phonological reduction in English idiomatic expressions. *Cognitive Linguistics*, *19*, 585–603.

Savage, M. (2010). *Identities and social change in Britain since 1940*. Oxford: Oxford University Press.

Stefanowitsch, A., & Gries, S. Th. (2003). Collostructions: Investigating the interaction between words and constructions. *International Journal of Corpus Linguistics*, *8*, 209–243.

Stenström, A-B., Andersen, G., & Kristine Hasund, I. (2002). *Trends in teenage talk: Corpus compilation, analysis and findings*. Amsterdam: John Benjamins.

Trudgill, P. (1995). *Sociolinguistics: An introduction to language and society* (4th ed.). London: Penguin.

Wells, J. C. (1982). *Accents of English: Volume 1, an introduction*. Cambridge: Cambridge University Press.

Winter, B., & Wieling, M. (2016). How to analyze linguistic change using mixed models, growth curve analysis and generalized additive modeling. *Journal of Language Evolution*, *1*(1), 7–18.

Zipf, G. K. (1965 [1935]). *Psycho-biology of language: An introduction to dynamic philology*. Cambridge, MA: MIT Press.