A Study of Non-Finite Forms of Anaphoric DO in the Spoken BNC

Kimberly Oger

A study of non-finite forms of anaphoric DO in the spoken BNC

Kimberly Oger
CeLiSo EA 7332, Sorbonne University
koger@hotmail.fr

RESUME :

Il existe en anglais britannique une forme non finie de DO intransitif, utilisée en contexte anaphorique, appelé " DO britannique ". Ce DO britannique est considéré par certains comme un " proprédicat " ou " substitut ". Il est rarement utilisé dans d'autres variétés d'anglais et les locuteurs américains, notamment, le jugent agrammatical.

Le " DO britannique " a jusqu'ici fait très peu l'objet de discussion dans la littérature, surtout comparé à d'autres formes d'anaphore faisant appel à DO telles que l'ellipse, par exemple. Dans l'ensemble, les quelques lignes consacrées à son sujet se limitent à quelques traits descriptifs basés sur des exemples inventés ou glanés ça et là dans des œuvres littéraires. Il n'a jamais été entrepris d'étudier le phénomène à partir d'un corpus d'exemples authentiques représentatif de l'utilisation de la langue en situation de communication.

Cet article présente la première étude réalisée sur DO britannique à partir de données empiriques, à savoir 486 occurrences prélevées dans la composante orale du British National Corpus (BNC). Les premiers résultats de cette étude sont présentés et commentés. Ils tendent à contredire nombre d'affirmations formulées par le passé au sujet du DO britannique, et donnent lieu à de nouvelles interrogations.

Mots-clés : Do britannique - anaphore verbale - DO proprédicat - linguistique de corpus - BNC oral

ABSTRACT:

"British DO" (sometimes referred to as "propredicate" or "substitute" DO) is a non-finite form of intransitive DO, used as an anaphor in conversational British English. It is rarely used in other varieties of English, and American speakers, in particular, tend to consider its usage ungrammatical. British DO has received very little attention in the literature, particularly in comparison with other anaphoric phenomena involving DO, such as ellipsis. What has been written has been mainly descriptive in nature and based on contrived or literary examples. No in-depth analysis has been carried out on a corpus based on real data that is representative of English usage.

In this article I present the first study based on empirical data, i.e. 486 occurrences of British DO taken from the spoken BNC. I also discuss the initial results of this study, which tend to disprove much of what has heretofore been written on the subject, and raise further questions to be addressed.

Key words: British DO - verbal anaphora - propredicate DO - corpus linguistics - spoken BNC

1. Introduction: Definition and exclusions

"British DO" refers to non-finite forms of anaphoric DO, used mainly in British English (henceforth BrE). It is rarely used in other varieties of English, and American speakers, in particular, tend to consider its usage ungrammatical. According to Biber, *et al*. (1999: 430-432), British DO is more frequent in conversation than in any other register. It is relatively frequent in fiction, and rare in journalistic and academic writing.[1]

As part of their discussion on verbal anaphora, Halliday & Hasan (1976) provide the following examples to illustrate what they call "substitute"[2] DO:

(1) a. … the words did not come the same as they used to **do**.

b. 'I don't know the meaning of half those long words, and, what's more, I don't believe you **do** either!'

c. He never really succeeded in his ambitions. He might have **done**, one felt, had it not been for the restlessness of his nature.

However, if we look closely at these examples, we notice that they exemplify three different usages of DO. In (1a), DO appears in a comparative context, and more precisely in a comparative clause introduced by *as*. Although some speakers of American English (henceforth AmE) might consider the use of DO in constructions such as this infelicitous on occasion, it is undeniably productive in all varieties of English, including AmE. The following selection of examples, taken from the Corpus of Contemporary American English (Davies, 2008-), attests to this:

(2) a. Senator John McCain stays in a thin-walled Crystal City high-rise, with jets from National Airport shrieking overhead. The senior Senator from New York, Charles Schumer, bunked for years with four guys. Others sleep in their offices and shower in the gym, as House majority leader Dick Armey **used to do**.   (*Time*)

b. Rarely did a serial killer cross state lines as he **had done**, and when one did, and subsequently escalated at the rate of Julian's pace, he usually began to get sloppy.
(J. Frank, *Hunting Julian*)

c. One thing is clear to me. We are all getting too old, and we may not age as gracefully as Dad **has done**.   (W.J. Boyne, *Hypersonic Thunder: A Novel of the Jet Age*)

d. The RCMP believed Miyoshi helped the boys plan the murder of the Rafay family, and they wanted him to give a full confession on tape just like his friends **had done**.
(CBS, *48 hours*)

Although (1b) might be considered anaphoric by some (Halliday & Hasan clearly consider it to be a substitute form, as do Biber *et al*., 1999), this is a finite form of DO and should, in my opinion, be analyzed as an auxiliary used in post-auxiliary ellipsis (PAE). This can easily be

---

[1] Although I contest the data presented by Biber, *et al*. on the grounds that it concerns finite forms of DO rather than non-finite forms, a preliminary check using the BYU-BNC (Davies, 2004-) has confirmed the overall tendency described by the authors.

[2] The phenomenon under study has been referred to in the literature by different appellations, *e.g*. "substitute DO" (Halliday & Hassan, 1976), "propredicate DO" (Butters, 1983), "intransitive substitute verb DO" (Quirk, *et al*., 1985), to name a few. Indeed, the precise grammatical nature of DO (lexical verb, auxiliary or proform) in this type of usage is still under debate. However, the data presented in this article has little or no bearing on the matter. Until I dispose of sufficient relevant data to allow for a more thorough analysis of this question, I have elected to take a neutral stance, and thus refer the phenomenon simply as "British DO" throughout this article.

demonstrated by inverting the polarity of the sentence: *I believe you **don't** [know the meaning of those long words] either.*[3]

Only (1c) actually contains an occurrence of "British DO". It is the only example of the three that would be considered ungrammatical by speakers of AmE. Furthermore, it is not part of a comparative construction and cannot be considered a case of PAE (**He might have **done** succeed in his ambitions*[4]). However, whenever British DO is mentioned in the literature, it is generally considered on a par with other usages of DO in anaphoric contexts, such as (1a,b).

Based on these preliminary observations, I propose to define British DO according to the following criteria:

1. Intransitive form of DO

2. Restricted to non-finite forms of DO (*do*, *done*, *doing*)

3. Used as an anaphor

4. Essentially restricted to conversational BrE

5. Completely absent from standard AmE

6. Not the same as DO in comparative contexts

2. Motivation and methodology

British DO as such has received very little attention in the literature, and most references to this phenomenon are mainly descriptive in nature.[5] It is not uncommon to come across contradictory assertions concerning its use. This may be explained by the fact that any discussion is essentially based on contrived or literary examples, and thus the claims made are mainly based on intuition. As far as I know, there has been no in-depth study of British DO based on empirical data. The first linguist that I know of to attempt such a study was Miller (2002). However, he admits that he didn't have enough data to draw any statistically significant conclusions about its use. Sharifzadeh (2012) also made an attempt to compile a corpus in order to study British DO, but was only able to collect 39 occurrences from the British National Corpus (henceforth BNC) at that time.[6]

---

[3] In this example, the part of the sentence that can be elided is placed in square brackets.

[4] An anonymous reviewer of this article has suggested that this might be comparable to one of the characteristics of PAE, *viz.* that it is impossible to keep DO when putting back the antecedent without changing the meaning of the sentence. *He did* is not strictly equivalent to *He did succeed in his ambitions*. Keeping DO necessarily creates a polemic/emphatic effect which is not present under normal circumstances in the elliptical version.
I argue however, that **He might have done succeed in his ambitions* is infelicitous under any circumstances. Despite the numerous characteristics that British DO shares with PAE, this constitutes, in my opinion, a relevant argument against considering British DO as merely a dialectal equivalent to PAE. This discussion is to be the object of future publication.

[5] A few linguists, such as Chalcraft (2006), Haddican (2007), Aelbrecht (2010), and Baltin (2012), have proposed analyses of the nature of DO within a generative framework, but again, this is outside the scope of this paper, which deals with the linguistic and socio-linguistic conditions in which British DO is used.

[6] Sharifzadeh describes in detail her sampling method in her thesis.

My aim has been to carry out an in-depth study of this particular phenomenon, by analyzing a corpus of examples based on real data that is representative of the language. I thus chose to examine the spoken component of the BNC with the aim of identifying all the occurrences of British DO that it contains.

The spoken BNC contains over 10 million words of spoken BrE, collected in the early 1990s. It is meant to be balanced for variables such as age, gender, social class and dialect region. The data contained in the source files (*i.e.* the written transcriptions) also provide information about the situational context of each portion of recorded speech. This is all information that I wished to include in my study in order to determine socio-linguistic variables which might affect the use of British DO. Finally, many of the original recordings of the BNC have been digitized and made freely available to researchers, thus making it possible to check transcriptions and study intonation patterns.

In order to compile a corpus of occurrences, it was necessary to design corpus queries that would presumably return most, if not all, instances of British DO. Based on what has been written on the syntactical context in which British DO appears, the initial queries were designed to find instances of:

- HAVE + *done*, with the possibility of up to 3 words separating the two elements

- modal + *do*, with the possibility of up to 3 words separating the two elements

These were subsequently completed with queries intended to return marginal results (*be doing*, *to do*).

Results to queries made on the BNC were systematically sorted and analyzed. Furthermore, all seemingly relevant occurrences were listened to in the corresponding recordings of the BNC. No occurrences which could not be verified aurally were retained for the study.

The resulting corpus of examples used for this initial study comprises 443 complete occurrences, *viz.* relevant occurrences complete with corresponding antecedents, and 43 occurrences for which the antecedents have not been identified, for a total of 486 relevant occurrences of British DO. Occurrences for which the antecedents could not be identified were included in the analysis of the actual British DO constructions as well as the socio-linguistic study. However, they were of necessity excluded from the analyses of antecedents.

This paper presents the initial results of this corpus-based study of British DO, and discusses how they compare to what has been written in the literature. I will begin with a general discussion of frequencies and specificities to different situation types as well as to sociolinguistic aspects regarding the use of British DO. Following that, I will look at colligations and collocations, and will conclude with an overview of antecedent verbs.

3. Results and discussion

3.1.   Frequencies *vs*. specificities

The first aspect that I looked at was the sex of speakers who use British DO. Relative frequencies are presented in figure 1.
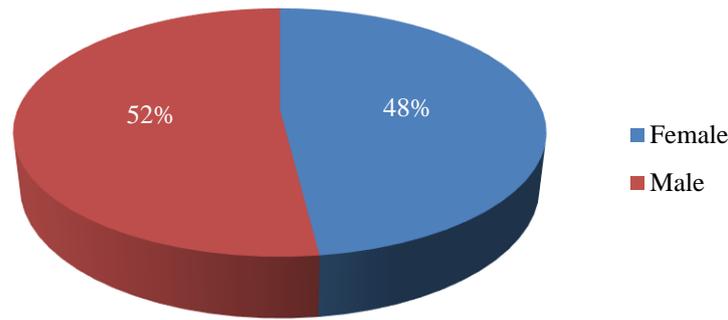
Figure 1: Relative frequencies of British DO according to sex of speakers

Figure 1 shows that occurrences of British DO are fairly evenly distributed between speakers who are identified as "female" and those who are identified as "male" in the spoken BNC. However, relative frequencies such as these do not take into account the sizes of the parts of the corpus corresponding to each attribute. In order to obtain a more accurate representation of each aspect studied, I have preferred to use the Textometry platform (TXM) to calculate specificity scores of British DO usage according to the hypergeometric model proposed by Lafon (1980). This method is used to measure frequency variations within a corpus that has been divided into parts (*e.g.* female and male speakers) or subcorpora, and to determine whether or not the frequency observed in a given part may be considered normal compared to random distribution.

Specificity scores thus are calculated based on four main variables:[7]

- the frequency (*f*) of occurrences of a phenomenon within a given part;

- the total frequency (*F*) of occurrences of the same phenomenon in the entire corpus

- the size (*t*) of the part that is under consideration (*viz.* the total number of words in that part)

- the size (*T*) of the entire corpus (total number of words)

Lafon's model is used to determine the probability that, given the total number of occurrences throughout the corpus, the phenomenon would occur at least *f* times in a particular subcorpus through random distribution. This probability is expressed in the form of a specificity score, which is calculated by the integrated R module in TXM. The higher the score, the lower the probability that the phenomenon would occur at least *f* times in that subcorpus due to random distribution. Thus, a positive score corresponds to over-representation of the phenomenon in the subcorpus, whereas a negative score indicates under-representation. Threshold scores are set at -2 and +2, and any scores that fall between them indicate that the actual distribution may be considered "normal", *i.e.* approaching random distribution.

Thus, the specificity analysis of British DO according to the sex of speakers, presented in Figure 2, shows a score of 3.5 for female speakers and a score of -3.5 for males.[8] In other words, the probability of obtaining the same number of occurrences in the "female" subcorpus due to random

---

[7] Lafon's original formula was:

log P(X=fi) = log f! + log (T-f)! + log ti! + log (T-ti)! - log T! - log fi! - log (f-fi)! - log (ti-fi)! - log (T-f-ti+fi)!

For more information, *cf*. http://www.persee.fr/web/revues/home/prescript/article/mots_0243-6450_1980_num_1_1_1008

[8] The sum of specificity scores is always 0. There being only two parts compared here (female and male), the fact that their scores are equal distances from 0 is normal.

distribution would be approximately 1‰ (0.1%), whereas the probability of finding so few occurrences in the "male" corpus would also be approximately 1‰.
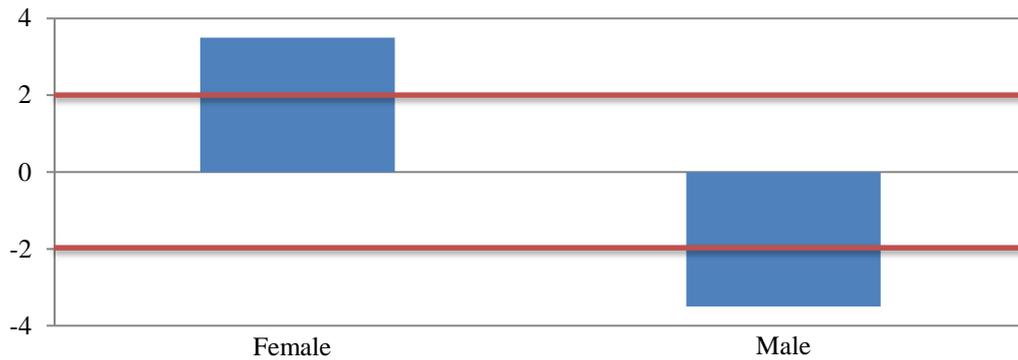
Figure 2: Specificities of British DO according to sex of speakers

These scores lead to the conclusion that, despite the fact that more occurrences of British DO are attributed to male speakers, given the total number of words attributed to females and males respectively in the spoken BNC, women would tend to be more likely to use British DO than men.

3.2. Main socio-linguistic aspects of British DO

3.2.1. Social classes

Butters (1983) has asserted that in literary works, British DO is predominantly used by "upper middle-class, often elegant characters", and implies that this is reflective of society. Furthermore, Kato & Butters (1987) consider British DO to be a characteristic of "cultivated conversational British English".

In order to verify these assertions I examined the social classes of speakers of the BNC who use British DO, wherever this attribute is indicated.[9] Figure 3 and following present not only the specificity scores for British DO according to social class, but also indicate the total number of occurrences for each class.

---

[9] Occurrences for which a given socio-linguistic attribute is "unknown" were not taken into account in the statistical analysis for that attribute, unless otherwise specified.
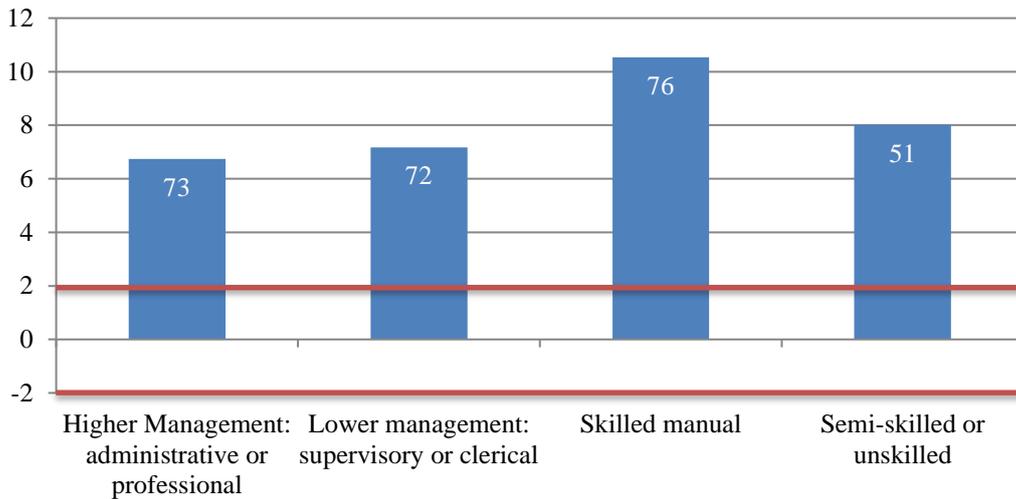
Figure 3: Specificities of British DO according to social class of speakers

If Butters is correct, one would expect to find a strong over-representation of occurrences of British DO among speakers of the "higher management" category, which would be closest to his "upper middle-class" characters. However, the specificity scores represented in Figure 3 show this category of speakers to be the least likely to use British DO. Indeed, British DO would seem to be more highly specific to skilled manual, semi-skilled and unskilled workers, which is in direct contradiction to Butters's hypotheses.

However, there are a few things which should be taken into consideration when interpreting these scores. First of all, there are a large number of speakers in the BNC for whom the value for the social class attribute is "unknown". These specificity scores were therefore not calculated from the entire corpus of examples. They only take into account the limited number of occurrences (272) for which the "social class" was indicated in the BNC.

As a result, in figure 3, British DO is presented as being over-represented in all categories. This is indeed the case relative to the rest of the corpus, for which "social class" is not indicated. However, that particular relation is irrelevant to this study. I suggest, therefore, that these specificity scores should be taken as relative, rather than absolute values.

Finally, the specificity scores indicated in figure 3 range from 6.74 to 10.54, which represents a delta of less than 4 points, which is the delta of the threshold scores. This means that the difference may be considered by some as statistically inconclusive.

These scores must therefore be taken with a great deal of precaution and any conclusions drawn from them must be verified by other statistical means.

3.2.2. Dialect regions

Some evidence has suggested that the usage of British DO might be dialectal in nature, even within BrE. Figure 4 presents the specificity analysis for British DO according to the dialect regions of BNC speakers.
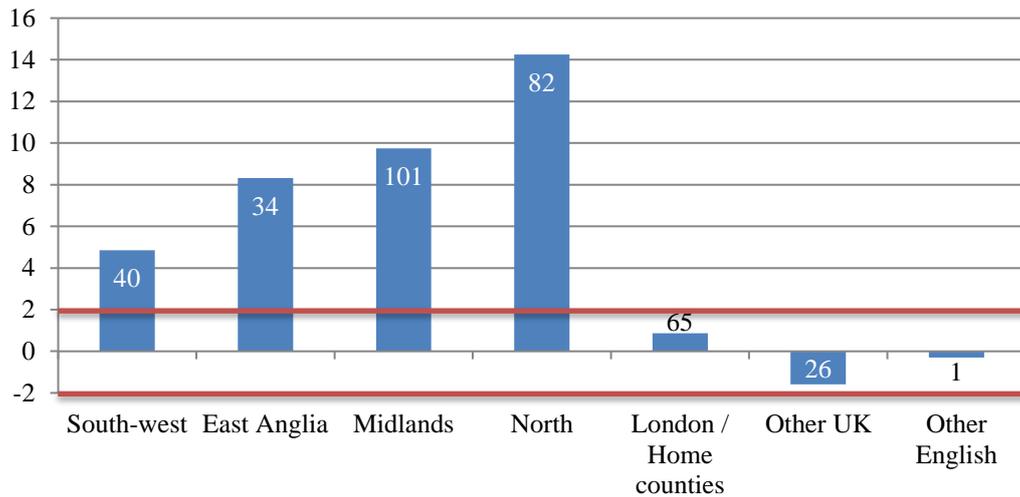
Figure 4: Specificities of British DO according to dialect regions of speakers

We can draw three very interesting conclusions from Figure 4. First of all, we can observe that British DO is mainly restricted to England. The specificity score for other UK countries (Scotland, Ireland, Wales) indicates that it is slightly underused in these countries. This has been suggested in comments to a post on the subject in Murphy's linguistic blog "Separated by a common language"[10]:

> mollymooly said...
> The Irish are with the Scots on this: "I could do" et al usually seem English English rather than British English. I think I do add a "do" on occasion, but more often I don't do :p I'm not sure if it's prosody or clarity that triggers it for me.
> 27 September, 2007 12:13

Secondly, it would seem that the farther north one lives in England, the more likely one is to use British DO. Finally, in spite of the fairly high raw frequency of occurrences involving speakers in and around London, British DO does not seem at all typical of the language of this region.

One question that has yet to be answered is whether there is a correlation between social class and dialect region among the speakers who use British DO. For instance, it would seem likely to find a higher concentration of skilled manual, semi-skilled and unskilled workers in the industrial regions of Northern England than farther south. If such a correlation were shown to exist, then the next step would be to try to determine which variable has more influence on the use of British DO. This aspect is currently undergoing further investigation.

3.2.3. Diachronic variation indicated by speaker age

According to Butters (1983) as well as Kato & Butters (1987), British DO has become more and more predominant in BrE since the end of the 18th Century (period corresponding to the

---

[10] http://separatedbyacommonlanguage.blogspot.fr/2007/09/pro-predicate-do-and-verb-phrase.html [Consulted on 16 June 2015].

American Revolution) when it was practically non-existent, to the point of "approaching the categorical" (Kato & Butters 1987)[11] in the second half of the 20th Century.

Although the BNC is a synchronic corpus, it was compiled only shortly after the aforementioned articles were written. Thus one might hope to find traces of such rapid diachronic variation in the ages of the speakers who use British DO.
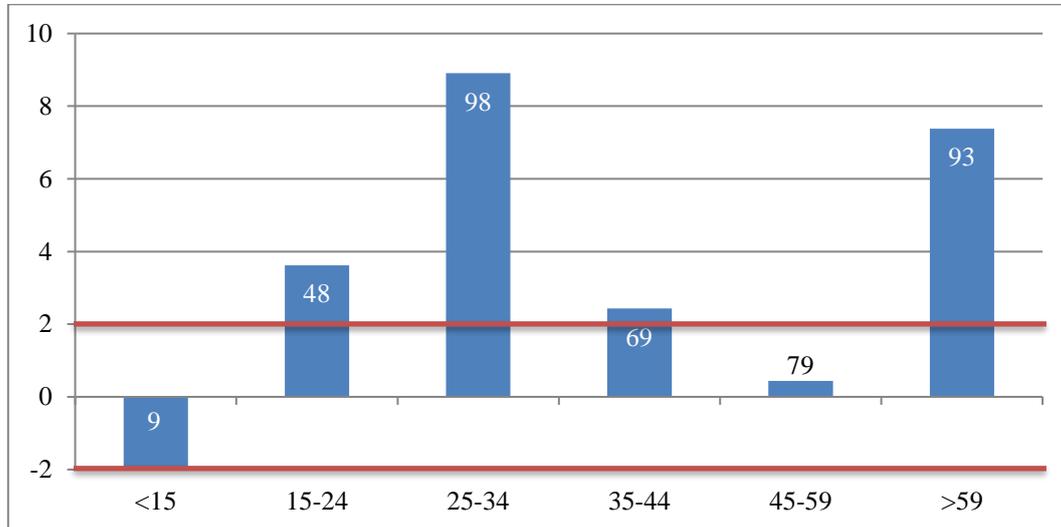

Figure 5: Specificities of British DO according to age of speakers

Unfortunately, as is shown in figure 5, there is no such indication in the spoken BNC. In the case of rapid diachronic variation, one would expect British DO to be overused by younger speakers, and underused by older ones. Discounting the "under 15" category of speakers, many of whom are very small children whose language development may not yet include British DO, the use of British DO seems to increase up to the mid-thirties, and might even be considered characteristic of the 25-34 age group if it weren't for the speakers who are over 59 years old. Indeed, British DO is highly over-represented in this group with an approximate probability of 1 in 10 million of finding at least as many occurrences among these speakers due to random distribution.

It is difficult to determine what conclusions (if any) can be drawn from the ages of speakers at this point. However, they do not appear to confirm the diachronic variation described by Butters and Kato & Butters. This will have to be checked by other means.

3.2.4. Situations

It is generally accepted that British DO is essentially limited to the spoken language. However, little attention has been given to specific registers (other than by Kato & Butters 1987, perhaps, who mention "cultivated conversational English"). Therefore, I divided the situations and

---

[11] The authors seem to base their assessment on a claim made by Butters (1983: 5): "In [Iris Murdoch's 1970 novel] A Fairly Honorable Defeat, done is retained in every possible instance." In my own examination of the same novel, I have identified 13 instances where DO is retained, and 66 instances of PAE where DO could potentially have been retained, but wasn't, (*e.g. p. 39: "Came back! But you said he'd gone away for the week-end. Is he likely to come back?" -- "Well, he might. And I wouldn't want him to find me here all naked and shivering."*). This in itself would tend to disprove Kato & Butters's assertion.

activities indicated for each recording into ten categories. The specificity scores for these categories are represented in figure 6.
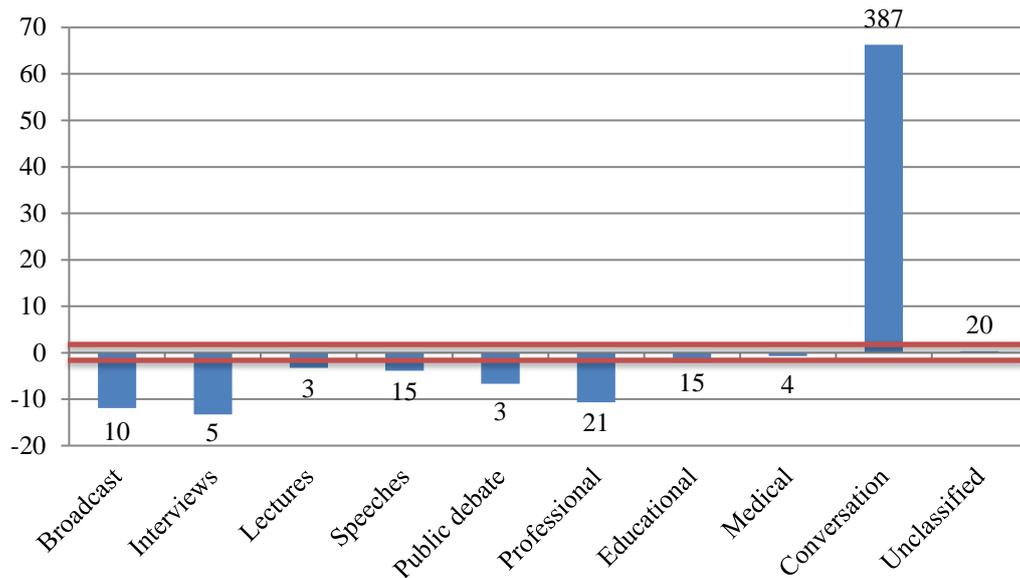
Figure 6: Specificities of British DO according to situation type

As is shown in figure 6, British DO tends to be relatively underused in all types of situations except casual conversation. The extremely high specificity score for this category is all the more exceptional in that the conversation subcorpus is by far the largest of the ten subcorpora, as it represents nearly half of the spoken BNC in number of words.

British DO appears to be underused in more formal contexts such as public debates and interviews, in television and radio broadcasts, as well as in professional contexts.

Moreover, the "speeches" subcorpus is divided into two parts: "scripted" (including sermons) and "unscripted". It is interesting to note that of the 15 occurrences of this subcorpus, none are taken from scripted speeches or sermons. All 15 are from unscripted speeches. This would tend to further indicate that when a speaker prepares in advance what he is going to say, he will be more likely to avoid using British DO.

Thus, British DO clearly appears to be characteristic of informal conversation and highly spontaneous interaction.

3.3. Linguistic aspects of British DO

3.3.1. Forms

Huddleston & Pullum (2002: 1525) consider British DO to be generally limited to perfective constructions and that the past participle *done* is the most frequently used form. However, they also state that all non-finite forms are acceptable, including the *–ing* form.

Figure 7 presents the relative frequencies of the three forms of British DO in the corpus.
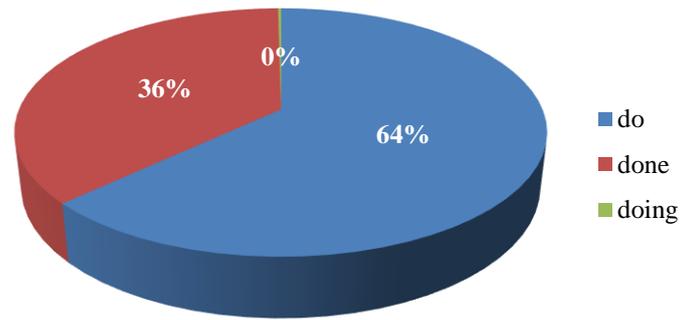
Figure 7: Relative frequencies of forms of British DO

As can be seen in figure 7, the past participle is far from being the most frequently used form of British DO in the spoken BNC. It is found in just over a third of all occurrences, whereas the base form *do* is nearly twice as frequent. From this we can conclude that British DO is by no means limited to, or even most common in, perfective constructions.

As for the *–ing* form, I have found only one relevant occurrence, in the following example.

(3) Ann (PS01U)   [3738]   Yeah!
     [3739]   They're driving round and say, oh we'll have that!
     [3740]   Straight away.
     [3741]   And what's, what happens is ... er ... if you're, if you're on scene
              they're not supposed to take it ... clamp it
Alec (PS01T)   [3742]   Right.
Ann (PS01U)   [3743]   but they **were doing** ... you know, and then you'd got to go ...
              and claim it back.   KB2

Thus, although this form is attested, it appears to be extremely rare.

3.3.2. Colligations

According to Quirk *et al*. (1985: 875), British DO can be preceded by a modal auxiliary, by HAVE in perfective constructions, or by a series of auxiliaries. Figure 8 indicates the raw frequencies of the different types of constructions found in the spoken BNC.
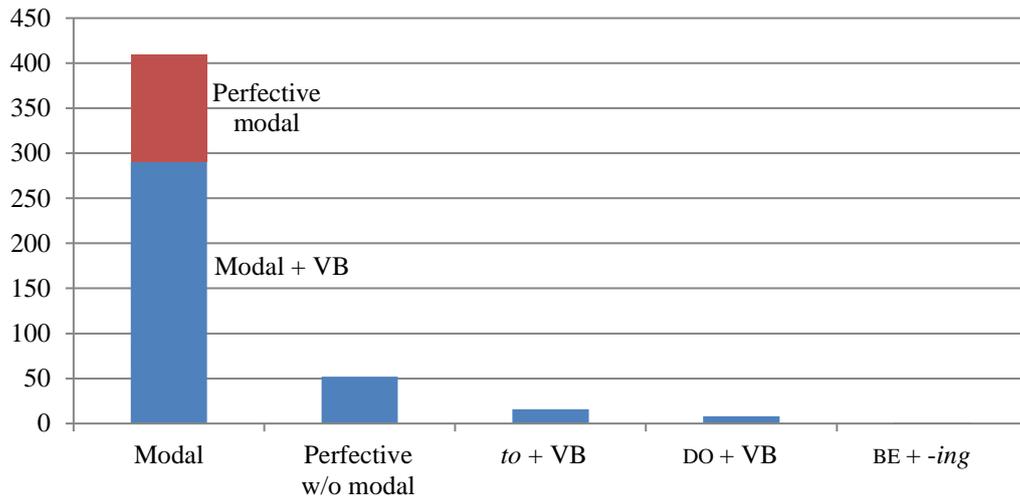
Figure 8: British DO constructions (raw frequencies)

Here we observe that in the immense majority of occurrences, British DO is preceded by a modal auxiliary, with or without HAVE. Moreover, whether or not they are preceded by a modal, perfective constructions are less frequent overall than simple modal constructions (*i.e.* without HAVE). This would explain why the past participle is less frequent than *do*.

As far as other constructions are concerned, they are by far more rare than modal and perfective constructions. They include the single aforementioned occurrence of the BE + *-ing* construction, 8 occurrences of constructions with auxiliary DO (4 "*did do*", 3 "*don't do*" and 1 "*do do*"), and 16 occurrences of "*to do*".

However, Quirk *et al*. (*ibid*) also claim that British DO does not appear in non-finite clauses, and cannot be preceded by TO. The corpus findings contradict this assertion. British DO does appear with a limited number of constructions that require the presence of *to*:

have to	5
be forced to	2
be going to	2
appear to	1
be able to	1
be bound to	1
ought to	1
try to	1
used to	1
want to	1

One may note that, through modality or aspect, nearly all of these are comparable to another, more frequent construction used with British DO (*eg. have to* and *must*). It has yet to be determined whether this lies within the realm of coincidence or is a significant linguistic factor licensing this type of British DO construction.

3.3.3. Collocations

The corpus shows that British DO occurs most frequently with modal auxiliaries, followed by perfective constructions. Figure 9 indicates in detail which words most frequently introduce British DO constructions.
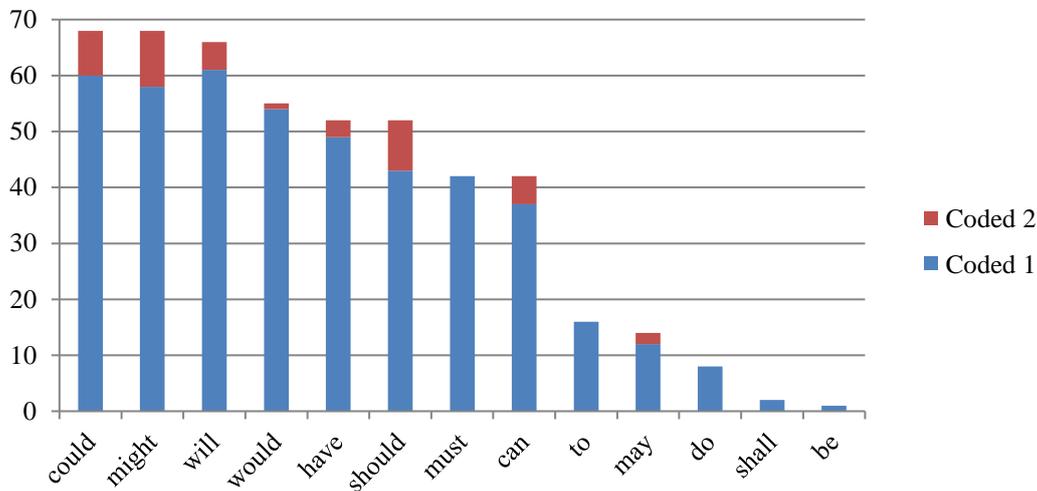
Figure 9: Heads of British DO constructions (raw frequencies)

The collocations in figure 9 can be sorted into five groups. The most frequently used words to introduce British DO constructions are the modal auxiliaries *could*, *might* and *will*, followed by a second group comprised of *would* and *should*, as well as HAVE in non-modal perfective constructions. Next we have *must* and *can*, then *may* and *to*. Finally, *shall* is found with the rare constructions involving DO and BE.

This order can be compared to that of overall occurrences of the same modals in the spoken BNC.

Table 1: Modal auxiliaries in the Spoken BNC: Overall frequencies / occurrences in British DO constructions

hw	BNC Spoken*		British DO	
	Frequency	Rank	Frequency	Rank
can	49,916	2	42	7
could	20,157	4	68	1
may	4,929	8	14	8
might	8,354	6	68	1
must	6,079	7	42	6
shall	2,889	9	2	9
should	12,087	5	52	5
will	56,878	1	66	3
would	46,246	3	55	4

The figures in table 1 underline the fact that British DO can be considered a rare phenomenon, at least in the spoken BNC. Indeed, less than 0.2% of all instances of modal auxiliaries in the corpus occur in British DO constructions. Another interesting observation to be made is that the relative frequencies of modals do not follow the same pattern when associated with

DO. The most frequent modals found with DO in the corpus, *could* and *might*, are only ranked 4th and 6th respectively in terms of overall frequencies. Conversely, the most frequent modals overall, *will* and *can*, rank 3rd and 7th respectively when they are associated with British DO. One possible explanation for this may be related to the apparent ease with which British DO is associated with epistemic modality, presented in figure 10, which contrasts with relatively infrequent epistemic uses of *will* and *can*.
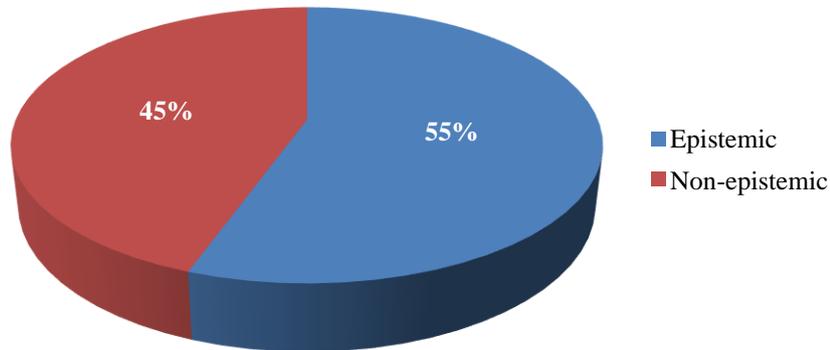
Figure 10: Values of modals associated with British DO

Overall, the corpus contains 409 occurrences of British DO in modal constructions. Figure 10 indicates the values of those modals that are thus associated with DO and highlights the fact that modals receive an epistemic interpretation slightly more often than they do a non-epistemic interpretation when associated with British DO.

These questions are currently being explored further.

3.4.   Antecedents of British DO

One of the main characteristics of anaphoric lexical DO is said to be that it can only have actions, *viz.* dynamic and agentive verbs, as its antecedents. Figure 11 presents the different types of verbs used as antecedents for British DO in the corpus.[12]
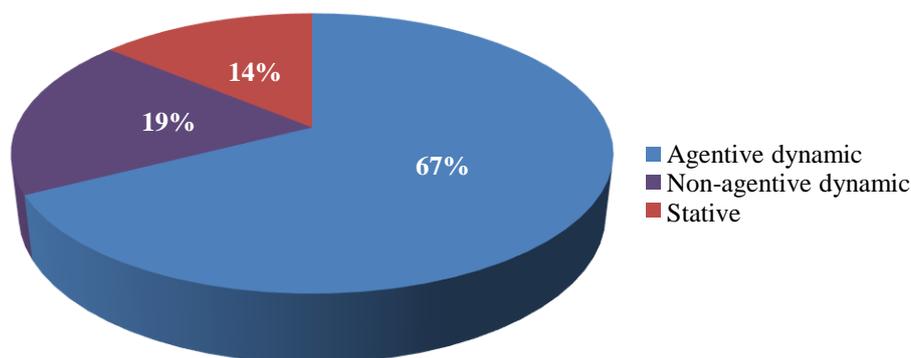
Figure 11: Relative frequencies of different types of antecedents of British DO

---

[12] As a reminder, only 443 examples were used for this part of the study.

3.4.1. Dynamic *vs*. stative antecedents

Unsurprisingly, about two-thirds of the antecedents refer to processes that are indeed dynamic and agentive. However, figure 11 also indicates that nearly 20% refer to dynamic non-agentive processes. This is rather unexpected relative to the literature, according to which this type of antecedent is more likely to be associated with *do so*. Indeed, in many of these cases, British DO could actually be replaced by *do so*. This is an aspect that will undergo further investigation in future work.

Finally, figure 11 reveals an even more unexpected characteristic of British DO, in that 14% of the antecedents are actually stative verbs. This is a non-negligible amount (61 occurrences), and cannot, in my opinion, be considered an anomaly, nor even an extremely rare phenomenon.

These are a few examples to illustrate the different types of antecedents encountered in the corpus:

(4) Antecedents referring to agentive dynamic processes
Adrian (PS0W7)   [2230] Shall I <u>tick it</u>?
Terence (PS0W2)   [2231] **Can do** if you want.   KE2

(5) Antecedents referring to non-agentive dynamic processes
Sanjay (PS0YA)   [213]   You <u>get sore bones</u> sitting down on anything.
Andrew (PS028)   [214]   Oh **can do**, yeah, yeah, yeah.   KB5

(6) Stative antecedents
Merielle (PS0LM)   [1876] They don't <u>know</u> where it's come from, they have never
                            have done and they never will.   KD8

Note that in (6), British DO is immediately followed by *will* without *do*. It would also have been possible to obtain *They never have and they never will do*, or even both auxiliaries with or without DO, without any difference in meaning. Here again is a topic for further research.

Regarding stative antecedents, Sharifzadeh (2012) notes that, when British DO is encountered outside comparative contexts, the stative verbs found in antecedents in her corpus can generally have dynamic uses in other contexts (namely whenever their subject is assigned the semantic or thematic role of agent). In my corpus, however, I have been able to identify several instances of what she calls "pure" statives, as in example (6). These I have labeled stative antecedents. All stative verbs that have dynamic interpretations in the antecedent have been categorized as dynamic antecedents.

3.4.2. Antecedent BE

This is a particularly interesting phenomenon that has been revealed through this study.

It is generally accepted that DO cannot have as its antecedent an auxiliary or lexical BE. This is true in cases of PAE as well as verbal anaphora. A few linguists have specifically mentioned this aspect:

Cotte (1997):   *Ne se substitue pas à un auxiliaire ou à BE lexical*
                 Does not substitute for an auxiliary or for lexical BE

Huddleston & Pullum (2002): The antecedent cannot have an auxiliary verb as its head

Sharifzadeh (2012): *Ne reprend ni les modaux, ni BE lexical*
                     Can have neither modals nor lexical BE as its antecedent

However, I have been able to identify some examples which are in direct contradiction with these assertions. Indeed, there are 5 occurrences in the corpus where BE is the antecedent verb of British DO. I have determined three different interpretations among these occurrences of lexical BE, and in turn assigned each to its corresponding category.

(7)   BE with an agentive dynamic interpretation
       Kathleen (PS1FC) [7810] And I bet he's <u>been down to see ducks</u>.
                          [7811] ... He **would of done**.   KCX

In (7), the antecedent can be interpreted as "*gone down to see t'ducks*"[13]. This clearly refers to a dynamic process, and the agentive role of the subject is reinforced by the adverbial "*to see t'ducks*". Thus this example has been classified with agentive dynamic antecedents.

(8)   BE with a non-agentive dynamic interpretation
       Richard (PS1K9) [1708] I mean what I'd like to do to be, to be honest, I don't know if
                               we **will do** is er <u>be at er, at er Borg by about half four, five</u>
                               <u>o'clock</u>   KDR

Here the antecedent takes the syntactic form of a pseudo-cleft construction. The relative *what I'd like to do* is identified as *be at Borg by about half four, five o'clock*, which occurs after British DO. Hence the textual antecedent of British DO is the relative clause. However, in order to interpret the referent of this, the addressee must depend on the defining element. Thus the interpretation assigned, albeit indirectly,[14] to the antecedent of British DO is "*arrive at Borg*". This is a non-agentive dynamic process, or an achievement. Indeed, the antecedent refers to an event which, although it may result from a series of actions or activities (*i.e.* agentive processes), is not in itself produced by an agent.

(9)   Stative BE
       Albert (PS01A) [3691] He probably agreed not to go into it like, you know, or ...
                               didn't want to go into it.
       Ada (PS01C)   [3692] He probably [...]
       Albert (PS01A) [3693] But he'd have been, he'd have <u>been better</u>, well ... well he
                               wouldn't have done.
       Ada (PS01C)   [3694] Yeah, the only the time he'd been paid
       Albert (PS01A) [3695] You'd have <u>been better off</u>.
       Ada (PS01C)   [3696] is when he has no money.
                       [3697] And ... he had.
                       [3698] If he said he had no money he had.   KB1

---

[13] According to my informant from Leeds, this utterance has most likely been misunderstood by the transcriber. Although I keep the original transcription in my example, I have chosen to modify it slightly in my discussion to reflect our interpretation of the speaker's words. This has no effect on my analysis of this occurrence of British DO.

[14] As British DO occurs between the relative clause and the identifying element "*be at Borg by...*", it is unclear whether the use of British DO is licensed by *do what* in the previous clause or by semantic reference provided by the following clause. One reviewer has suggested that the former is more likely, as it would be the case in the variant: "This is what I'd like to do but I don't know if we will do". This is a valid point, and it is difficult to determine whether (8') would be acceptable.

(8')   ?I'd like to be at Borg by about half four, five o'clock, but I don't know if we will do.

However, my analysis here is based on the premise that, "when we refer (and predicate) in natural language use, we are focussing upon the referents of our referring expressions, not, under normal circumstances, on their phonetic or graphical form." (Cornish, 2010).

Finally and most remarkably, the BE in the antecedent *been better off*[15] of (9) is a copular BE, a pure stative. As this is the only occurrence of British DO in the corpus with a copular BE as its antecedent, we must consider the possibility that this might in fact be a performance error.[16] However, we must also bear in mind that we are dealing with a rare phenomenon, and that even a single occurrence in a given corpus has the potential to provide insight into its usage. I have therefore decided to include this occurrence in my study and assigned it to the category of stative antecedents.

It is, of course, impossible to draw any significant conclusions based on so few examples. Nevertheless, the results of my initial study do suggest that, although this is extremely rare, contrary to what has been stated in the literature, and unlike auxiliary DO in PAE phenomena, British DO is indeed capable of having lexical BE as its antecedent.

3.4.3. Other remarkable types of antecedents

Not all antecedents can readily be labelled and sorted into categories. Among those encountered in the corpus are:

(10) Antecedents which refer to more than one process (split antecedents)
Arthur (PS50T) [425] you know from that, when the lime went
Paula (PS50V) [426] Yeah
Arthur (PS50T) [427] be sitting in the chair wide awake and suddenly cower and
  cover himself up
A. (PS50U) [428] That's suppose I think
Unknown speaker (KP1PSUNK) [...]
Arthur (PS50T) [429] he's wide awake for no reason, crying for weeks afterwards
Paula (PS50V) [430] Oh I think, well he **would do**
Arthur (PS50T) [431] Why?
A. (PS50U) [432] It's like it's shock   KP1[17]

Note that split antecedents may very well refer to different types of process, making it impossible to assign them to a single category of antecedents.

(11) Antecedents without an explicit verb
  a. Tim (PS088) [14336] If you get a match and you strike it on something else it
    would still
  Dorothy (PS087) [14337] Yes, it **can do**
  Marjorie (PS0XR) Yes, it can.   KBW[17]

  b. Evelyn (PS03T) [11283] Coffee?
  Arthur (PS03S) [11284] I **could do**.
  Evelyn (PS03T) [11285] Do you want some cake with it?
  Arthur (PS03S) [11286] No, it's ... I've had enough cake.   KBB

The antecedent in (11a) is left unfinished by the speaker, but the process that is being referred to is already part of the common ground as it has been the subject of the conversation. Thus

---

[15] The speaker interrupts himself line [3693], but completes his thought line [3695]. Thus the antecedent is identified as *been better off*.

[16] Indeed, my informant in Leeds is not convinced that the speaker does not say "he wouldn't have *been*".

[17] This is a corrected transcription of the passage, based on the audio recordings.

it can be inferred by the addressee and need not be mentioned explicitly. Similarly, in (11b) the mere mention of coffee as a question (indicated by rising intonation) in the situational context implies an offer. The question of whether the speaker had in mind "will you have", "would you like" or "do you want" is irrelevant. The process that is referred to by British DO is suggested by the situational context, and that is enough to license its use. This raises the question of whether British DO can in theory be used exophorically, and if so, why this type of usage has not been attested.

4. Provisional conclusions and future research

The initial results of this study of British DO based on empirical data have made it possible to draw several conclusions, but they also raise several questions.

As far as dialect is concerned, British DO appears to be essentially limited to BrE, and even to "English English", and seems to be progressively more and more common from the south to the north of England.

It is clearly characteristic of informal conversational English and spontaneous situations. It also seems inconsistent with what Butters calls "cultivated conversation". This we can deduce from the social classes of speakers, and particularly from the relatively high specificity of British DO to the working classes which one would not necessarily associate with the characters Butters had in mind.

These two aspects have been relatively well summed up by an informant who once said, in reference to an example containing an occurrence of British DO: "A bit informal, but we'd say that in Leeds, no problem!"

From a purely linguistic point of view, this study has revealed that British DO is most productive with modal auxiliaries, but that perfective forms are also very common. British DO constructions almost always indicate either modal or aspectual contrast from their antecedents. Finally, this study has shown that all types of antecedents are possible with British DO, including statives and lexical BE, as well as multiple processes or processes that are merely suggested.

Obviously, there is still much research and analysis to be done in order to verify and complete these conclusions. A more in-depth study is currently underway and includes several different aspects that have been mentioned here, as well as the many questions that have been raised and as of yet left unanswered.

5. Resources and References

RESOURCES:

The British National Corpus, version 3 (BNC XML Edition), 2007. Distributed by Oxford University Computing Services on behalf of the BNC Consortium. URL: http://www.natcorp.ox.ac.uk/.

Coleman, J., L. Baghai-Ravary, J. Pybus, and S. Grau (2012). *Audio BNC: the audio edition of the Spoken British National Corpus*. Phonetics Laboratory, University of Oxford. http://www.phon.ox.ac.uk/AudioBNC.

Davies, Mark. (2004-). BYU-BNC. (Based on the British National Corpus from Oxford University Press). Available online at http://www.english-corpora.org/bnc/.

Davies, Mark. (2008-) *The Corpus of Contemporary American English: 560 million words, 1990-present*. Available online at http://www.english-corpora.org/coca/.

Heiden, S. (2010). The TXM Platform : Building Open-Source Textual Analysis Software Compatible with the TEI Encoding Scheme. In K. I. Ryo Otoguro (Ed.), *24th Pacific Asia Conference on Language, Information and Computation - PACLIC24* (p. 389-398). Institute for Digital Enhancement of Cognitive Development, Waseda University, Sendai, Japan. Online.

REFERENCES:

Aelbrecht, Lobke. *The Syntactic Licensing of Ellipsis*. Vol. 149, John Benjamins Publishing Company, 2010. doi:10.1075/la.149.

Baltin, Mark. "Deletion versus Pro-Forms: An Overly Simple Dichotomy?" *Natural Language & Linguistic Theory*, vol. 30, no. 2, 2012, pp. 381–423.

Chalcraft, Faye. "Do-Doubling in Varieties of English." *Workshop on Syntactic Doubling, Amsterdam*, 2006.

Cornish, Francis. "Anaphora: Text-Based or Discourse-Dependent?: Functionalist vs. Formalist Accounts." *Functions of Language*, vol. 17, no. 2, 2010, pp. 207–41. doi:10.1075/fol.17.2.03cor.

Cotte, Pierre. *Grammaire Linguistique*. Didier érudition - CNED, 1997.

Biber, Douglas, et al. *Longman Grammar of Spoken and Written English*. Harlow: Pearson Education Limited, 1999.

Butters, Ronald R. "Syntactic Change in British English Propredicates." *Journal of English Linguistics*, vol. 16, no. 1, 1983, pp. 1–6. doi:10.1177/007542428301600101.

Haddican, Bill. "The Structural Deficiency of Verbal Pro-Forms." *Linguistic Inquiry*, vol. 38, no. 3, 2007, pp. 539–47. doi:10.1162/ling.2007.38.3.539.

Halliday, M. A. K., and Ruqaiya Hasan. *Cohesion in English*. Harlow: Pearson Education Limited, 1976.

Huddleston, Rodney D., and Geoffrey K. Pullum. *The Cambridge Grammar of the English Language*. Cambridge: Cambridge University Press, 2002.

Kato, Kazuo, and Ronald R. Butters. "American Instances of Propredicate Do." *Journal of English Linguistics*, vol. 20, no. 2, 1987, pp. 212–16. doi:10.1177/007542428702000205.

Lafon, Pierre. "Sur la variabilité de la fréquence des formes dans un corpus." *Mots*, vol. 1, no. 1, 1980, pp. 127–65. doi:10.3406/mots.1980.1008.

Miller, Philip. "Les emplois non finis de *do* auxiliaire." *Construire et reconstruire en linguistique anglaise : Syntaxe et sémantique*, edited by Claude Delmas, Saint-Etienne: Publications de l'université de Saint-Etienne, 2002, pp. 185–98.

Murphy, Lynn. "Pro-Predicate Do and Verb Phrase Ellipsis." *Separated by a Common Language* [blog], http://separatedbyacommonlanguage.blogspot.fr/2007/09/pro-predicate-do-and-verb-phrase.html. September 27, 2007. Accessed June 16, 2015.

Quirk, Randolph, et al. *A Comprehensive Grammar of the English Language*. London: Longman, 1985.

Sharifzadeh, Saghie. *Recherche Sur Do (Lexique et Grammaire).* Doctoral thesis, Université Paris-Sorbonne, 2012. http://www.theses.fr/2012PA040248.