

Sociolinguistic variation at the grammatical/discourse level

Demonstrative clefts in spoken British English

Andreea S. Calude

University of Waikato, New Zealand

This paper brings together the study of sociolinguistic variation and the area of grammatical analysis by investigating demonstrative cleft constructions in spoken British English such as *That's what I wanted to talk about* and *This is where I saw him*. Using the Spoken BNC2014S, I ask whether speaker characteristics, including gender, age, education and occupation, might be correlated with the use of demonstrative clefts and with various aspects of their structure (preference for the distal or proximal demonstrative pronoun, use of negative polarity, and use of stance adverbs). Findings suggest that in British English, demonstrative cleft use is more likely to be present in the speech of males compared to females, working adults in higher-skilled occupations compared to semi-skilled adults, and in adults of middle age compared to younger adults. This work shows that even highly abstract grammatical constructions can be sensitive to speaker preferences and linguistic communicative style.

Keywords: demonstrative clefts, sociolinguistic variation, spoken British English, anaphora

1. Introduction

This paper analyses the use of demonstrative clefts in spoken British English with the goal of documenting sociolinguistic patterns of variation. Like other cleft types (see, among others, Collins 1991 and Lambrecht 2001 for a general description, Miller & Weinert 1998 and Calude 2009 for analyses of their occurrence in spoken language, and Mair 2013, Mair & Winkle 2012, and Pattern 2012, 2013 for diachronic accounts), demonstrative clefts instantiate a grammatical means for focusing a chosen constituent, termed the 'cleft constituent', deemed to be of interest to the speaker either because it is contrastive or because it encodes newsworthy information. Demonstrative clefts are closely related to reversed wh-

clefts (or pseudo-clefts) and are often identified as forming a coherent group with these, as seen for example in the work of Collins (1991).

Demonstrative clefts are typically regarded as a sub-type of reversed wh-clefts, namely the type of reversed wh-cleft in which the cleft constituent is expressed specifically by a demonstrative pronoun (either the proximal *this* or the distal *that*¹); a property responsible for giving the construction its name (Ball 1977 terms them ‘th-clefts’, but here I follow Biber et al. 1999 and Calude 2009 in using the term ‘demonstrative cleft’). However, while reversed wh-clefts typically have an “un-reversed” counterpart, demonstrative clefts do not, as given in Examples (1) to (4) below:

- (1) A book about clefts is what I was after. [reversed wh-cleft]
- (2) What I was after is a book about clefts. [wh-cleft]
- (3) That’s what I was after. [demonstrative cleft]
- (4) * What I was after is that. [‘un-reversed’ demonstrative cleft]

The demonstrative pronoun is an anaphor to previous or upcoming discourse or, in some cases, simultaneously points to both upcoming and previous discourse, as will be discussed in Section 3. In some fewer examples, the demonstrative pronoun serves as a spatial deictic marker to physically reference entities. Thus, unlike prototypical it-clefts or other types of wh-clefts, demonstrative clefts do not encapsulate the point of focus themselves, but merely act as pointers to it.

The majority of the discussion included in this paper concerns the use of demonstrative clefts which involve the relativizer *what* (termed ‘*what*-demonstrative clefts’), as found in the sample version of the Spoken BNC2014 (Love et al. this issue); see Examples (5a) and (5b) below. In Example (5a), the cleft *that’s what I decided to tell you* contains the distal demonstrative referencing anaphorically the preceding clause *oh yeah I was craving brownies earlier* (the cleft is given in bold and the referenced material is underlined). Example (5b) exhibits a cleft constituent with the proximal demonstrative *this is what I used to do when I first moved down here*, with cataphoric reference to a portion of upcoming discourse containing multiple clauses. The clefts in Examples (5a) and (5b) are termed ‘*that*-demonstrative clefts’ and ‘*this*-demonstrative clefts’, respectively.

(5a) BNCAC002

0447: would like to ride my phoenix down to Tesco

0448: I really really wanted brownies earlier

0447: that was random

0448: well it wasn't it wasn't if you followed my brain through its thought process my thought process went oh yeah he went to Tesco to get the cheesecake and then I was like oh desserts I love desserts you know I love desserts and then I was like oh yeah I was craving brownies earlier and **that's what I decided to tell you**

0447: oh okay

(5b) BNCAP040

0439: and that's the point like it doesn't matter if I work from my house your house I can work from home and just sit with --ANON-name-m and you can go off and have a day to yourself cos you get I think you 'll need that so

0440: I've got to have it yeah

0439: so I think that works really well as well

0440: mm

0439: if I can work from home on Fridays

0438: mm

0439: if I can work from home on Fridays

0440: yeah but the the other thing as well is **this is what I used to do when I first moved down here** and I know it sounds stupid but I didn't have any jobs whatsoever and I wasn't doing part time work or anything like that is I used to go to the butcher twice a week

0439: mm

0440: and I used to make meals for two or three days but

0439: yeah

0440: not spend loads of money like buy loads of stuff for the freezer

Previous accounts of demonstrative clefts describe their syntactic structure, formulaic usage, and function in discourse, but to my knowledge, no study has previously taken on a detailed quantitative approach to probe their use from a sociolinguistic perspective. This paper aims to address this gap. The paper is organised as follows. I begin by discussing variationist accounts of grammatical structures (Section 2), and then summarize the current literature on demonstrative clefts, with particular reference to spoken language (Section 3). Next, I introduce the current study, including details of the data, coding and methods used (Section 4), followed by findings from the statistical analysis (Section 5) and a discussion of these (Section 6). The paper concludes with a short summary and outlook for future research.

2. Sociolinguistic variation within grammar

The study of syntactic variables within a variationist approach has enjoyed comparatively less attention in the literature than studies investigating phonological variation (see discussions in Andersen 2001, Barbieri 2008, Macauley 1997, Cheshire 2005, and Meyerhoff 2013). There are a number of reasons for this, examined in detail by Cheshire (2005) and Pichler (2010). A most obvious reason for the gap is the low frequency of occurrence of many syntactic variables when compared to phonological variables. Low numbers lead to a lack of confidence in the patterns observed and little scope for generalisation, which is a desired outcome of sociolinguistic research.

There is also the issue that speakers are more conscious of syntactic variables than of phonological variables, at least in part, because these tend to be more “prescribed” in nature. For example, double negatives such as *I don't know nothing* are marked, and associated with low prestige and non-standard varieties, compared to the standard negative constructions *I don't know anything*. This makes syntactic variables more “loaded” and less neutral, and hence the (potential) social stratification observed for such variables cannot be simply reduced to speakers falling into using one or other variant from a pool of possible contenders. Here the choices are more conscious with speakers aligning themselves within a certain perceived non-neutral position (standard or non-standard, marked or unmarked); see Cheshire (2005), Barbieri (2008) and Meyerhoff (2013: 33) for a discussion. Of course, while this problem is not completely absent in studies of phonological variation, syntactic variation appears to suffer from it more acutely. Moreover, even among syntactic variables, there is variation across different syntactic constructions with respect to the strength of the link between variation and cues for social perception. In other words, not all syntactic variables are homogenous with respect to the salience of their association with higher or lower social status. In experimental data from Squires (2013), grammatical information relating to the use of “singular + *don't*” sentences is more strongly marked as low-status compared to “*there's* + plural” sentences. The high frequency of the latter in spoken language may have led speakers away from associating them with low-status in this linguistic medium.

Aside from concerns surrounding prestige and social perception, there is a broader problem with finding truly synonymous variants for syntactic variables; see, for instance, Romaine's (1984) criticism of the active and passive alternation investigated by Weiner &

Labov (1983). This difficulty has led to uncertainty regarding the theoretical validity of the syntactic variable (in the Labovian sense) as a working concept: can such a variable indeed be found, and if so, what might it look like (Romaine 1984, Lavandera 1978)? Currently, it is agreed that given the “fuzziness” of syntactic variables, “a rough semantic equivalence or discourse function equivalence” is sufficient for positing a syntactic variable (Meyerhoff 2006: 402).

More recent accounts by Barbieri (2008) and Serrano and Olivia (2010) write in support of a move beyond the Labovian sociolinguistic variable, and propose instead that accounts of syntactic variation focus more on a difference of linguistic style:

In sum, formal variants are not (as long posited by the mainstream variationist view) different ways to say the same thing, which should in fact be a theoretical impossibility, but rather *different ways to say different things* (Halliday et al., 1968; Beaugrande and Dressler, 1999, p. 103). Styles are the different ways to shape, organize and communicate what is in the mind of speakers. Our concept of linguistic style is quite far from a neutral consideration of variables and variants in the traditional Labovian fashion: the alleged need to view variants as synonymous becomes senseless in light of the most recent theories on style and syntactic variation. What really needs to be undertaken from this point on is the analysis of different ways of speaking, as had already been proposed from ethnographic sociolinguistics (Hymes, 1974). (Serrano and Oliva 2011:150-151)

The implication is that what unifies the variants investigated is not a syntactic or semantic equivalence, but rather a stylistic-equivalence. While of course, equivalence is still required among variables at some level or other, the question is what “level” should this equivalence be at? One problem with pursuing equivalence between structures at the discourse level is that it can be investigated at different levels of granularity, rendering comparisons between different studies difficult to gauge.

In spite of the struggles and complexities raised by a variationist approach to syntactic phenomena, the current literature suggests an appetite for pursuing this line of inquiry. Studies such as Pichler (2010) offer a number of suggestions as a way forward. She calls for (i) investigations of larger data sets, which are likely to yield higher rates of frequency for the variables investigated (a luxury that is luckily made possible today but which was not twenty years ago, when Romaine and Lavandera were writing), (ii) the use of multivariate statistical measures, which can control for multiple factors at once within the same model rather than

testing each factor separately, (iii) the use of relative rather than absolute frequencies, for example, *X* number of occurrences per *Y* total of words in a corpus, and (iv) the integration of discourse function within the analysis.

3. Demonstrative clefts

I now turn back to demonstrative clefts with a brief review of current understanding of this construction. Perhaps owing to its rigid word order and a lack of both productive case marking and topic particles, English is rich in cleft types, as shown in Example (6) below (see Lambrecht 2001: 493-510 for a detailed overview of cleft types). There is disagreement among researchers as to where precisely the boundary between a cleft and a non-cleft might be drawn (see discussion in Collins 1991).

- | | | |
|-----|--|---------------------|
| (6) | a. It is his money that she was after. | [it-cleft] |
| | b. What I like about him most of all is his humour. | [wh-cleft] |
| | b. His humour and wit is what people like most about him. | [reversed wh-cleft] |
| | c. It is not that I hate him, I just can't stand to see him gloat. | [inferential cleft] |
| | d. All this appliance does is grunt and moan. | [all-cleft] |
| | e. It's been a month since we have seen her. | [since-cleft] |
| | f. There's the bar situation he's trying to sort out. | [there-cleft] |

The demonstrative cleft is rather uncontroversially welcomed into the cleft family (Collins 2004, Huddleston and Pullum 2002, Lambrecht 2001, Hedberg 2000, Biber et al 1999, Miller 1996, Miller and Weinert 1998, Trotta 2000 among many others). The defining characteristic of the demonstrative cleft is its referential nature, as achieved through the use of the distal or proximal demonstrative pronoun (*that* or *this*). It has been shown that the cleft constituent can have anaphoric, cataphoric, or exophoric reference, but that it can also reference both anaphoric and cataphoric material at the same time, or even have no explicit reference at all (see Calude 2008 for examples and discussion). The anaphoric use is the most common and the exophoric the least frequently encountered one (at least in the New Zealand English corpus data which is, to my knowledge, the only data from which precise frequencies of each type of demonstrative cleft were obtained).

Where the material they reference can be identified, discourse-deictic demonstrative clefts can be used to reference various types of expressions, ranging from a single phrase to a

whole clause or a longer clause complex portion. The discourse material references may be uttered by the speaker producing the cleft or by other participants present, and it could be located up to several turns before the turn in which the cleft is uttered (Calude 2008, 2009).

These findings suggest a great deal of flexibility of reference in demonstrative clefts. The exact difference in use between proximal demonstrative clefts (*this*-demonstrative clefts) and distal demonstrative clefts (*that*-demonstrative clefts) has not been conclusively identified (if indeed there is one). One suggestion is that *this*-demonstrative clefts tend to be cataphoric, while *that*-demonstrative clefts are predominantly anaphoric (Miller 1996, Miller and Weinert 1998). However, *this*-demonstrative clefts are infrequent compared to *that*-demonstrative clefts, making reliable generalisations difficult to ascertain. In agreement with examples from Diessel (1999) and Calude (2009), the Spoken BNC2014S data also shows that *this*-demonstrative clefts can have both cataphoric and anaphoric reference, as do *that*-demonstrative clefts, see examples (7)-(10) (as before, the referenced expression in each cleft is underlined and the cleft construction is given in bold):

(7) BNCAP001 – anaphoric *this*-demonstrative cleft

- 0440: yeah
0439: my hair is a bit shitty at the bottom what a nice horse speckly oh no maybe it's just dirty was it speckly or dirty? I couldn't tell
0440: don't know oh we nearly ended in the oh god two more
0439: more horses **this is what happens when you live in the countryside**
0440: oh
0439: combine harvester are they allowed to do that?
0440: they shouldn't do no
0439: doesn't look like they're taking any notice of the fact that there's a car behind you oh no now they are oh look what the the horses have like weird things round their ankles like plastic

(8) BNCBE004 – cataphoric *this*-demonstrative cleft

- 0323: and I wanted something with aloe vera in it
0320: yeah
0323: and there's some Nivea stuff which I bought
0320: mm
0323: but she said to me and th- I **this is what I like about the girls in there as well**
they're not about selling stuff that's expensive they 're about selling stuff that that's good for you and that they use
0320: yeah yeah yeah
0323: and she was like the Banana Boat aloe vera gel is awesome

(9) BNCAJN005 – anaphoric *that*-demonstrative cleft

- 0200: oh it's really hot it's burning me
0187: take it off you
0387: oh cos it takes ages I just want to get clean
0200: it is literally moving the shower head
0387: yeah but you've got to wait and you get cold
0200: or you can burn
0187: or you burn or you get frozen
0200: that **that's what I hate**
0387: I can embrace the cold
0200: I can't handle that cold I cannot handle

(10) BNCAP001 – cataphoric *that*-demonstrative cleft

- 0391: no I don't think so I 've erm I 've had hot cross buns tonight after that we had chicken
curry for for er tea
0392: so did I
0399: save a bit for tomorrow
0391: I thought we were eating out tomorrow
0399: **that's what I mean** save some room for tomorrow
0391: ah that's what I am thinking I don't really fancy
0392: ah why don't we go to the garden centre --ANON-place tomorrow because --ANON-
name-f is going to go there with --ANON-name-m's sister

Given that there does not appear to be a strict and absolute link between phoricity and choice of demonstrative pronoun (*this* or *that*), one possibility may be that this choice is socially conditioned, with certain types of speakers preferring one or the other of the demonstrative pronouns. If this is the case, then in a sufficiently large sample of cleft data, one might expect to see probabilistic trends of association between social variables (such as gender, age or social class – it is *a priori* unclear which if any of these do indeed apply here) and *this*-versus *that*-demonstrative clefts. This question is addressed in Section 5.2.

Finally, demonstrative clefts have been noted for their formulaic character (Biber et al. 1999: 961, Calude 2009, Pavesi 2016, Weinert and Miller 1996, Miller and Weinert 1998) and for their high frequency in spoken discourse, particularly spontaneous spoken conversation (Biber et al. 1999, Calude 2008, 2009, Miller and Weinert 1998). This formulaic character indicates the construction's "word-like" behaviour, that is, having a fixed structure, exhibiting a lack of pauses within the cleft construction, and being expressed as a phonologically cohesive unit. From a sample of 205 demonstrative cleft constructions, Calude (2009: 69) finds that over 90% of these followed the schema "*That's* +

what/why/where + personal pronoun/name + verb of cognition/communication/movement”. While this is the usual pattern observed, some demonstrative clefts exhibit variation in their otherwise fixed schema, and they occur with various adverbs, such as *probably*, *certainly*, *basically*, *definitely*, *clearly*, *also*, and *essentially*. These adverbs encode the speaker’s commitment or stance to what is being asserted in the cleft about the referenced material. Consider Examples (11) and (12) below.

In Example (11), the speaker is using the cleft to reference the previous speaker’s comment about being forced to drive at most forty miles per hour, and to use this as an inference for why they might be experiencing queues. Immediately after the assertion, the speaker realises that there are in fact no queues after all, but regardless, the cleft attributes what was initially thought to be a queue to the lower speed limit. The tentative nature of the link between the two is expressed by means of the adverb *probably*.

(11) BNCVNA010

0421: then at --ANON-place roundabout take the third exit onto the --ANON-place
0423: so it is yeah
0421: yeah so you're gonna branch left before you go to the roundabout
0423: that thing on my phone's not saying to go left
0421: >> no
0423: oh look at all these queues but why we have to go forty mile an hour along here? it's
silly
0421: **that's probably why there's queues** oh there's no queues
0423: >> exactly
0421: wow look at that lake
0423: I thought it was some massive thing
0423: not very big is it?
0421: quite big

In Example (12), the speaker is using the cleft as an expression of agreement with the previous speaker. Here, the scope of the material referenced by the demonstrative pronoun (*that*) is vague: it might be the final few clauses uttered (*it has that kind of heart it doesn't have a very strong central heart I think that's just my feeling*), or it could be referencing the entirety of the previous speaker’s turn. At any rate, the cleft signals approval of the points being made, and the speaker’s position is strengthened by the adverb *exactly*.

(12) BNCRW015

0261: so that's kind of like living in --ANON-place or in --ANON-place
 0262: yeah
 0261: erm so when you go to places other places in Europe I mean --ANON-place feels more like a European place because it's got some of kind of older heritage to it and stuff like that and and you feel that it's it's kind of there erm and --ANON-place is a bit kind of bits and pieces about all that yeah? not as coherent so not not feeling like that it doesn't feel like it it has that kind of heart it doesn't have a very strong central heart I think that's just my feeling
 0262: no **that's exactly what I was saying twenty minutes ago**
 0261: >> mm mm so erm if you then take the environmental side of it a place with hills a place with some kind of water it's what we talked about you know a place that isn't necessarily a big city could be a big city or maybe not but certainly it's got kind of a central heart to it
 0262: yeah where the people are friendly

A second type of variation to the schema presented above is the use of negation within the cleft's predicate, as shown in Example (13). The cleft is used in a similar vein to the construction in Example (12), namely as a marker of agreement with the previous speaker's overall message of dissatisfaction with the level of noise and disruption. But in addition, it also signals solidarity (being able to relate to the speaker's feelings and offering sympathy) and of commitment to and (indirect) involvement in the narrative (giving voice to the perceived inner thoughts that the speaker might be having relative to the situation described). The validation and echoing provided by speaker 0058 is expressed by means of a negative polarity cleft (*that's not*) whose negation marker (*not*) is perhaps symbolic of the conflict situation described by the previous speaker, but not directly expressing an actual disagreement or conflict with this speaker's contribution.

(13) BSCSG042

0058: Right
 0057: And they've taken the rooms of the people who were just here for a semester last semester
 0058: Right
 0057: And they're noisy as anything at all kinds of anti-social hours
 0058: Mm
 0057: And it's really really annoying like I've been woken up about six or seven times
 0058: Oh **that's not what you want**
 0057: By them playing music and stuff at like really inappropriate hours um
 0058: That's --UNCLEARWORD

0057: S- but like yeah we have sub-wardens there all the time and we have people there from eight till six during the day so

Demonstrative clefts are referentially low in content, which makes them a versatile and cognitively un-taxing means for regulating discourse, either by referring to the direct or indirect speech of other participants, or by functioning as carriers of explanatory or evaluative remarks (Calude 2008). It is this global discourse management role, coupled with an informal character (Biber et al. 1999) attributed to their loose and imprecise connection to the surrounding discourse (recall that at times, it is not possible to pinpoint exactly what the demonstrative pronoun references) that affords the demonstrative clefts the kind of flexibility, which seems so useful and fitting in spoken language. It is thus perhaps no surprise that this cleft type should be found so commonly in this language genre.

Its versatile function in discourse also makes the demonstrative cleft a great contender for investigating links between possible sociolinguistic variation and grammatical patterns. On the basis of 205 demonstrative cleft constructions found in New Zealand English, it was observed that women used the clefts significantly more than men, and that Māori (indigenous people of New Zealand) used them more than Pākehā (European New Zealanders). Unfortunately, each of the social factors was tested separately, so it is not possible to be sure that the results are representative. Also, any deviations from the typical demonstrative cleft schema found were infrequent in that sample. A larger sample of data might be used to test whether such deviations (e.g. the use of stance markers) could be linked to specific speaker profiles. Hence the New Zealand English study brings to attention the possibility of testing these hypotheses in larger datasets and other varieties of English, and using multivariate statistics (as called for by other researchers, including Pichler 2010), which forms the aim of the present study.

4. Data and methods

The Spoken BNC2014 Sample (Spoken BNC2014S; see this issue's Introduction) analysed here comprises approximately 5 million words of conversation from 376 identifiable speakers, uttered over the course of 567 interactions, and was made available via Lancaster University's CQPweb server (Hardie 2012). The data contains unplanned, spontaneous spoken language, typically interactions between two speakers who know each other well,

conducted in various settings (their own homes, in their cars, at cafes, and so on). The speakers vary with respect to gender, age, highest qualification, occupation, social group, (self-reported) accent, and regional dialect and this information is recorded for each participant, making it available for analysis. The corpus is searchable by regular expressions, and I searched for “(that|this) {be} * (wh*|how)” in order to extract all instances of demonstrative clefts.

After excluding 435 constructions which were not clefts, such as those in Example (14) below, the remaining 5,741 demonstrative cleft constructions were coded for analysis. The construction in Example (14a) gives a biclausal clause complex in which the copula is not followed by a typical headless relative clause, but rather a headed relative clause.² Example (14b) contains two completely separate clauses, uttered in different turns and by different speakers, which the search picked up by the mere fact that the words occur adjacent to each other (but they do not form a clause complex together). Thus the structure in these two examples does not follow the demonstrative cleft schema of “demonstrative pronoun + copula + wh-word + relative clause” (see Calude 2009: 131).

(14a) BNCAC002

0448: you've gotta be you have to be

0447: well not really

0448: yeah you actually participate in sport by choice **that is someone who gets along with it**

0447: mm true

0448: well you can't be an ex-professional footballer and say you don't like exercise oh my god all the fat people would kill you

0447: I don't really like exercise I didn't enjoy the fitness part of it

(14b) BNCAC002

0448: well I think it's only fair that you pick something I don't wanna do but actually it will benefit my life massively and make me say oh I 'm so glad you did that cos that's what's gonna happen when you watch Game of Thrones you're gonna say oh thank you for making me watch that **this is amazing**

0447: **how long** is Game of Thrones?

0448: well it's a TV show it's an hour long what do you think?

Due to the large number of examples extracted, it was not possible to manually inspect all the 5,741 constructions³ in order to ensure they were all actually clefts, but as discussed above,

certain constructions were excluded from the sample on the basis of exhibiting the structure “that’s + adjective/noun + wh-word” (as below).

In light of the discussion in Section 3 regarding certain speaker characteristics being linked to a relatively higher use of demonstrative clefts (namely, females and Māori speakers in the New Zealand English data), these variables were tested to see if similar trends can be observed in British English. The results obtained from analysing the whole sample of demonstrative clefts are discussed in Section 5.1.

A second intended focus of the study was to contrast *this*-demonstrative clefts and *that*-demonstrative clefts as well as deviations from the formulaic schema. In order to investigate this, I wanted to keep as much of the remaining structure of the cleft constructions constant (and therefore comparable) as possible. To this end, the analysis centres on demonstrative clefts whose cleft clause is introduced by the wh-word *what* (termed *what*-demonstrative clefts from hereon in) because these were the most frequent type, as illustrated in Figure 1.⁴ Hence the analysis in Section 5.2 focuses on the resulting subset of 3,102 constructions, uttered by 280 participants (and 2 by unknown ones), suggesting that 75% of the recorded speakers used a *what*-demonstrative cleft at some point during their interactions.

[Figure 1 here]

Figure 1. Plot of *this*- and *that*-demonstrative clefts occurring with various relativizing words

Each speaker identified as having uttered (at least) one cleft was coded for their gender, highest qualification, age, and occupation, in accordance with the Spoken BNC2014S metadata. The regional dialect variable was excluded from the analysis; this was coded in various ways, which either showed extensive skewedness (towards South-East British English), or the distinctions made were too detailed to provide a coherent view of the spread of data.

The gender category is relatively straightforward and the coding was kept as given in the Spoken BNC2014S, distinguishing between male and female participants. However, both the highest qualification and the age categories were adjusted for the analysis.

In the Spoken BNC2014S, the “highest qualification” speaker metadata categories were not included in CQPweb’s restricted query function (they are, however, available for the full version of the corpus – see Love et al. this issue). So that the data could be analysed according to speaker qualification, I was granted access to the original non-standardized entries for this category for each of the speakers in the Spoken BNC2014S. The non-

standardized responses were presented as follows: college, college/6th form, graduate, university, postgraduate, secondary school, university graduate, and none indicated. In order to avoid possible overlap between the various categories and also to maximally differentiate between participants' different levels of education, four major groups were posited as a means of standardization: no formal education (none indicated), school (secondary school), university entrance (college and college/6th form), and university (university, graduate, postgraduate, and university graduate).

The age categories provided by the Spoken BNC2014S are more fine-grained than the ones used in this analysis (0-10, 11-18, 19-29, 30-39, 40-49, 50-59, 60-69, 70-79, 80-89, 90-99) with some being highly skewed in number of participants. For example, there were many more 19-29 year old speakers compared to every other group and in particular to 80-89 and 90-99 year olds. This made comparisons difficult, so, the speakers were combined into fewer age groups, which correlated roughly with major life-stages, yielding the following three groups: 1-29 (young adult speakers), 30-59 (working adults), 60-99 (older speakers).

Finally, the occupation categories were coded as specified in the Spoken BNC2014S ranging from A to E: A (higher managerial, administrative and professional), B (intermediate managerial, administrative and professional), C1 (supervisory, clerical and junior managerial, administrative and professional), C2 (skilled manual labourers), D (semi-skilled and unskilled manual workers), and E (state pensioners, casual and lowest grade workers, unemployed and state beneficiaries).

5. Demonstrative clefts in spoken British English

This section reports the results from the statistical analyses of the four models built to test speaker preferences for all demonstrative clefts (Section 5.1), and for *what*-demonstrative clefts specifically (Section 5.2), comparing the use of *this*-demonstrative clefts and *that*-demonstrative clefts, the use of positive and negative polarity, and the use of stance adverbs.

Each model is a logistic regression, that is, a Generalised Linear Model with a binomial distribution modelling the chance of an event (for instance, the use of a cleft) versus a non-event (the use of a non-cleft construction). The binomial distribution is used because the data contains a set number of trials (here, utterances⁵) where the outcomes are either an event or a non-event, and any two trials with the exact same conditions, that is, the same speaker with their associated sociolinguistic characteristics, have the same probability of

producing an event. It is useful to note that these models are not meant for predictive purposes, but are built in order to test the influence of variables. In other words, I would not assume that the model could accurately predict for any given speaker, how many clefts of a certain type might be uttered, but rather the hypothesis tested here relates to whether a certain type of speaker (e.g. male/female, younger/older) might be statistically more likely to use a given cleft type. The models are discussed in turn, beginning with the use of (all) demonstrative clefts.

5.1 All demonstrative clefts

The first logistic regression tests whether the use of demonstrative clefts can be linked to specific speaker attributes. I tested the four characteristics outlined in Section 4, namely, gender, age, highest qualification, and occupation. In this model, the use of a demonstrative cleft constitutes an event and the use of a non-cleft utterance constitutes a non-event (obtained by subtracting the total number of demonstrative clefts for each speaker from their total number of utterances).

The model was initially trimmed to only include significant factors, thereby leaving three factors: occupation (Chisq=46.116, df=5, $p<0.0001$), age (Chisq=17.541, df=2, $p<0.0001$) and (borderline) education (Chisq=7.475, df=3, $p=0.058$). There was no significant difference in performance between the full model and the reduced model (Chisq=2.9994, df=1, $p=0.083$).⁶ However upon inspecting the model diagnostics, three influential observations (e.g. outliers)⁷ were discovered and removed. When the model was re-built without these observations, all four factors became significant. This time there were no significant influential observations (Cooks distance ≥ 1) and there were no obvious patterns in the residual plot.

The odds ratio was used to measure the size of the effect for the use of demonstrative clefts compared to overall number of utterances uttered for each speaker, as shown below in Table 1. The table shows that, in general, 30-59 year old speakers use more demonstrative clefts than younger speakers (but this trend is not seen for the oldest group), speakers with school education use clefts more than those without schooling (but this effect was not seen for university educated speakers), male speakers use demonstrative clefts more than female speakers, and speakers in lower occupation groups (D) tend to use more demonstrative clefts

than those in the highest occupation group (once again, the effect is not detected for the lowest occupation group, E).

Table 1. Confidence Intervals for significant predictors in the model

Predictors	2.5 %	97.5 %
(Intercept)	76.9917837	132.6969808
NewAge30_59	0.8722379	0.9995064
NewAge60_99	0.9047242	1.0847197
EducationSCHOOL	0.6040346	0.9859624
EducationSCHOOL_UNI_ENT	0.5454977	0.9126853
EducationUNI	0.6130906	1.0165025
GenderM	1.0411397	1.1800808
OccupationB	0.9381314	1.1029816
OccupationC1	0.8993610	1.1065086
OccupationC2	0.8932787	1.3378144
OccupationD	1.3814344	2.1423889
OccupationE	0.8923758	1.0961944

Figure 2 shows a graphical interpretation of the relationship between demonstrative clefts and total number of utterances, indicating that the more utterances a speaker uses, the more likely they are to employ a demonstrative cleft. This might seem intuitively obvious, but I believe it is not. Indeed, it is not a given that with an increase in number of utterances, there necessarily comes an increase in the use of any given construction (in our case, demonstrative clefts). In fact, this trend suggests: (i) that the construction is used widely and (ii) that, in spite of the speaker preferences observed, most speakers will eventually use a demonstrative cleft, given sufficient speaking time.

[Figure 2 here]

Figure 2. Use of demonstrative clefts vs. total number of utterances, for each speaker (n=376 speakers); on the left, raw numbers of clefts (per speaker), on the right logged values (per speaker)

5.2 What-demonstrative clefts

I now turn to the results obtained from the *what*-demonstrative clefts (demonstrative clefts which involve the wh-word *what*) extracted from the total sample of demonstrative clefts. This data was investigated for three different relationships as regards sociolinguistic variation: preference for *this*-demonstrative clefts or *that*-demonstrative clefts, negative versus positive polarity in clefts, and use of stance adverbs. Three models were built to test for stratification across speakers, as discussed below.

5.2.1 *This-demonstrative clefts and that-demonstrative clefts*

The first logistic regression was used to test for a correlation between speaker attributes and their use of a *this*-demonstrative cleft (*this is what I had in mind*) or *that*-demonstrative cleft (*that is what I had in mind*) type. As before, the model included age, gender, education level and occupation. It was then trimmed to only include significant factors, thereby leaving two factors: occupation (Chisq=13.516, df=5, $p=0.018998$) and age (Chisq=9.8038, df=2, $p<0.007$). There was no significant difference in performance between the full model and the reduced model (Chisq=17.126, df=10, $p=0.07162$). The model diagnostics showed no obvious patterns in the residual plot and no overly influential observations.

The odds ratio was used to measure the size of the effect for the use of *this*-demonstrative clefts compared to *that*-demonstrative clefts for each of the categories investigated, as shown below in Table 2. Speakers in occupation group B use comparatively more *this*-demonstrative clefts than those in group A, and speakers in the 30-59 and 60-89 age group use comparatively more *this*-demonstrative clefts than *that*-demonstrative clefts than those in the younger 1-29 age group.

Table 2. Confidence Intervals for significant predictors in the model

Predictors	2.5 %	97.5 %
(Intercept)	0.05037704	0.1054201
OccupationB	1.02232791	2.1084748
OccupationC1	0.88356954	2.1547245
OccupationC2	0.18964205	2.4482606
OccupationD	0.05517921	1.1733766
OccupationE	0.49982496	1.3009219
NewAge30_59	1.09527501	2.0617584
NewAge60_89	1.36873366	3.1852430

Figure 3 shows that speakers tend to use more *that*-demonstrative clefts than *this*-demonstrative clefts, and some speakers only use *that*-demonstrative clefts.

[Figure 3 here]

Figure 3. Use of *that*-demonstrative clefts compared to *this*-demonstrative clefts, for each speaker (n=280 speakers)

5.2.2 Negative and positive what-demonstrative clefts

The second logistic regression tested for effects involving positive and negative polarity, as shown in Example (15) below.

- | | | |
|------|---|-----------------------------|
| (15) | a. That's not what you want really. | [negative cleft], BNCJC019 |
| | b. That's not what the book was about. | [negative cleft], BNCRW036 |
| | c. That's not what you do is it? | [negative cleft], BNCCB005 |
| | d. Oh yeah that's what I decided to tell you. | [positive cleft], BNCAC002 |
| | e. That's like what you do. | [positive cleft], BNCAC002 |
| | f. That's what she's done. | [positive cleft], BNCAJN004 |

As before, the model originally included age, gender, education level and occupation and was then trimmed to only include the significant factor of age (Chisq=22.253, df=2, $p < 0.0001$). There was no significant difference in performance between the full model and the reduced model (Chisq=122.32, df=4, $p = 0.234$). The model diagnostics showed no obvious patterns in the residual plot and no overly influential observations.

The odds ratio was used to measure the size of the effect for polarity in *what*-demonstrative clefts, as shown in Table 3. Younger speakers (1-29) tend to use more negative *what*-demonstrative clefts than either of the two older groups.

Table 3. Confidence intervals for significant predictors in the model

Predictors	2.5 %	97.5 %
(Intercept)	21.365385	38.046487
NewAge30_59	2.023105	8.098204
NewAge60_89	1.399697	7.090836

5.2.3 Stance adverbs and what-demonstrative clefts

A small proportion of the *what*-demonstrative cleft constructions (5%) used contained various stance adverbs, such as *about*, *approximately*, *absolutely*, *generally*, *clearly*, *obviously*; see Figure 4. As discussed in Section 3, these adverbs relate the speaker's stance towards the link that is being made between the referenced material (by the demonstrative pronoun) and the proposition following the *wh*-word.

[Figure 4 here]

Figure 4. Adverbs used in *what*-demonstrative demonstrative clefts (n=164 clefts)

The final regression tested for a correlation between speaker characteristics and the use of stance adverbs. As before, the full model was trimmed to only include significant factors, thereby leaving one factor: occupation (Chisq=21.597, df=5, $p < 0.0001$). There was no significant difference in performance between the full model and the reduced model (Chisq=1.842, df=4, $p = 0.7648$). As before, the model diagnostics showed no obvious patterns in the residual plot and no overly influential observations.

The odds ratio was used to measure the size of the effect for the use of stance adverbs, as shown below in Table 4. In general, speakers from group A used more adverbs compared to speakers from the lower occupation groups (with the exception of speakers in groups C2 and E).

Table 4. Confidence intervals for significant predictors in the model

Predictors	2.5 %	97.5 %
(Intercept)	23.7298852	57.9813564
OccupationB	0.2249474	0.6205401
OccupationC1	0.2057686	0.6993255
OccupationC2	0.1320300	1.0475959
OccupationD	0.1050968	0.8464688
OccupationE	0.3525267	1.0914307

Each of the two most frequent adverbs was tested separately to see if the results were matched and, once again, occupation was the only significant predictor for both the use of *probably* and the use of *exactly*. However, surprisingly, this time the correlations went the other way, with the highest occupation group (group A) using the adverbs less than members of groups B and C; see the odds ratio in Table 5.

Table 5. Confidence Intervals for significant predictors in the model

Use of stance adverb: <i>exactly</i>			Use of stance adverb <i>probably</i>		
Predictors	2.5 %	97.5 %	Predictors	2.5 %	97.5 %
(Intercept)	0.002437084	0.01469784	(Intercept)	0.0004505965	0.008398075
OccupationB	1.532574611	10.94157473	OccupationB	1.1451843272	27.142543338
OccupationC1	1.703260728	15.08275245	OccupationC1	1.4596190291	43.639750653
OccupationC2	0.648100409	21.75146230	OccupationC2	1.3634820623	97.277357878
OccupationD	0.143194474	17.58476927	OccupationD	0.3200785470	73.719564505
OccupationE	0.331919642	3.85397249	OccupationE	0.6376942990	18.998554107

6. Discussion

The results described in Section 5 indicate that, in within a reasonably long stretch of conversational English, most speakers will use at least one demonstrative cleft. Significantly, certain types of speakers prefer demonstrative clefts, namely working adults (30-59) use them significantly more than young adults (1-29), speakers with schooling prefer them to those without, and male speakers prefer them to female speakers.

This latter finding seems to go against the observed trend in the New Zealand English corpus (although that analysis only looked at each factor at a time, and did not control for the other factors). It is also contrary to accounts of girls/female speakers paying more attention to interpersonal relationships in contrast to boys/male speakers being more content-oriented (cf. Cheshire 2005 and discussion therein). If male speakers are indeed more concerned with information transfer, then demonstrative clefts must be understood to fulfil a sufficiently vital role in discourse management such that their absence would negatively impact the successful exchange of information. Alternatively, the idea that male speakers tend to information transfer while female speakers focus more on developing interpersonal relationships might be premature. Perhaps both genders tend to both functions, but they have different preferences for how these functions are achieved. For now, it is difficult to tell which alternative is more realistic, however, the hypothesis that there is a difference in preference across genders as regards how interpersonal relationships are maintained is worthy of further consideration.

Recall that at first the analysis did not find gender to be a significant predictor in the model, but that, once several outlier speakers were removed, gender did become significant. This highlights the danger of using small data samples and of allowing outliers to obscure

general patterns. It also points to the fact that cleft use might be an idiosyncratic choice exhibiting idiolectal variation (Barlow 2013). While in general, larger data samples are preferred to smaller ones, they do come at a price: there is less room for coding a more precise, detailed view of discourse function, a point I will return to below.

It is interesting to note the preferred use of demonstrative clefts by older speakers rather than younger ones in the context of previous findings which link younger speakers to what Barbieri (2008) sums up as “trendy” features, for example slang, swear words, attitudinal and personal affect adjectives, discourse markers (cf. Barbieri 2008), and “be like” quotatives (cf. Cheshire et al. 2001). In this study, young speakers were also found to use comparatively fewer *this*-demonstrative clefts than *that*-demonstrative clefts compared to older speakers. This could be signalling a time-apparent change, but further data is required to be sure. Where younger speakers did show a significant preference over older speakers was in the use of negative demonstrative *what*-demonstrative clefts. Despite their informality (Biber et al. 1999), it seems that demonstrative clefts are generally not seen as particularly attitudinal, slang-like, or trendy. However, the function of negative demonstrative clefts in discourse requires further attention.

At first glance, it might be reasonable to assume that negative *what*-demonstrative clefts might be used to express disagreement or divergence between speakers, but as also argued earlier in Example (13), this is not always the case. A manual examination of the 68 examples of negative *what*-demonstrative clefts suggests that while indeed in some cases negative *what*-demonstrative clefts are used as a way of disagreeing with a previously made utterance, this is in fact not a common use of the construction.

More frequently, speakers use negative *what*-demonstrative clefts either to reject or counteract an assumption which arises from a narrative – an assumption which they oftentimes *themselves* introduce in the discourse. For instance, in Example (16) speaker 0303 explicitly states an assumption which others supposedly make about their state of mind, and then uses the cleft to deny the validity of this assumption.

(16) BNCRW022

- 0303: and there are people like --ANON-name-m and there are people like my dad who are on the end of the spectrum
- 0262: >> yes yeah
- 0303: and people who think well everyone's mad if I say oh my dad was mad they go oh yes that's like me I suffer from depression
- 0262: yeah

0303: no no no **that's not what I mean**
0262: >> no no you're not know what I 'm talking about
0303: >> no
0262: yeah
0303: no it's it's it's slightly different

One salient speaker characteristic seems to come into play in various ways regarding demonstrative *what*-demonstrative clefts, namely that of speaker's occupation. Speakers from intermediate, managerial, administrative and professional occupations prefer *this*-demonstrative clefts to *that*-demonstrative clefts compared to speakers from higher managerial administrative and professional occupations (the highest occupation group), and they also use comparatively more instances of the stance adverbs *probably* and *exactly* within their clefts. The use of adverbs is rather complex to interpret, with some contradicting results regarding the exact direction of preferences. Given the function of stance adverbs, as shown in Examples (11) and (12) (Section 3), the departure from the typical demonstrative cleft schema to one which includes stance adverbs reflects a higher degree of solidarity with the addressee, or strong engagement between the speaker and the addressee. Stance adverbs will often express the speaker's strong approval of the addressee's position, or their sympathetic outlook towards the addressee's perspective. It follows then that if occupation is taken to be a loose measure of social class, the finding regarding the use of *probably* and *exactly* would be an agreement with Cheshire (2005) and Macaulay (2002), reporting that "middle-class speakers have an independent, speaker-oriented speech style that contrasts with a working-class collaborative, addressee-oriented style" (Cheshire 2005: 497).

The demonstrative cleft analysis presented here brings a number of advantages over some of the previous studies concerned with variationist accounts of syntactic phenomena. First, the construction investigated does not come with any prescriptive baggage, and the variants tested are not associated with either high or low prestige, or with standard or non-standard use (unlike, for example, the zero- versus *that*-relatives in Tagliamonte et al. 2005, the verb agreement and various negation constructions in Cheshire 1999, the *were/weren't* forms investigated by Cheshire & Fox 2009, or the inflected future versus periphrastic future analysed in Sankoff & Wagner 2006). Given the demonstrative cleft's abstract schema compared to individual lexical items (such as negative particles, or slang terms), it seems unlikely that demonstrative clefts would be explicitly conscious to speakers, and there is a good chance that the construction can avoid the potential problem of being treated as a single lexical item rather than as a syntactic construction (a solution which has been posited for

variables such as the negative concord; see Meyerhoff 2013: 32). Third, given the size of the Spoken BNC2014S and the frequent occurrence of the demonstrative cleft in conversation, the analysis has sufficient tokens from which statistical analyses can infer patterns of use.

Despite the wealth of data analysed and the minimization (if not altogether elimination) of concerns regarding prestige, speaker awareness, and social perception, this study is not without limitations. In taking a variationist approach to demonstrative clefts, I am faced again with questions regarding the nature of the “syntactic variable”. As discussed in Section 2 and also by Geeraerts and Kristiansen (2014), the methodological problem raised by Lavandera (1978) regarding semantic equivalence still stands: can the variants compared here be seen as semantically equivalent forms which different speakers choose to use, according to the social identity within which they align themselves? Even if one replaces the (overly strong) semantic equivalence with equivalence at a higher linguistic level, namely equivalence in discourse function, the question still stands, albeit in a different form.

As a discourse tool, the demonstrative cleft exhibits low referential content, bearing an over-arching role of discourse-management and of orienting the speaker and their addressee(s) in the discourse. While in general terms this is its broad discourse-function, several sub-functions, more specific and more fine-grained can, of course, be identified (see Calude 2008, 2009). But how much discourse functional equivalence is sufficient in order to reliably study their variation across speakers (or genres, for instance)? One might argue that no two constructions will have perfect discourse functional equivalence, in the same way that no two lexical items exhibit true synonymy. As regards *what*-demonstrative clefts, the waters become even murkier perhaps when looking at positive or negative *what*-demonstrative clefts due to (subtle) differences in their function. Serrano and Olivia’s (2010) notion of ‘communicative style’, which assumes that one cannot capture grammar without considering discourse in a more holistic manner, and how a particular construction fits into it (as opposed to blindly looking to “equate” formula constructs to each other), is a useful way of capturing the “broad-brush approach” (Cheshire 2005: 480) adopted in this analysis of demonstrative clefts.

Demonstrative clefts are just one of the constructions available to English speakers for managing and organising discourse, within a co-operative and interactive environment (cf. Calude 2009). They allow participants to reference previously uttered content and ideas, to take the floor in order to move the conversation forward, and to introduce their own stance and evaluation of what is being discussed. Because these functions can be fulfilled by other means as well, or depending on speaker preference, simply not necessarily explicitly stated at

all, the use of demonstrative clefts can and should be viewed as a stylistic choice which speakers employ or not, depending on their preference. What this analysis shows is that indeed, there is a group preference for certain types of demonstrative clefts.

7. Conclusion

I propose in this paper that demonstrative clefts exhibit speaker variation, in regard to both their use and their internal structure (polarity, stance adverbs, choice of demonstrative pronoun). The size of the statistical effect is likely to be small so that it is not possible to predict *a priori* who will use a demonstrative cleft and in which situation. However, the variation patterns observed suggest that linguistic repertoires of certain social groups can be detected not just for linguistic variants which clearly align within distinct social perceptions of standard/non-standard usage or prestige, but that such patterns can also be detected at the level of more abstract linguistic structures which speakers are most probably unaware of. Alongside other studies which are increasingly being undertaken in this area, I hope to have shown that the grammar and discourse level offer fertile testing ground for variationist accounts of language phenomena. The study of spoken language is currently experiencing an unprecedented and much overdue expansion, afforded by larger corpora, such as the Spoken BNC2014, which provide a rich empirical resource that future research can draw on.

Acknowledgments

I would like to thank Robbie Love, Vaclav Brezina and Karin Aijmer for their support with this project; Robbie Love, Tony McEnery and Andrew Hardie for access to the Spoken BNC2014S corpus and their ongoing technical support; Steven Miller for his help with statistics; Jim Miller, Gerald Delahunty, Tana Wu and Daryl MacDonald for reading drafts of the paper; and the two anonymous referees for their suggestions and improvements. Any remaining errors are, of course, my own.

Notes

1. While demonstrative clefts can in theory be introduced by plural demonstrative pronouns *these/those*, this is not instantiated in real data.
2. Although Collins (1991: 29) does include certain “*that’s + noun + wh-word*” structures in his cleft definition, namely where the noun is a pro-noun (e.g. *way, thing, someone, something*), I follow the majority of the literature on reversed wh-clefts and do not include these here.
3. However, using a random number generator, I manually checked 100 random examples from the 5,741 constructions identified and only one of these was not a true cleft (namely the following construction: *almost all the same but there are a couple that are what do they call it ? F- ten to fifteen*, in file BSCCB003). Assuming that the 100 random examples are representative of the data, only 1% (or 57 constructions) might indeed not be true clefts.
4. All statistical analyses are done in *R* (R Development Team 2010) and certain plots (in Figures 1 and 4) use the *ggplot* package (Wickham 2009).
5. Even though a cleft construction need not necessarily equate to an entire utterance, keeping track of the use of utterances helps to control for total opportunities that a given speaker might have had to use a cleft construction.
6. If the education is taken out, the reduced model does perform worse than the full model, suggesting that despite its low p-value, the education factor does explain some of the variation observed (whereas, if education is left in the model, the reduced model does not perform significantly worse than the full model). It could be that there is some overlap in what education and occupation are actually measuring since the two might be correlated to some degree. However, this potential collinearity is only a problem if both factors were taken out at the same time (say on the grounds that they are not significant). But, in this case, the results point to the opposite effect, that in fact they are both needed. Hence regardless of the potential overlap between them, they still measure sufficiently different “aspects” of the variation that they are both required.
7. The three outliers were: a male speaker, 0-29, who produced only 1 cleft (and 186 non-cleft utterances), a female speaker, 30-59, who produced 72 clefts (and 12,125 non-cleft

utterances), and a female speaker, 60-69, who produced 89 clefts (and 14,696 non-cleft utterances).

References

- Andersen, G. (2001). *Pragmatic Markers and Sociolinguistic Variation: A Relevance-theoretic Approach to the Language of Adolescents*. Amsterdam: John Benjamins.
- Ball, C. (1977). Th-Clefts. *Pennsylvania Review of Linguistics*, 2, 57-69.
- Barbieri, F. (2008). Patterns of Age-Based Linguistic Variation in American English. *Journal of sociolinguistics*, 12(1), 58-88.
- Barlow, M. (2013). Individual differences and usage-based grammar. *International Journal of Corpus Linguistics*, 18(4), 443-478.
- De Beaugrande, R., & Dressler, W.U. (1999). *Introducción a la lingüística del texto*. Barcelona: Ariel.
- Biber, D., Stig, J., Geoffrey, L., Susan, C., Edward, F., & Quirk, R. (1999). *Longman Grammar of Spoken and Written English*. Vol. 2, Boston: MIT Press.
- Calude, A. (2008). Demonstrative Clefts and Double Cleft Constructions in Spontaneous Spoken English. *Studia Linguistica*, 62(1), 78-118.
- Calude, A. (2009a). Formulaic Tendencies of Demonstrative Clefts in Spoken English. In R. Corrigan, E.A. Moravcsik, H. Quali & K.M. Wheatley (Eds.), *Formulaic Language* (pp. 55-76). Philadelphia: John Benjamins.
- Calude, A. (2009b). *Cleft Constructions in Spoken English*. Berlin: VDM Verlag.
- Cheshire, J. (1999). Taming the Vernacular: Some Repercussions for the Study of Syntactic Variation and Spoken Grammar. *Cuadernos de Filología Inglesa*, 8 (Retrieved from <http://revistas.um.es/cfi/issue/view/5351>, last accessed 10 July 2017).
- Cheshire, J. (2005). Syntactic Variation and Beyond: Gender and Social Class Variation in the Use of Discourse-New Markers. *Journal of Sociolinguistics*, 9(4), 479-508.
- Cheshire, J. & Fox, S. (2009). Was/Were Variation: A Perspective from London. *Language Variation and Change*, 21(1), 1-38.
- Cheshire, J., Paul K., Fox, S., & Torgersen, E. (2011). Contact, the Feature Pool and the Speech Community: The Emergence of Multicultural London English. *Journal of Sociolinguistics*, 15(2), 151-96.
- Collins, P. (2015). *Cleft and Pseudo-Cleft Constructions in English* (2nd ed.). Vol. 6, London: Routledge.
- Collins, P. (2004). Reversed What-Clefts in English: Information Structure and Discourse Function. *Australian Review of Applied Linguistics*, 27(2), 63-74.

- Diessel, H. (1999). *Demonstratives: Form, Function and Grammaticalization*. Vol. 42, Amsterdam: John Benjamins.
- Geeraerts, D., & Kristiansen, G. (2014). Cognitive Linguistics and Language Variation. In J. Littlemore & J. Taylor (Eds.), *The Bloomsbury Companion to Cognitive Linguistics* (pp. 202-217). London: Bloomsbury.
- Halliday, M.A.K., McIntosh, A., & Strevens, P. (1968). The users and uses of language. In J. Fishman (Ed.), *Readings in the sociology of language* (pp. 139-169). The Hague: Mouton.
- Hardie, A. (2012). CQPweb – combining power, flexibility and usability in a corpus analysis tool. *International Journal of Corpus Linguistics*, 17(3), 380-409.
- Hedberg, N. (2000). The Referential Status of Clefts. *Language*, 76(4), 891-920.
- Huddleston, R., & Pullum, G. K. (2002). *The Cambridge Grammar of English Language*. Cambridge: Cambridge University Press.
- Hymes, D. (2003). *Foundations in sociolinguistics: An ethnographic approach*. London: Psychology Press.
- Lambrecht, K. (2001). A Framework for the Analysis of Cleft Constructions. *Linguistics*, 39(3), 463-516.
- Lavandera, B. R. (1978). Where Does the Sociolinguistic Variable Stop? *Language in Society*, 7(2), 171-82.
- Love, R., Dembry, C., Hardie, A., Brezina, V. and McEnery, T. “The Spoken BNC2014 - designing and building a spoken corpus of everyday conversations.” *International Journal of Corpus Linguistics* 22, no. 3 (2017).
- Macaulay, R. (1997). *Standards and Variation in Urban Speech: Examples from Lowland Scots*. Vol. 20, Amsterdam: John Benjamins.
- Macaulay, R. (2002). Extremely Interesting, Very Interesting, or Only Quite Interesting? Adverbs and Social Class. *Journal of Sociolinguistics*, 6(3), 398-417.
- Mair, C. (2013). Writing the Corpus-Based History of Spoken English: The Elusive Past of a Cleft Construction. *Language and Computers*, 77, 11-29.
- Mair, C., & Winkle, C. (2012). Change from to-Infinitive to Bare Infinitive in Specificational Cleft Sentences. In M. Hundt & U. Gut (Eds.), *Mapping Unity and Diversity World-wide: Corpus-based Studies of New Englishes* Vol 43 (pp. 243-262). Amsterdam: John Benjamins.
- Meyerhoff, M. (2006). Syntactic Variation. In K. Brown (Ed.), *Encyclopedia of Language Linguistics* (pp. 402-404). Amsterdam: Elsevier.
- Meyerhoff, M. (2013). Syntactic Variation and Change: The Variationist Framework and Language Contact. In I. Léglise & C. Chamoreau (Eds.), *The Interplay of Variation and Change in Contact Settings* (pp. 23-51). Amsterdam: John Benjamins.
- Miller, J. (1996). Clefts, Particles and Word Order in Languages of Europe. *Language Sciences*, 18(1), 111-125.

- Miller, J. & Weinert, R. (1998). *Spontaneous Spoken Language: Syntax and Discourse*. Oxford: Oxford University Press.
- Patten, A. (2013). *The English It-Cleft: A Constructional Account and a Diachronic Investigation*. Berlin: Mouton de Gruyter, 2013.
- Patten, A. (2012). The Historical Development of the It-Cleft: A Comparison of Two Different Approaches. *Studies in Language*, 36(3), 548-575.
- Pavesi, M. (2016). Formulaicity in and across Film Dialogue: Clefts as Translational Routines. *Across Languages and Cultures*, 17(1), 99-121.
- R Development Core Team. (201). *R: A language and environment for statistical computing*. Vienna: R Foundation for Statistical Computing. ISBN 3-900051-07-0.
- Romaine, S. (1984). On the Problem of Syntactic Variation and Pragmatic Meaning in Sociolinguistic Theory. *Folia Linguistica*, 18(3-4), 409-438.
- Sankoff, G. & Wagner, S.E. (2006). Age-Grading in Retrograde Movement: The Inflected Future in Montréal French. *University of Pennsylvania Working Papers in Linguistics*, 12(2), 16 (Retrieved from <http://repository.upenn.edu/pwpl/vol12/iss2/>, last accessed 10 July 2017).
- Serrano, M.J., & Oliva, M.A.A. (2011). Syntactic Variation and Communicative Style. *Language Sciences*, 33(1), 138-153.
- Squires, Lauren. "It Don't Go Both Ways: Limited Bidirectionality in Sociolinguistic Perception." *Journal of Sociolinguistics* 17, no. 2 (2013): 200-37.
- Tagliamonte, S., Smith, J., & Lawrence, H. (2005). No Taming the Vernacular! Insights from the Relatives in Northern Britain. *Language Variation and Change*, 17(1), 75-112.
- Trotta, J. (2000). *Wh-Clauses in English: Aspects of Theory and Description*. Amsterdam: Rodopi.
- Weiner, E. J., & Labov, W. (1983). Constraints on the Agentless Passive. *Journal of linguistics*, 19(1), 29-58.
- Weinert, R., & Miller, J. (1996). Cleft Constructions in Spoken Language. *Journal of Pragmatics*, 25(2), 173-206.
- Wickham, H. (2009). *ggplot2: Elegant Graphics for Data Analysis*. New York: Springer Science.

Author's address

Name Andreea S. Calude

Department or school School of General and Applied Linguistics

Institution University of Waikato

Address Knighton Road

City and postcode Hamilton, Private Bag 3105

Country New Zealand

Email address andreea.calude@waikato.ac.nz