Please cite the Published Version

Recent change in modality in informal spoken British English: 1990s – 2010s[1]

ROBBIE LOVE

Aston University

NIALL CURRY

Coventry University

Abstract

Studies in modality comprise a complex canon of functional, formal, sociological, and diachronic analyses of language. The current understanding of how English language speakers use modality is unclear; while some research argues that core modal auxiliaries are in decline, they are reported as increasing elsewhere. A lack of contemporary and representative spoken language data has rendered it difficult to reconcile such differing perspectives. To address this issue, this article presents a diachronic study of modality using the Spoken BNC2014 and the spoken component of the BNC1994. We investigate the frequency of core modal auxiliaries, semi-modals, and lexical modality-indicating devices (MIDs), as well as the modal functions of the core modal auxiliaries, in informal spoken British English, between the 1990s and 2010s. The results of the analysis are manifold. We find that core modal auxiliaries appear to be in decline, while semi-modals and lexical MIDs appear relatively stable. However, on a form-by-form basis, there is significant evidence of both increases and decreases in the use of individual expressions within each modal set. As a result, this study problematises form-based studies of change, and illustrates the value and coherence that functional analyses of modality can afford future work.

Keywords: Modality; Spoken British English; Language Change; Spoken Grammar; Spoken Corpora.

1 INTRODUCTION

Modality in English language studies is by now a well-established field of linguistic enquiry. A canon of research exists that documents functional categorisations, formal properties, and ongoing change in the expression of modality. Such research centres on the semantic functions of modality, which some researchers (e.g. Krug 2000; Palmer 2001, 2003; Facchinetti & Palmer 2003; Leech 2003) categorise as epistemic, deontic, and dynamic. Categorically, modality in English is usually conceptualised as pertaining to the core modal auxiliaries (e.g. *can*), semi-modals (e.g. *ought to*), and lexical modals (e.g. *certainly*); although, of these sets of modal expressions, only the core modal auxiliaries are consistently categorised as such in the literature. These sets of modal expressions have been at the centre of studies of language and language change, with seminal research debating whether the core modal auxiliaries are in decline (e.g. Leech 2003; 2011; 2013; Millar 2009). Yet, as a linguistic phenomenon, modality remains relevant and pertinent, with many avenues for further research, not least because of the ongoing technical and theoretical developments in the field of corpus linguistics.

Typically, research on modality has centred on written language, owing to the limited availability of representative spoken language data. However, with the advent of new spoken corpora, such as the Spoken British National Corpus 2014 (Spoken BNC2014, Love et al. 2017; McEnery et al. 2017) and the London-Lund Corpus 2 (Põldvere et al. forthcoming), there is further scope to better understand and document spoken modality in British English as well as diachronic changes when compared to the Spoken BNC1994 (BNC Consortium 2007), for example. Such a diachronic perspective will offer an important contribution to research on modality, as there is an evident dearth in knowledge surrounding contemporary spoken use of modality in British English, as well as an understanding of how this use may have changed overtime. Moreover, while it has long been recognised that modality can be classified according to semantic function (von Wright 1951), research in the field tends to formally group and classify modality into sets on lexical

syntactic grounds (core modal auxiliary, semi-modal, lexical modal; Leech 2003; 2013). Consequently, such studies of modality typically make claims surrounding modal behaviour and change with reference to these sets, which may limit our understanding of general functional behaviour and change in English modality. Advances in corpus pragmatics encourage that both form–function and function–form relationships (Aijmer & Rühlemann 2014; O'Keeffe 2018; Curry forthcoming) be considered in order to gain a more nuanced perspective of the language being studied. To-date, a comprehensive perspective on both formal and functional change in modality in contemporary British English is absent from the literature.

This study seeks to contribute such a perspective through a diachronic investigation of modality using a relatively new corpus of contemporary spoken British English, the Spoken BNC2014. The Spoken BNC2014 comprises 11.5 million tokens[2] of transcribed informal spoken British English, as recorded by hundreds of participating members of the public in the UK (mainly in England) (Love 2020). By comparing the Spoken BNC2014 to its predecessor, the Spoken BNC1994 (BNC Consortium 2007), it is possible to explore recent change in modality in informal British English between the 1990s and 2010s.

To do this, this article presents a literature review underpinning our theoretical perspective on modality, research on documented change in written and spoken modality, and the current state of the art of modality research in spoken British English (Section 2). Subsequently, Section 3 presents the data and methodology, outlining the corpus data, the modality-indicating devices studied, and the methodology for analysing the modal expressions in terms of their frequency and modal function (epistemic, deontic, dynamic). Section 4 presents our findings regarding changes in the use of the modal expressions, which suggest that changes in modality are not consistent within formal sets, and that by considering modals on a case-by-case basis, it is possible to see

---

[2] A simple definition of a token is "any single, particular instance of an individual word in a text or corpus" (McEnery & Hardie 2012: 252); in many corpora, there are also non-word tokens, such as punctuation, but we accept that *token* can (in many instances) be used as a near-synonym for *word*.

more coherent trends in modal function use. These findings are discussed in Section 5, which is followed by a brief conclusion outlining the empirical and methodological contributions of this paper and highlighting future directions for the study of modality.

2 LITERATURE REVIEW

In this section, we discuss a range of definitions for modality and its classifications, leading us to make the case that conceptual inconsistency in the literature creates a challenge for diachronic studies of modality such as the present study. We then survey corpus research into modality in written and spoken English, before presenting our Research Questions.

2.1 *Modality in English: forms and functions*

Modality is the linguistic means of indicating a speaker's attitude or point of view on a state of the world (Carter & McCarthy 2006: 638). It is widely considered to have two sub-types: grammatical mood (verb inflections) and a lexical 'modal system' (Palmer 2003: 2-3); Present-Day English (PDE) is considered not to have a grammatical mood, but rather a modal system (Palmer 2003: 3). While several members of the English modal system – modal auxiliaries and clitics – 'enjoy a rather advanced grammaticalized status' (Krug 2000: 40), and have received a lot of attention in research, there are also the less grammaticalized (i.e. lexical) 'modal constructions' (Krug 2000: 40) (e.g. *I think*) as well as the so-called 'semi-modals' (Leech 2003: 229) (e.g. *used to*). In this paper, we refer to any item that functions to express modality, regardless of form, as a *modality-indicating device* (MID), a term we adopt from Mubarak (2015).

Modality is typically classified according to function; this convention can be attributed to von Wright (1951), who proposed four categories of 'modal logic' (p. 1): *alethic, epistemic, deontic,* and *existential*. Alethic modality is concerned with 'the modes in which a proposition is (or is not) true'; epistemic modality concerns 'the modes of knowing'; deontic modality the 'modes

of obligation', and existential modality 'the modes of existence' (von Wright 1951: 1-2). In subsequent research, the 'major distinction' came to be between epistemic and deontic modality (Krug 2000: 41), with the other two categories receiving less attention, possibly because epistemic modality could be said to have subsumed alethic modality, and because existential modal expressions are less common. This approach can also be seen in work by Coates (1983), Perkins (1983), Nordlinger & Traugott (1997), Palmer (2001) and Fairclough (2003).

According to Palmer (2003: 7), epistemic modality 'is concerned solely with the speaker's attitude to [the] status of the proposition' (e.g. *They **may** be in the office*), while deontic (root) modality is concerned with the subject's ability to do something as permitted by an external source (e.g. *They **can** come in now*). Palmer (2001) classifies epistemic modality as a type of 'propositional modality' (p. 24); this is contrasted to 'event modality' (p. 70), which comprises deontic modality and another type: *dynamic* modality, which is concerned with 'the subject's own (internal) ability' (e.g. *They **can** run very fast*) (Palmer 2003: 7).

An alternative framework for modality is described by Bybee et al. (1991), who distinguish three types of modality: *agent-oriented*, *epistemic* and *speaker-oriented* (p. 23), but exclude deontic modality, as it 'cuts across the modality domain in a way that is not cross-linguistically valid' (p. 23). Despite this framework appearing to represent a different view of modality, Krug (2000: 42) suggests that Bybee et al.'s framework does, nonetheless, adhere to the traditional view of the modal system, since 'the concept of agent-oriented modality overlaps to a great extent with the concept of deontic modality'.

How modality research has addressed and labelled such functions has been inconsistent, with a range of terminology with overlapping senses used to categorise modal function. On the other hand, generally speaking, it appears that epistemic, deontic and dynamic modalities are accounted for in some way across most studies. Furthermore, there is similar inconsistency when it comes to the categorisation of MIDs across the three formal categories of core modal auxiliary,

semi-modal and lexical modal expression. For example, those who follow Quirk et al. (1985) might draw distinctions between semi-modals and lexical modals by arguing that the likes of marginal modals, semi-auxiliaries, and modal idioms are not lexical items and therefore would not be considered lexical MIDs. A similar view is shared by Bolinger (1980: 297), who argues that verbs that perform modality (e.g. *need*) and have some form of infinitive complement can be seen as modal auxiliaries or semi-modals. However, Carter & McCarthy (2006) do not share this view, and instead draw a distinction between semi-modals and lexical modal verbs by identifying shared syntactic features between core modal verbs and semi-modals (*e.g.* the lack of auxiliaries in forming negatives or questions). Moreover, Leech (2013) identifies categories such as 'emergent' modals; includes *need* and *ought to* among core modals; and includes a range of lexical modal verbs among his set of lexical modal expressions. Making sense of these varied categorisations has proven difficult, and the incoherence between the ways in which MIDs are formally grouped poses challenges for comparing findings across a range of studies. In this article, we endeavour to consider both formal and functional categorisations of MIDs, and ultimately argue that functional perspectives may better serve to avoid such incoherence. We discuss our approach to the categorisation of both modal forms and functions in Section 3.

2.2 *Corpus research on modality in British English*

The following section reviews several corpus studies that have investigated modality in recent British English.

2.2.1 *Modality in written British English*

Due to the wider availability of written (as opposed to spoken) English corpora, corpus-based research into change in modality in contemporary British English has been concerned mostly with written English. The research in this area points towards a general pattern of decline from the

middle of the 20th century. Leech (2003) examined change in the use of modal auxiliaries (e.g. *may, should, must*) and semi-modals / 'emergent' modals (e.g. *be able to, be going to, have got to*) in British and American written English between the 1960s and 1990s, using four corpora: LOB, Brown, F-LOB and Frown (the 'Brown quartet', Bowie et al. 2013: 58). Leech found that modal auxiliaries had decreased significantly in frequency in written British English, and that, simultaneously, semi-modals had increased significantly. This finding was questioned by Millar (2009) who, by investigating the much larger TIME Magazine corpus (Davies 2007), found an overall pattern in growth between the 1920s and the 2000s. In response to Millar (2009), Leech (2011) asserted that the variation observed in the TIME Magazine corpus (comprising just one genre of written American English) could not be assumed to be representative of English in general, and (using newly-compiled Brown Family corpora from the 1900s and 1930s) argued that modal verbs had decreased in usage in British and American English during the latter half of the 20th century, having peaked somewhere between the 1930s and 1960s. In addition, beyond the core modals and semi-modals, Leech (2013) provided further evidence of a decline in modality by examining a set of nearly 40 'lexical expressions of modality' (p. 108; e.g. *be obliged to, certainly, perhaps, seem*), observing a decrease of almost 12% over a 75-year period.

In offering possible explanations for the decline of core and lexical modality (and the rise of semi-modals) in written English, Leech (2003) demonstrates that modal semantics is often involved in frequency change (i.e. the change in frequency of a particular semantic function drives an overall change in frequency). For example, the decline of *may* and *should* can be explained by 'a trend towards monosemy', whereby the dominant function becomes even more dominant and the minor functions become less frequent (p. 234). Furthermore, Leech (2013) draws upon the theories of colloquialisation (e.g. Mair 1997; Hundt & Mair 1999; Leech et al. 2009) and grammaticalisation (e.g. Hopper 1991; Rohdenburg 1995; Krug 2000) to present

the most plausible explanation…that grammaticalization of the emergent modals in speech has been associated with increasing frequency, progressively leading to competition with the core modals, which consequently have been undergoing decline in recent English (Leech 2013: 114).

Although we are only interested in spoken (and not written) British English in this article, our brief review of research on modality in written English suggests, through the theory of colloquialisation, that changes in spoken English modality may lead the way for changes in written modality. Therefore, the overall trends observed in written British English are worth discussing, as the changes we observe in present-day spoken English may serve to predict near-future developments in written English. However, it is worth remembering that the verbs that Leech includes in these formal sets of semi-modals and lexical expressions of modality differ from those used in other research on modal forms, rendering much of the literature in this area difficult to compare.

2.2.2 *Modality in spoken British English*

Although most of Leech's (2003) research focussed on written English, he also conducted an analysis of modality in spoken British English, using two 80,000-word samples from the Diachronic Corpus of Present-day Spoken English (DCPSE; Aarts et al. 2002), one each from the London-Lund Corpus (LLC; Svartvik 1990) and ICE-GB (Nelson et al. 2002) sub-corpora. Considering the dearth of available spoken data from time periods comparable to the LOB/Brown (1960s) and F-LOB/Frown (1990s) corpora, the use of such small datasets is understandable; nonetheless, the extent to which we ought to generalise Leech's (2003: 231) finding–a 17.3% decrease in the use of core modals between the DCPSE samples–should be considered. This limitation is discussed by Leech (2013), who shows that, when frequencies of the core modals and semi-modals are combined, there is only a non-significant decrease in overall modal frequency

between the DCPSE sample, 'so we must assume that the case for a 'modality deficit' in the spoken data…is unproven' (Leech 2013: 107). The same caution can be applied to the claims of Smith (2003), who used the same corpus samples to analyse the use of the 'obligation/necessity markers' (i.e. deontic modality): *must, need*, *(have) got to*, *have to* and *need to*; in the spoken data, Smith (2003) reports a decrease of 11.7% between the DCPSE samples. Leech's recognition of the impact of formal groupings on observable trends is important, as the MIDs that constitute these formals sets may not behave homogeneously.

A much larger sample of spoken British English from the early 1990s was made available in the form of the Spoken BNC1994 (BNC Consortium 2007), facilitating fresh studies of modality with a more solid empirical foundation. Paradis (2003), for example, examined the functions of adverb *really* in the LLC and in COLT (the Bergen Corpus of London Teenage Language, Haselrud & Stenström 1995), which is a sub-corpus of the Spoken BNC1994. She finds that '*really* is pragmatically conditioned by the speaker's wish to qualify an expression epistemically with judgements of truth' (Paradis 2003: 214). Nokkonen (2006) also used COLT, in comparison with the LLC (spoken) and LOB and FLOB (written) corpora, to examine the semantic functions of the semi-modal *need to*. She found that, of the four corpora, COLT had the strongest deontic examples of *need to* and also the most examples of the newly emerging epistemic function. Another example is Verhulst et al. (2013), who used random samples of *should, ought to* and *be supposed to* from the BNC1994 to refine theoretical approaches to root necessity.

Leech (2013) used the demographically-sampled sub-corpus of the Spoken BNC1994 and the Santa Barbara Corpus of Spoken American English (SBCSAE; Du Bois et al. 2000-2005) to compare modality in 1990s British and American English conversation. He found 'a much higher incidence' (p. 111) of semi-modals in these corpora compared to the much smaller DCPSE samples; this adds considerable weight to the observed pattern of increasing competition between the core and semi-modals. Leech (2013) also conducted an apparent-time analysis of modals in the

BNC data (using the age groups in the speaker metadata), which very clearly shows a rise in usage of semi-modals from the oldest to the youngest speakers (p. 113), suggesting that the semi-modals have become more popular over time. Of course, it is worth noting that among Leech's semi-modal are forms such as *going to, have to, need to, got to, be supposed to, be able to*, while his core modals include the typical nine auxiliary verbs we might expect as well as *ought* and *need*.

Due to a dearth of available data, few studies have investigated modality in British English as spoken any later than the early 1990s, and most of those that have done so (e.g. Tagliamonte 2004; Fehringer & Corrigan 2015) have investigated a specific regional dialect of British English (York and Tyneside, respectively) rather than a national sample. However, Baker & Heritage (forthcoming) conduct a diachronic analysis of the modal *may* by comparing the Spoken BNC1994 and the Spoken BNC2014 (Love et al. 2017). They find that the overall usage of *may* is lower in the 2010s data, but that frequency is highly variable across speaker age groups. They also find some evidence of a functional shift in *may*, where polite requests (e.g. *may I have some milk*) have given way to hedging propositions (e.g. *you may want to go*).

In summary, it seems certain that there has been a decline in the usage of core and lexical MIDs, and a rise in semi-modals, in written British English, over the course of the 20th century. There is some evidence of a more extreme version of this pattern in conversational British English (Leech 2013), and there appear to be other functional effects that are most salient in speech (e.g. Nokkonen 2006). While some research has been conducted on modality in regional varieties of 21st century spoken British English (e.g. Fehringer & Corrigan 2015), or on a specific modal in national corpora (Baker & Heritage, forthcoming), what is lacking is a general perspective on how modality in spoken British English has changed since the 1990s, and whether the attested patterns of the 1960s – 1990s have continued since then. The aim of this paper is to investigate the expression of modality in spoken British English in the 1990s and 2010s. The release of the Spoken BNC2014 (Love et al. 2017) facilitates the comparison of informal spoken British English with its

spoken counterpart from the BNC1994 (BNC Consortium 2007), affording a comparison of data from the 1990s and 2010s. By isolating one register of speech, we minimise unwanted variation that may be caused by genre differences (cf. Bowie et al. 2013). The trade-off of this approach is a sacrifice of genre representativeness; we acknowledge that we are not speaking of 'spoken English' in general but of a specific register – informal spoken British English (as spoken mainly in England; Love 2020).

2.3 *Research Questions*

This study aims to provide a broad perspective on how modality may have changed in informal spoken British English since the 1990s. We explore this using the following Research Questions:

- RQ1. How has the frequency of modality-indicating devices (MIDs), categorised into three sets (core modal auxiliaries, semi-modals and lexical MIDs), changed in informal spoken British English between the 1990s and 2010s?

- RQ2. How have the modal functions of the core modal auxiliaries in informal spoken British English changed between the 1990s and 2010s?

RQ1 aims to explore how the frequencies of core modals, semi-modals and lexical expressions of modality have changed between the two sampling points. RQ2 is intended to set in motion a body of work exploring the functional factors which may be at play in explaining any observed frequency changes; our starting point in this article is to explore the modal functions of the core modal auxiliaries, with a view to expanding our functional approach to other MIDs in future research (e.g. Love & Curry forthcoming). From a macro perspective, our approach to RQ2 is onomasiological in that we are interested in how the function(s) of modality are expressed through the modal auxiliaries and how this may have changed between the 1990s and 2010s in informal spoken British English.

At this stage, we wish to acknowledge the limitations of our approach, much in the same way as we seek to interrogate and problematise prior approaches to the formal and functional investigation of modality. Firstly, it should be noted that we do not consider how the three sets of modality-indicating devices (modal auxiliary, semi-modal and lexical MID) correspond with one another. In future studies, engaging with any such correspondences would offer a more comprehensive onomasiological approach. Secondly, by only considering the core modal auxiliaries in our functional analysis, we are necessarily restricting the observations we can make about how modality is expressed, and how it may have changed, in recent spoken British English.

To summarise, RQ1 is concerned with formally analysing three sets of MIDs in terms of frequency differences between the Spoken BNC1994 and the Spoken BNC2014. RQ2 is concerned with modal function; but, for reasons of space, and our desire to ensure ample opportunity to problematise the treatment of MIDs as 'sets' in our own and others' research, we only consider the functions of the core modal auxiliaries. As mentioned, we aim to complement the analysis of core modal auxiliary functions in future work by investigating the functions of semi-modals and other lexical expressions of modality, in order to gain a fuller understanding of the semantics of modality in 21st century informal British English conversation.

3 DATA AND METHODOLOGY

In this section, we discuss our use of the spoken BNC corpora and our procedure for selecting and analysing the MIDs investigated in this study.

3.1 *Corpus data*

The corpora used in this study are the spoken components of the two British National Corpora, which were sampled from the 1990s and 2010s, respectively. The first is the demographically-sampled sub-corpus of the Spoken BNC1994 (BNC Consortium 2007), which was recorded in

1991–1993 among 1,408 speakers across 153 conversations. The second is the Spoken BNC2014 (Love et al. 2017; McEnery et al. 2017), which was recorded in 2012–2016 among 668 speakers across 1,251 texts. Both corpora comprise solely informal conversational data, recorded mostly among family and friends, and can be said to represent informal spoken British English as spoken mostly in England (the representativeness of both corpora is discussed in detail by Love 2020).

The corpora were accessed and analysed using Sketch Engine (Kilgarriff et al. 2014). In Sketch Engine, the Spoken BNC1994DS (demographically-sampled part) comprises 4,896,645 tokens, while the Spoken BNC2014 comprises 11,832,933 tokens. Both corpora are tagged using the English Penn Treebank tagset (Marcus et al. 1993).

In our analysis, we sought to mitigate against a known limitation of both the Spoken BNC1994DS the Spoken BNC2014, which is the presence of speaker IDs that contribute a very small or very large number of tokens to the corpora. This issue is discussed in detail by Sönning & Krug (forthcoming), who clearly illustrate the challenges that individual speaker under- and over-representations can create. To mitigate against this limitation, we firstly identified and excluded speakers in both corpora who contributed fewer than 500 tokens each. The arbitrary cut-off point of 500 tokens was chosen to maximise the opportunity for each speaker to make a meaningful individual contribution to the data, while avoiding the removal of a substantial portion of each corpus. We identified and excluded a total of 435 speaker IDs in the Spoken BNC1994DS that contribute less than 500 tokens each. By contrast, in the Spoken BNC2014, only 28 speaker IDs of this type were found and excluded. Secondly, we used Welch's t-test to account for individual variation among the remaining speakers (Brezina 2018), including those who contribute relatively large token counts. The application of Welch's t-test is discussed in more detail later in the article. Table 1 summarises the data used in this study.

Table 1. Token and speaker counts for the corpora before and after removal of speakers accounting for fewer than 500 tokens each.

	Spoken BNC1994DS		Spoken BNC2014	
	Token count	Speaker count	Token count	Speaker count
Original material	4,896,645	1,408	11,832,933	668
Removed material	73,576	435	7,314	28
Study material	4,823,069	973	11,825,619	640

3.2 *Methodological approach*

Owing to the diachronic nature of this study, we adopted an integrated horizontal and vertical reading of the corpus data to allow for the effective investigation of language within clearly defined and temporally situated data (cf. Kohnen 2014). We conducted a quantitative analysis of the usage of core modals, semi-modals and a sample of lexical MIDs in both corpora. Then, we conducted a detailed, qualitative corpus study (cf. Verhulst et al. 2013) of the functions of the core modals in both corpora, which allowed us to explore the relationship between shifts in frequency and shifts in the functional use of these forms.

3.2.1 *Selecting the modality-indicating devices*

To address RQ1, we first identified the three categories of MID: (1) core modals, (2) semi-modals and (3) lexical modal expressions and created lists of members of each category (see Table 2). For all MIDs listed in the table, we searched for any morphologically-related forms (e.g. when searching for *possible*, we retrieved the forms *impossibility, impossible, impossibly, possibilities, possibility, possible* and *possibly*). We also retrieved any negative forms (e.g. *oughtn't*).

Table 2. The modality-indicating devices (MIDs) examined in the study.

Modal category	Members
Core modal	*can, could, may, might, must, shall, should, will, would*
Semi-modal	*dare, need [aux], ought to, used to*
Lexical MID	*able, allow, certain, going to, have to, need [main verb, noun], possible, probable, sure, want (to)*

In deciding which forms to assign to each category, we noted the inconsistency with which these distinctions have been made in the literature. Although the nine core modal auxiliaries are fairly consistently reported as such, other forms have also been considered as core modals. Leech (2003: 226), for example, includes *ought to* and *need* among the core modals in his study, on the basis that they are identified as such in the Quirk et al. (1985) grammar. However, a more recent corpus-based grammar (Carter & McCarthy 2006: 657) lists *ought to* and *need* as semi-modals (alongside *dare* and *used to*). In this article, we adopt Carter & McCarthy's (2006) categorisations as, when compared to those of Quirk et al. (1985), Carter & McCarthy's (2006) categorisations, arguably, better suit our data. This is because Carter & McCarthy's (2006) categorisations are based on an analysis of larger and more recently-compiled corpora of spoken English than Quirk et al. (1985), e.g. CANCODE (Carter & McCarthy 2005).

As for the lexical MIDs, we noted Leech's observation that 'lexical modality devices are so numerous that it is scarcely possible to list them exhaustively' (2013: 108). This meant that certain decisions had to be made in order to identify a reasonable selection of lexical MIDs to analyse and present herein. Given the attestation that modality is largely identified by its function, and that lexical MIDs can take the form of verbs, verb constructions, nouns, adjectives, and adverbs (Carter & McCarthy 2006), the ten lexical MIDs were chosen to reflect a range of

epistemic and root modality as well as a range of forms. These MIDs are also frequent in both corpora and have been studied elsewhere (Carter & McCarthy 2006; Leech 2013; Keizer 2018), offering sufficient data to analyse and contextualise them with previous studies. This reduced focus does make it impossible to talk generally about changes in lexical modality and there is an evident need to expand the study of lexical MIDs to dig more deeply into contemporary usage of modal expressions. This is an issue to which we return in Section 4.1.

3.2.2 *Frequency comparison of MIDs across corpora*

For the core modals, we used part-of-speech (POS) tagging to assist our retrieval, by searching for each form tagged as modal (tag: MD). Each core modal was sampled using a 95% confidence (+/- 5%) sample (Israel 1996; Moinester & Gottfried 2014)[3] to avoid arbitrary selection of sample sizes, and the samples were analysed to distinguish core-modals from items erroneously tagged as modal.

For the semi-modals, we adopted Carter & McCarthy (2006) and searched for *ought to*, *need*, *dare*, and *used to* in their various forms and syntactic positioning. As with core modals, semi-modals were analysed to distinguish semi-modals from lexical modals (in the case of *need* and *dare*) and to determine any non-modal forms retrieved in the search. In the case of *dare*, which was very infrequent, it was possible to analyse all examples in both corpora. For the remaining items, 95% confidence (+/- 5%) samples were extracted and analysed. As noted in Table 2, *need* is included in our study as both a semi-modal and a lexical MID. In our analysis, this distinction was made based on syntactic behaviour. Instances of need were categorised as semi-modal when they act as an auxiliary to a main verb (e.g. *I needn't go round that way*), while instances where need is the main verb were categorised as lexical (e.g. *we need some bananas*). In addition, nominal uses of *need* were classed as lexical MIDs, although these are very rare (e.g. *he has no need to*).

---

[3] This was calculated using confidence sampling software: https://www.surveysystem.com/sscalc.htm (last accessed 05 September 2020).

For lexical modal expressions, a function-to-form approach was necessary to identify relevant forms. Following Aijmer & Rühlemann (2014), O'Keeffe (2018) and Curry (forthcoming), for this initial study of lexical modality, we identified lexical MIDs based on the formal findings from previous analyses (Carter & McCarthy 2006; Leech 2013; Keizer 2018).

As mentioned, we endeavoured to ensure that we investigated forms that are heavily associated with the expression of modality. For the core modal auxiliaries, we relied on part-of-speech tagging to isolate modal-functioning forms. For the semi-modals and lexical MIDs, 95% confidence (+/- 5%) samples were extracted and analysed to distinguish between MID and non-MID forms. We found that, for both sets, the forms functioned modally to a very high extent. For the semi-modals, 99.4% (1990s) and 99.5% (2010s) of cases in our samples were found to express modality. For the lexical MIDs, 99.2% (1990s) and 98.2% (2010s) of cases in our samples were found to express modality, in the 1990s and 2010s, respectively). Across the two sets, this sampling procedure involved the qualitative coding of a total of 1,429 concordance lines. Given that (a) the proportion of non-modal usage (e.g. *I dare you*) is so low across all samples, and (b) adjusting the relative frequency per speaker for each form to remove the small proportion of non-modal usage would be methodologically challenging, we decided not to adjust the frequencies. In making this decision, we acknowledge the presence of a very small proportion of non-modal usage within our reported findings.

The relative frequency of each of the MIDs (per speaker) was retrieved from both corpora and then compared statistically using Welch's independent samples t-test (Welch 1947) and the Cohen's *d* effect size measure (Cohen 1988). Welch's t-test was used to compare the mean relative frequency of each MIDs across all speakers in each corpus, thus taking into account individual speaker variation (Brezina 2018: 187); this approach has been shown to better reflect the reality of the data when compared to an aggregate data methodology that analyses the data wholesale and uses a statistic like log-likelihood (Brezina & Meyerhoff 2014). Cohen's *d* was then used to

evaluate the size of the frequency difference between the corpora (Brezina 2018: 190). The combination of Welch's t-test and Cohen's *d* is recommended by Brezina (2018) for comparing the occurrence of linguistic variables between two groups of speakers (or corpora).

3.2.3 *Semantic function analysis of core modal auxiliaries*

To address RQ2, we manually categorised random samples of each of the core modals in both corpora according to modal function. The size of the sample taken for each core modal (Table 3) was determined by 95% confidence samples (+/-5%).

Table 3. Sizes of each random sample of the core modals.

Core modal	Spoken BNC1994DS		Spoken BNC2014	
	Raw frequency	Sample size	Raw frequency	Sample size
can	21,404	378	56,366	382
could	7,422	367	22,197	378
may	560	240	1,421	303
might	3,287	347	10,207	371
must	2,843	341	4,840	357
shall	1,413	308	2,135	327
should	4,051	354	10,507	371

will	26,282	380	46,851	382
would	15,211	376	50,659	382
TOTAL	82,473	3,091	205,183	3,253

The samples were categorised qualitatively according to modal function via close inspection of concordance lines, which were used to gain an understanding of the immediate linguistic context in which each MID occurred. Simple category labels, which conveyed the basic modal sense of the MID (e.g. *permission*), were used during the analysis, and later attributed to the three broad modality types discussed in Section 2.2.1 (epistemic, deontic and dynamic modality). Table 4 shows each of the modal function categories used in the analysis and their corresponding definition that guided our categorisation.

Table 4. Category labels and corresponding modality types used in the qualitative semantic analysis.

Modal Function Category	Definition
epistemic	hypotheticality, possibility
deontic	instruction, intention, obligation, offering, permission, preference, promising, refusal, request, suggestion
dynamic	ability, habit
none	N/A, unclear

The *N/A* (not applicable) label was used for instances where tagging errors introduced non-modal forms (e.g. the month of May). The *unclear* label was used for cases where the immediate linguistic context afforded by the concordance lines was not sufficient to determine the modal function.

The qualitative analysis was split evenly between co-authors, with regular reviewing of each other's categorisations used to maximise inter-rater reliability. An inter-rater reliability test was also conducted, with both raters categorising the same set of 600 concordance lines, containing an approximately even number of core modals from both corpora. Across all samples, the mean rate of agreement for the broad modal function was 94.42% (kappa coefficient = 0.93, Cohen 1960; this indicates 'almost perfect' agreement, Landis & Koch 1977: 165).

For each core modal, we conducted two measures of the distribution of modality type. The first follows Leech (2003) and reports the percentage of each modal function as a proportion of the total sample of the core modals. However, this alone can be misleading, as such an approach does not take into account changes in the overall usage of the functions in the corpora. For example, Leech (2003: 232) reports that, between the LLC and ICE-GB samples of spoken British English, the epistemic function of *may* increases from 45% to 82% as a proportion of all instances of *may*. This, according to Leech (2003: 234), is evidence of 'a common tendency for the dominant sense in the early 60s to be even more dominant in the early 90s'. However, if instead of only considering the internal distribution of the functions, we consider the corpus frequency of the functions (i.e. the relative frequency of the function as a proportion of the entire corpus size, per million tokens), then it is revealed that the epistemic function of *may* actually decreases from a relative frequency of 488 per million tokens (LLC) to 388 per million tokens (ICE-GB), as part of an overall decrease in the usage of *may* between the two corpora. In real terms, speakers use the epistemic function of *may* less, not more, because they utter the word *may* less frequently overall. This approach provides an alternative view of the functional patterns of the modals, which, we argue, is as important to the interpretation of the functions as the internal distribution. Therefore, we report both types of findings in our analysis. In our case, the corpus frequency of the core modal functions is determined by scaling up the proportions observed in the random samples to extrapolate their frequency across all instances.

4 FINDINGS

In this section, we present the findings of our analysis, starting with the formal analysis of the core modals, semi-modals and lexical MIDS, before turning to the functional analysis of the core modals.

4.1 *Frequency analysis of the core modals, semi-modals and lexical MIDs*

Between the corpora, there is a statistically significant difference in the use of the core modals (as a set), $t(1422.7) = 2.82$, $p = 0.005$, with the Spoken BNC2014 (mean: 16,954.5 per million tokens) containing relatively fewer instances than the Spoken BNC1994DS (mean: 17,695.8 per million tokens). The size of the effect is minimal, $d = 0.15$, 95% CI [0.04, 0.25]. This appears to provide evidence that the occurrence of the core modals has decreased in informal spoken British English between the 1990s and 2010s.

Table 5. Comparing frequency (per million tokens) of the core modals between the Spoken BNC1994DS ('1990s') and Spoken BNC2014 ('2010s') (significant p-values in bold).

Core modal	Mean rel. freq. per speaker (1990s)	Mean rel. freq. per speaker (2010s)	Direction	*p*-value (Welch)	Cohen's *d*	95% confidence interval	Effect
could	1587.95	1836.29	up	**< 0.001**	-0.21	[-0.31, -0.1]	small
might	698.45	798.08	up	**0.01**	-0.13	[-0.24, -0.03]	minimum
would	3074.10	4268.58	up	**< 0.001**	-0.59	[-0.7, -0.48]	medium
must	565.91	353.93	down	**< 0.001**	0.37	[0.26, 0.47]	small

shall	295.36	180.27	down	**< 0.001**	0.27	[0.17, 0.38]	small
will	5662.69	3756.09	down	**< 0.001**	0.74	[0.64, 0.85]	medium
can	4779.58	4780.81	up	0.992	0	[-0.1, 0.1]	minimum
may	127.45	124.68	down	0.867	0.01	[-0.1, 0.11]	minimum
should	868.29	855.82	down	0.763	0.02	[-0.09. 0.12]	minimum

Individually, six out of the nine core modals differ significantly in frequency between the corpora (Table 5). Core modals *could, might* and *would* have a significantly higher frequency in the Spoken BNC2014, whereas *must, shall* and *will* have a significantly lower frequency. The core modals *can, may* and *should* do not differ significantly in frequency between the corpora; the frequency of these core modals appears to have remained stable over time. The most substantial differences among the core modals are the increase of *would* and the decrease of *will*, the latter of which has the highest effect size ($d = 0.74$), and appears to be the driver behind the overall decrease of the core modals as a set.

With regards to the semi-modals, Table 6 shows that two of the four semi-modals differ in frequency significantly between the corpora; both *ought to* and *need* have a significantly lower frequency in the Spoken BNC2014 compared to the Spoken BNC1994DS. In contrast, there is not a significant difference in frequency for *dare* and *used to* between the corpora.

Table 6. *Comparing frequency (per million tokens) of the semi-modals between the Spoken BNC1994DS ('1990s') and Spoken BNC2014 ('2010s') (significant p-values in bold).*

Semi-modal	Mean rel. freq. per	Mean rel. freq. per	Directi on	*p*-value (Welch)	Cohen's *d*	95% confidence interval	Effect

	speaker (1990s)	speaker (2010s)					
ought to	66.88	23.97	down	**<0.001**	0.26	[0.15, 0.36]	small
dare	33.1	27.89	down	0.613	0.03	[-0.08, 0.13]	minimum
need [aux]	20.53†	0†	down	**<0.001‡**	N/A	N/A	N/A
used to	565.29	637.28	up	0.105	-0.08	[-0.19, 0.02]	minimum

†These values were extrapolated from the sample data for *need*. ‡This *p*-value was calculated using log-likelihood,[4] rather than Welch's t-test, since individual speaker frequencies were not available, due to some instances of *need* being classified as lexical.

As a set, the semi-modals have a slightly higher frequency in the Spoken BNC2014 (689.14 per million tokens) compared to the Spoken BNC1994DS (685.8 per million tokens). This difference is not significant ($t(4257.17) = -0.44$, $p = 0.663$; excluding *need*, the frequencies of which were extrapolated from samples), so there is not enough evidence to suggest that the semi-modals as a set have risen in usage over time.

Turning to the lexical MIDs (Table 7), seven MIDs have significantly different frequencies between the corpora. Five MIDs (*able to, need, possible, probable* and *sure*) show evidence of a significant increase in usage over time, while the MIDs *have to* and *want to* have decreased significantly. The difference with the largest effect size is the increase in usage of *probabl**, the frequency of which (per speaker) in the Spoken BNC2014 is almost double its frequency in the Spoken BNC1994DS.

---

[4] Log-likelihood and effect size calculator, Lancaster University: http://ucrel.lancs.ac.uk/llwizard.html (last accessed September 2020).

Table 7. Comparing frequency (per million tokens) of the lexical MIDs between the Spoken BNC1994DS ('1990s') and Spoken BNC2014 ('2010s') (significant p-values in bold).

Lexical MID	Mean rel. freq. per speaker (1990s)	Mean rel. freq. per speaker (2010s)	Directio n	*p*-value (Welch)	Cohen's *d*	95% confidence interval	Effect
able	175.58	228.09	up	**0.006**	-0.14	[-0.25, -0.04]	minimum
need	728.99†	1326.95†	up	**<0.001‡**	N/A	N/A	N/A
possible	99.93	144.29	up	**0.001**	-0.17	[-0.28, -0.07]	minimum
probable	540.33	986.25	up	**<0.001**	-0.64	[-0.74, -0.53]	medium
sure	436.9	567.38	up	**<0.001**	-0.23	[-0.33, -0.12]	small
have to	2448.03	2266.21	down	**0.019**	0.12	[0.02, 0.22]	minimum
want	3254.05	2528.88	down	**<0.001**	0.26	[0.18, 0.39]	small
allow	109.45	138.1	up	0.065	-0.09	[-0.2, 0.01]	minimum
certain	66.73	86.25	up	0.1	-0.09	[-0.19, 0.02]	minimum
going to	1898.97	1858.36	down	0.553	0.03	[-0.07, 0.13]	minimum

†These values were extrapolated from the sample data for *need*. ‡This *p*-value was calculated using log-likelihood, rather than Welch's t-test, since individual speaker frequencies were not available, due to some instances of *need* being classified as semi-modal.

As a set, the lexical MIDs have a higher frequency in the Spoken BNC2014 (10130.76 per million tokens) compared to the Spoken BNC1994DS (9758.96 per million tokens). This difference is not significant (*t*(12823.48) = 0.95, *p* = 0.341; excluding *need*, the frequencies of which were extrapolated from samples), so there is not enough evidence to suggest that the lexical

MIDs as a set have risen in usage over time. As will be discussed, the divergent trends observed among these items leads us to suggest that grouping the lexical MIDs in this way may not be helpful for investigating change in modality.

4.2 *Functional analysis of the core modals*

This section reports on the overall functional findings for the core modals as a set (see Appendix for functional data for each individual core modal). Starting with the internal distribution of modal functions (Figure 1), there appears to be a divergence between epistemic and deontic modality; between the 1990s and 2010s, epistemic modality rises from 50.9% to 52.8%, while deontic modality falls from 40.9% to 38.8% (dynamic modality remains stable at 8.2% and 8.3%, respectively). However, these differences are not significant (epistemic: $t(15.56) = -0.14$, $p = 0.89$; deontic: $t(15.37) = 0.15$, $p = 0.881$; dynamic: $t(15.78) = -0.02$, $p = 0.982$), so there is not enough evidence to claim that the internal distribution of modal functions has changed between the 1990s and 2010s. Despite this, it should be noted that this pattern is replicated by five of the modals (*can, may, might, must, will*; see Appendix).
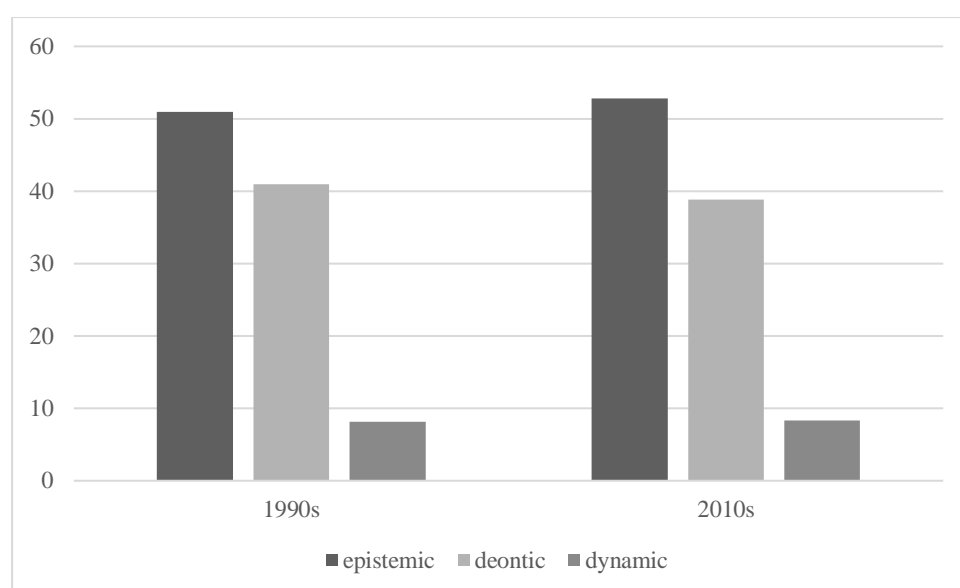

Figure 1. Internal distribution (%) of modal functions across all core modal samples.

As discussed in Section 3.2.3, considering the internal distribution of functions alone has the potential to be misleading, so we also calculated the mean relative frequency of modality types across all core modals (Figure 2).
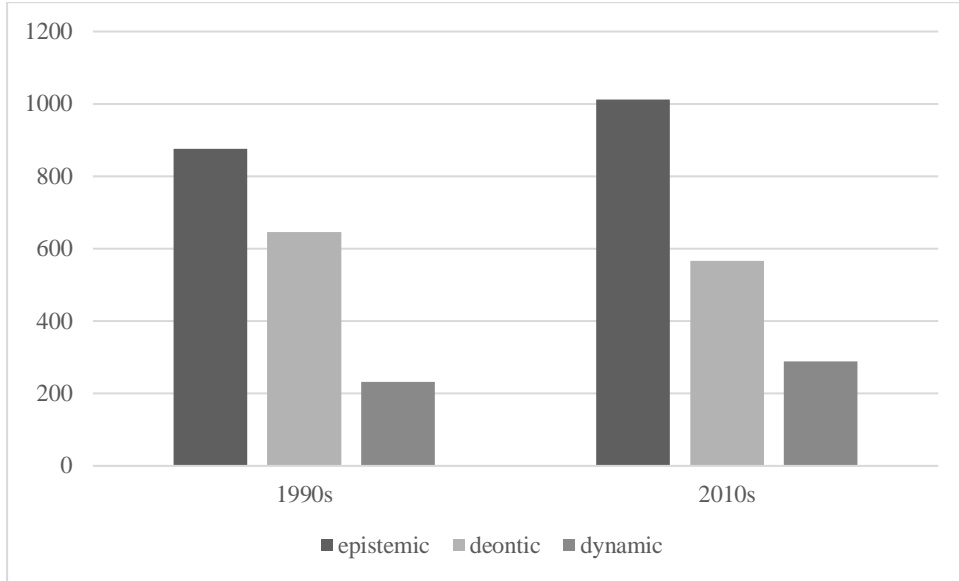

Figure 2. Mean relative corpus frequency (per million tokens) of modal functions across the core modals (extrapolated from samples).

In this case, taking into account relative frequency (i.e. a normalised frequency that allows the comparison of findings from corpora of different sizes) appears to support and amplify the divergence between epistemic and deontic modality. However, once again, these differences are not significant (epistemic: $t(15.67) = -0.28$, $p = 0.78$; deontic: $t(13.8) = 0.25$), $p = 0.808$; dynamic: $t(15.97) = -0.24$, $p = 0.81$).

5 DISCUSSION AND CONCLUSION

In discussing the results, Section 5.1 addresses the findings of the frequency analysis of core modal auxiliaries, semi-modals, and lexical MIDs (RQ1), with a view to interrogating how these findings

may reflect changes (and lack thereof) in modality in spoken British English. Secondly, the findings of the functional analysis of core modals (RQ2) are discussed, considering both formally constrained functional changes and more general changes in the functional behaviour of modality in the core modal auxiliaries. Finally, in Section 5.2, we discuss how this research could be extended in future work.

5.1 *Discussion of findings*

Generally, the presence of core modal forms in informal spoken British English has decreased between the 1990s and 2010s. This decrease is significant and would appear to support Leech's view that core modal auxiliaries are in decline (2003; 2011). Arguably, such a view may not be surprising, as with the decline of core modal auxiliaries in a range of written genres and registers (Leech 2013), the theory of colloquialisation would have predicted that this decline had already been taking place in informal spoken English. However, this argument appears only to be valid when considering the core modal auxiliaries as a homogenous set. Interestingly, our findings illustrate clear divergences within this set, rendering a claim of declining core modal auxiliaries somewhat problematic. While the overall trend of formal decline is significant, the modal forms *could, might* and *would* have significantly increased over time. Similarly, the forms *must, shall* and *will* have significantly decreased, while the remaining core modal auxiliary forms appear to be stable. Therefore*, could, might,* and *would* better reflect Millar's (2009) proposition of increase while *must, shall* and *will* support Leech's stance on core modal auxiliaries in decline (2003; 2011). Furthermore, it appears that the modal *will* plays a key role in determining this general state of decline in core modal auxiliaries, given its large effect size. Essentially, our findings problematise the treatment of core modal auxiliaries as a homogenous set, bound by formal and grammatical

categorisation, given that their semantic function arguably plays a more important role in determining their use. We shall return to this issue later in this section.

With regards to the semi-modals, our results indicate that, as a set, there is not sufficient evidence to suggest that they have changed in frequency over time. However, within the set, both the forms *ought to* and *need* have decreased significantly. This reflects Carter & McCarthy's (2006: 657-660) observation that both forms occur rather infrequently. Evidently, these forms are continuing to decline, according to our results. Conversely, there does not appear to be a significant difference in the frequency of *dare* and *used to* between the 1990s and 2010s. While the wholesale comparison of the set of four semi-modals reflects a very slight increase, this increase is not significant. Leech (2003) discusses the increase in semi-/'emergent' modals, which our findings may appear to contradict. However, the issue here pertains to the label of semi-modals, which is inconsistent in the literature. For the sake of comparability elsewhere, it is important to keep in mind that, while Leech (2003) included a range of forms, including *be going to, be to, (had) better, (have) got to, have to, need to, want to* and *used to* (p. 229), the semi-modals in our study encapsulate only *dare*, *need*, *ought to*, and *used to*. Therefore, it is challenging to discuss this formal set, and it would be better to consider these four forms independently. Further, Leech's focus (predominantly) on written English renders comparability of the findings incommensurable.

For lexical modality, the significant increase in the usage of *able, need (to), possible, probable* and *sure* is noteworthy. In Leech (2013), lexical modality saw a decline of 11.93%, to which Leech attributed a cause of 'modality deficit'. For Leech, this deficit was potentially owing to colloquialisation, with the implication that spoken language would also reflect a decrease in lexical modality. However, according to our data, this does not appear to be the case. That being said, it must be noted that our study is based on only ten lexical MIDs and is by no means exhaustive. Nonetheless, this contrast is an interesting observation. While five lexical MIDs

significantly increased, two (*have to* and *want to*) decreased significantly. Therefore, as with the core modal auxiliaries and semi-modals, the set of lexical MIDs is not homogenous and while, on a case-by-case basis, individual lexical MIDs reflect significant changes over time, the overall use of lexical MIDS studied herein does not appear to have changed significantly. This further challenges the view of lexical MIDs as a homogenous group, defined by their alignment with degrees of grammaticality and lexicality.

Overall, a key problematisation that has emerged in this discussion pertains to the issue of the lack of homogeneity in the literature in the conception of modals as forming sets. In considering MIDs individually, there is evidence of significant increases and decreases in core modal auxiliaries, semi-modals, and lexical MIDs. However, in trying to align our findings on modality to a range of studies, the lack of homogeneity in approaches to categorising modal forms in the literature, and within our own study, means that it is challenging to bring together a wide collection of studies to underpin the discussion of modality and language change in informal spoken British English. Arguably, making general observations about changes in core, semi-, and lexical modality 'sets' can obfuscate and distort changes that occur on a case-by-case basis, formally, and on a functional basis, more generally. Given that modality is a semantic phenomenon (von Wright 1951), a functional perspective–not only semantic but also morphosyntactic–may better serve to offer insight as to whether change in the use of modality is taking place. Therefore, a clear question emerges as to whether a comparison of the use of modal function might produce a more coherent understanding of the behaviour of modals in contemporary British English; based on our initial findings regarding core modal forms, we suggest that future studies move away from the formal categories (core, semi- and lexical MIDs) and instead adopt functional categorisations of modal forms.

In the case of the core modal auxiliaries, the functional analyses revealed a number of shifts in functional distribution. For *can*, *may*, *might, must*, and *will*, there appears to be a slight shift in

modal function, with a greater proportion of each verb performing epistemic modality and a lesser proportion reflecting deontic modality. For the remaining core modal auxiliaries (*could*, *shall*, *should* and *would*), there are only marginal differences in the proportional distribution of function. Among these, *shall* and is noteworthy in that it appears to resist the possible trend towards epistemicity that we have noted above. In the 1990s data, *shall* is almost exclusively used in deontic functions (92.4% of our sample), and this remains the case in the 2010s data (98.2%). Likewise, *should* shifts towards a higher proportion of deontic usage (from 74.0% to 87.9%). Given the high (and apparently increasing) proportion of a single modal function in both *shall* and (to a lesser extent) *should*, an argument could be made for a movement towards monosemy in both cases (Leech 2003)–at least in terms of their broad modal function. However, overall (*shall* notwithstanding), there is no evidence of a general shift towards monosemy in core modal auxiliaries, as Leech (2003) suggests there is, given that most of the modal auxiliaries are observed to perform a range of modal functions.

When considering changes in epistemic, deontic, and dynamic modality in core modal auxiliaries in general, our results show a slight increase in epistemic modality and a slight decrease in deontic. While neither trend is significant, this does remain of interest. At a functional level, it is possible to observe a more coherent trend in modality than the analysis for formal sets affords. For example, although this study's semantic analysis centred on the core modals only, the semi-modals and lexical MIDs that significantly increased (excluding *need*) all show a tendency to reflect meanings attributed to epistemic modality (Carter & McCarthy 2006). Similarly, those semi-modals and lexical MIDs that have significantly declined typically reflect deontic modality meanings, according to Carter & McCarthy (2006); this is a claim we intend to investigate in future. Therefore, while the change in modal function is not significant, the functional perspective offers a coherent overview of modal functions in contemporary British English. It is possible that what we are observing is a change in process with a potential for increased use of epistemic

modality over time. Therefore, future studies of British English modality may benefit from focussing more (or even exclusively) on functional categories instead of formal categories to understand changes in the use of modality over time.

5.2 *Future directions*

This article represents the first stage of a body of work (e.g. Love & Curry forthcoming) that aims to provide a rigorous, inclusive and functional description of recent changes in modality, however expressed, in conversational British English. The contributions that this paper makes towards this goal are:

- a description of frequency differences among core, semi- and a selection of lexical MIDs between the 1990s and 2010s;
- a description of the modal functions of the core modal auxiliaries in the 1990s and 2010s.

The next stage of this work involves:

- extending the frequency analysis of lexical MIDs to include a larger and more representative number of forms;
- conducting apparent-time frequency analyses of all MIDs to complement the real-time analysis presented in this paper (cf. Leech 2013 and Baker & Heritage forthcoming);
- extending the functional analysis of MIDs to include semi-modals and lexical MIDs.

Beyond this, further work should consider changes in the distribution of modalised and non-modalised utterances over time (cf. Biber et al. 1999: 456). In addition, it should provide a complementary functional analysis of changes in the expression of modality between the 1990s and 2010s among other registers of British English, including various genres of writing and e-language. To facilitate this, we anticipate the public release of the written component of the

BNC2014, which complements the spoken component used in this study and provides a point of

comparison for the written component of the BNC1994.[5]

---

[5] The Written BNC2014 is due for full public release in 2020. http://cass.lancs.ac.uk/bnc2014/ [accessed 6 September 2020].

Authors' addresses:

Dr Robbie Love

School of Social Sciences and Humanities

Aston University

Aston Street

Birmingham

B4 7ET

UK

Email: r.love@aston.ac.uk

Dr Niall Curry

Centre for Academic Writing

Coventry University

Priory Street

Coventry

CV1 5FB

UK

Email: niall.curry@coventry.ac.uk

REFERENCES

Aarts, Bas, Joanne Close & Sean Wallis. 2013. Choices over time: methodological issues in investigating current change. In Bas Aarts, Joanne Close, Geoffrey Leech & Sean Wallis (eds.), *The verb phrase in English: Investigating recent language change with corpora*, 14–45. Cambridge: Cambridge University Press.

Aarts, Bas, Sean Wallis, Dirk Bury, Lesley Kirk, Yordanka Kostadinova–Kavalova, Ann Law & Gabriel Ozón. 2002. *DCPSE: The Diachronic Corpus of Present–Day Spoken English*. Distributed by Survey of English Usage, University College London. https://www.ucl.ac.uk/english–usage/projects/dcpse/index.htm [accessed 18 April 2021].

Aijmer, Karin & Christoph Rühlemann. 2014. Introduction. Corpus pragmatics: laying the foundations. In Karin Aijmer & and Christoph Rühlemann (eds.), *Corpus pragmatics*, 1–28. Cambridge: Cambridge University Press.

BNC Consortium. 2007. *The British National Corpus, version 3 BNC XML Edition*. Distributed by Bodleian Libraries, University of Oxford, on behalf of the BNC Consortium. http://www.natcorp.ox.ac.uk/ [accessed 18 April 2021].

Baker, Paul & Frazer Heritage. forthcoming. Corpus approaches to sociolinguistics. In Anne O'Keefe & Michael McCarthy (eds.), *The Routledge handbook of corpus linguistics* (2nd ed.). London: Routledge.

Biber, Douglas, Stig Johansson, Geoffrey Leech, Susan Conrad & Edward Finegan. 1999. *The Longman grammar of spoken and written English*. London: Longman.

Bolinger, Dwight. 1980. *Wanna* and the gradience of auxiliaries. In Gunter Brettschneider & Christian Lehmann (eds.), *Wege zur universalien forschung*, 292–99. Tübingen: Gunter Narr Verlag.

Bowie, Jill, Sean Wallis & Bas Aarts. 2013. Contemporary change in modal usage in spoken British English: mapping the impact of "genre". In Juana I. Marín–Arrese, Marta Carretero, Jorge Arús Hita & Johan van der Auwera (eds.), 57–94.

Brezina, Vaclav. 2018. *Statistics in corpus linguistics: A practical guide*. Cambridge University Press.

Brezina, Vaclav & Miriam Meyerhoff. 2014. Significant or random? A critical review of sociolinguistic generalisations based on large corpora. *International Journal of Corpus Linguistics* 19(1), 1–28.

Bybee, Joan, William Pagliuca & Revere D. Perkins. 1991. Back to the future. In Elizabeth Closs Traugott and Bernd Heine (eds.), Vol 2, 17–58.

Carter, Ronald & Michael McCarthy. 1995. Grammar and the spoken language. *Applied Linguistics* 16(2), 141-58.

Carter, Ronald & Michael McCarthy. 2006. *Cambridge grammar of English: A comprehensive guide to spoken and written English grammar and usage*. Cambridge: Cambridge University Press.

Coates, Jennifer. 1983. *The semantics of the modal auxiliaries*. London: Croom Helm.

Cohen, Jacob. 1960. A coefficient of agreement for nominal scales. *Educational and Psychological Measurement* 20(1), 37–46.

Cohen, Jacob. 1988. *Statistical power analysis for the behavioural sciences*. Hillsdale, NJ: Erlbaum.

Collins, Peter. 2009. *Modals and quasi–modals in English*. Amsterdam & New York: Rodopi.

Curry, Niall. forthcoming. Identifying and analysing indirect questions in English, French, and Spanish academic writing: Addressing methodological challenges in function–based corpus and contrastive analyses. *International Journal of Corpus Linguistics.*

Davies, Mark. 2007. *TIME Magazine Corpus 100 million words, 1920s–2000s.* http://corpus.byu.edu/ [accessed 18 April 2021].

Du Bois, John W., Wallace L. Chafe, Charles Meyer, Sandra A. Thompson, Robert Englebretson & Nii Martey. 2000–2005. *Santa Barbara Corpus of Spoken American English, Parts 1–4.* Philadelphia: Linguistic Data Consortium. https://www.linguistics.ucsb.edu/research/santa–barbara–corpus [accessed 18 April 2021].

Facchinetti, Roberta, Manfred Krug & Frank Palmer (eds.). 2003. *Modality in contemporary English.* Berlin: Mouton de Gruyter.

Fairclough, Norman. 2003. *Analysing discourse: textual analysis for social research.* New York: Routledge.

Fehringer, Carol & Karen Corrigan. 2015. 'You've got to sort of eh hoy the Geordie out': modals of obligation and necessity in fifty years of Tyneside English. *English Language and Linguistics* 19(2), 355–81.

Haslerud, Vibeke & Anna–Brita Stenström. 1995. The Bergen corpus of London teenager language COLT. In Geoffrey Leech, Greg Myers & Jenny Thomas (eds.), *Spoken English on computer*, 235–42. Harlow: Longman.

Hopper, P. J. 1991. On some principles of grammaticisation. In Elizabeth C. Traugott & Bernd Heine (eds.), Vol. 1, 17–36.

Hundt, Marianne & Christian Mair. 1999. "Agile" and "uptight" genres: The corpus–based approach to language change in progress. *International Journal of Corpus Linguistics* 4(2), 221–42.

Israel, Glenn D. 1996. *Determining sample size.* University of Florida: IFAS Extension, PEOD–6.
https://www.psycholosphere.com/Determining%20sample%20size%20by%20Glen%20Israel.pdf [accessed 18 April 2021].

Keizer, Evelien. 2018. Modal adverbs in FDG: Putting the theory to the test. *Open Linguistics*, 4(1), 356–90.

Kilgarriff, Adam, Vít Baisa, Jan Bušta, Miloš Jakubíček, Vojtěch Kovář, Jan Michelfeit, Pavel Rychlý & Vít Suchomel. 2014. The Sketch Engine: ten years on. *Lexicography* 1, 7–36. http://www.sketchengine.eu [accessed 25 August 2020].

Kohnen, Thomas. 2014. Speech acts: A diachronic perspective. In Karen Aijmer and Christoph Rühlemann (eds.), 52–83.

Krug, Manfred. 2000. *Emerging English modals: A corpus–based study of grammaticalisation*. Berlin & New York: Mouton de Gruyter.

Landis, J. Richard, & Gary G. Koch. 1977. The measurement of observer agreement for categorical data. *Biometrics* 33, 159–74.

Leech, Geoffrey. 1993. 100 million words of English. *English Today* 9–15.

Leech, Geoffrey. 2003. Modality on the move: The English modal auxiliaries 1961–1992. In Roberta Facchinetti, Manfred Krug & Frank Palmer (eds.), 223–40.

Leech, Geoffrey. 2011. The modals ARE declining. Reply to Neil Millar's 'Modal verbs in TIME: Frequency changes 1923–2006'. *International Journal of Corpus Linguistics* 16(4), 547–64.

Leech, Geoffrey. 2013. Where have all the modals gone? An essay on the declining frequency of core modal auxiliaries in recent standard English. In Juana I. Marín–Arrese, Marta Carretero, Jorge Arús Hita & Johan van der Auwera (eds.), 95–116.

Leech, Geoffrey, Marianne Hundt, Christian Mair, & Nick Smith. 2009. *Change in Contemporary English: A grammatical study*. Cambridge: Cambridge University Press.

Love, Robbie. 2020. *Overcoming challenges in corpus construction: The Spoken British National Corpus 2014*. New York: Routledge.

Love, Robbie & Niall Curry. forthcoming. *Modality in spoken British English: Recent change in casual conversation*. New York: Routledge.

Love, Robbie, Claire Dembry, Andrew Hardie, Vaclav Brezina, & Tony McEnery. 2017. The Spoken BNC2014: Designing and building a spoken corpus of everyday conversations. *International Journal of Corpus Linguistics*, 22(3), 319–44.

Mair, Christian. 1997. Parallel corpora: A real–time approach to the study of language change in progress. In Magnus Ljung (ed.), *Corpus–based studies in English: Papers from the Seventeenth International Conference on English Language Research on Computerized Corpora ICAME 17,* 195–209. Amsterdam: Rodopi.

Marcus, Mitchell P., Beatrice Santorini, & Mary Ann Marcinkiewicz. 1993. Building a large annotated corpus of English: The Penn Treebank. *Computational Linguistics*, 19(2), 313–30.

Marín–Arrese, Juana I., Marta Carretero, Jorge Arús Hita & Johan van der Auwera (eds.). 2013. *English modality: Core, periphery and* evidentiality. Berlin: Walter de Gruyter.

McEnery, Tony & Andrew Hardie. 2012. *Corpus linguistics: method, theory and practice*. Cambridge: Cambridge University Press.

McEnery, Tony, Robbie Love, Andrew Hardie, Vaclav Brezina, Claire Dembry & Matt Timperley. 2017. *The Spoken British National Corpus 2014.* corpora.lancs.ac.uk/bnc2014 [accessed 6 September 2020].

Millar, Neil. 2009. Modal verbs in TIME: Frequency changes 1923–2006. *International Journal of Corpus Linguistics*, 14(2), 191–220.

Moinester, Murray & Ruth Gottfried. 2014. Sample size estimation for correlations with pre–specified confidence interval. *The quantitative methods for psychology* 10(2), 124–30.

Mubarak, Rehana. 2015. *Politeness strategies in different varieties of English: De– and encoding of modality in speech acts from a translational, editorial and transcultural perspective considering methods of automatic text processing*. PhD dissertation, University of Vienna. http://othes.univie.ac.at/39453/1/2015–06–02_0968558.pdf [accessed 18 April 2021].

Nelson, Gerald, Sean Wallis & Bas Aarts. 2002. *Exploring natural language: Working with the British component of the International Corpus of English*. Amsterdam: John Benjamins.

Nokkonen, Soili. 2006. The semantic variation of NEED TO in four recent British English corpora. *International Journal of Corpus Linguistics* 11(1), 29-71.

Nordlinger, Rachel & Elizabeth Close Traugott. 1997. Scope and the development of epistemic modality: Evidence from *ought to*. *English Language and Linguistics* 1(2), 295–317.

O'Keeffe, Anne. 2018. Corpus–based function–to–form approaches. In Andreas H. Jucker, Klaus P. Schneider, & Wolfram Bublitz (eds.), *Methods in pragmatics*, 587–618. Berlin: Mouton de Gruyter.

Palmer, Frank. 2001. *Mood and modality*, 2nd ed. Cambridge: Cambridge University Press.

Palmer, Frank. 2003. Modality in English: Theoretical, descriptive and typological issues. In Roberta Facchinetti, Manfred Krug & Frank Palmer (eds.), 1–17.

Paradis, Carita. 2003. Between epistemic modality and degree: the case of *really*. In Roberta Facchinetti, Manfred Krug & Frank Palmer (eds.), 191–222.

Perkins, Michael R. 1983. *Modal Expressions in English*. Norwood, NJ: Ablex.

Põldvere, Nele, Victoria Johansson & Carita Paradis. Forthcoming. *A guide to the London-Lund Corpus 2 of spoken British English*. Lund Studies in English 121.

Quirk, Randolph, Sidney Greenbaum, Geoffrey Leech & Jan Svartvik. 1985. *A comprehensive grammar of the English language*. London: Longman.

Rohdenburg, Günter. 1995. On the replacement of finite complement clauses by infinitives in English. *English Studies* 16, 367–88.

Smith, Nick. 2003. Changes in the modals and semi–modals of strong obligation and epistemic necessity in recent British English. In Roberta Facchinetti, Manfred Krug & Frank Palmer (eds.), 241–66.

Sönning, Lukas & Manfred Krug. forthcoming. Comparing study designs and down–sampling strategies in corpus analysis: The importance of speaker meta–data. In Ole Schützler & Julia Schlüter (eds.), *Data and methods in corpus linguistics – comparative approaches*. Cambridge: Cambridge University Press.

Svartvik, Jan. (ed.). 1990. *The London–Lund Corpus of Spoken English: Description and research*. Lund: Lund University Press.

Tagliamonte, Sali A. 2004. *Have to, gotta, must?*: Grammaticalisation, variation and specialization in English deontic modality. In Hans Lindquist & Christian Mair (eds.), *Corpus approaches to grammaticalization in English*, 33–6. Amsterdam: John Benjamins.

Traugott, Elizabeth Closs & Bernd Heine (eds.). 1991. *Approaches to grammaticalization*. Amsterdam: John Benjamins.

Verhulst, An, Ilse Depraetere & Liesbet Heyvaert. 2013. Source and strength of modality: An empirical study of root *should*, *ought to* and *be supposed to* in Present–day British English. *Journal of Pragmatics* 55, 210–25.

von Wright, G. H. 1951. *An essay in modal logic*. Amsterdam: North Holland.

Welch, B. L. 1947. The generalization of 'Student's' Problem when several different population variances are involved. *Biometrika* 34(1–2), 28–35.

Appendix

Internal distribution and relative frequency of the modality types among the core modals in the Spoken BNC1994DS and Spoken BNC2014.

		Internal %		Relative frequency (corpus)	
Core modal	Modality type	Spoken BNC1994DS	Spoken BNC2014	Spoken BNC1994DS	Spoken BNC2014
can	Epistemic	36	43	1,455.80	1,996.41
	Deontic	28	25	1,127.07	1,172.89
	Dynamic	36	32	1,467.54	1,509.79
could	Epistemic	35	37	490.59	660.44
	Deontic	33	35	465.43	620.71
	Dynamic	32	28	444.46	501.53
may	Epistemic	70	81	74.50	86.06
	Deontic	30	19	31.93	20.62
	Dynamic	0	0	0.00	0.00
might	Epistemic	81	83	491.01	704.93
	Deontic	19	17	111.95	148.90
	Dynamic	0	0	0.00	0.00
must	Epistemic	76	88	420.05	355.40

	Deontic	24	12	129.65	48.15
	Dynamic	0	0	0.00	0.00
shall	Epistemic	7	2	19.02	3.31
	Deontic	92	98	243.51	177.23
	Dynamic	0	0	0.95	0.00
should	Epistemic	26	12	208.79	105.37
	Deontic	74	88	593.16	766.36
	Dynamic	0	0	0.00	0.00
will	Epistemic	49	59	2538.19	2281.68
	Deontic	50	39	2566.87	1493.46
	Dynamic	1	2	57.36	72.60
would	Epistemic	77	71	2182.25	2915.70
	Deontic	19	16	537.17	650.43
	Dynamic	4	13	117.51	515.85