Heike Pichler has recently completed her PhD on discourse variation at the Centre for Linguistic Research at the University of Aberdeen, where she is currently teaching sociolinguistics at undergraduate and postgraduate levels. Her research focuses on discourse markers and their formal, functional and social distribution in interaction.
ABSTRACT

This paper explores the functional and social meanings of the formal variants of the expressions *I DON’T KNOW* and *I DON’T THINK*. The qualitative analysis of the data shows that *I DON’T KNOW* and *I DON’T THINK* are highly routinized discourse formulae that perform multiple interpersonal and textual functions. The quantification of their phonetic and grammatical variants across function, age and gender reveals that the occurrence of non-localised and localised variants of *I DON’T KNOW* and *I DON’T THINK* is conditioned by different parameters. Non-localised variants show function-specific patterning: variants of *I DON’T KNOW* pattern with broad functional categories; variants of *I DON’T THINK* reflect fine-grained functional differences at a micro-level of analysis. Localised variants, by contrast, do not carry a functional load but evince orderly social heterogeneity. The results demonstrate that social variation in discourse extends beyond differential frequencies and strategic uses of discourse markers (DMs) to their formal encoding, and that function plays an important role in accounting for patterns of formal variation in discourse. The paper argues for a fuller integration of qualitative methods in the study of language variation and change in discourse, and calls for the form and function of discourse variables to be studied in greater detail.
1. Introduction

The publication of Deborah Schiffrin’s (1987) seminal work on discourse markers (henceforth DMs) in the late 1980s triggered an unprecedented interest in linguistic features which, for the purpose of this paper, are broadly defined as linguistic items or expressions that have little or no referential meaning but serve multiple pragmatic functions in the interpersonal and textual domains of discourse, i.e., the domains concerned with the expression of speaker attitudes and the structure of discourse as text (see further Section 4).\(^2\)

Scholars investigating DMs have been concerned with the description of their functional repertoires (e.g. Aijmer 2002; Erman 1987; Kärkkäinen 2003) as well as their diachronic development (e.g. Brinton 1996, 2006; Thompson and Mulac 1991; Traugott 1995). An abundance of variationist studies have examined the correlation of the strategic uses and relative frequencies of DMs with social variables, notably socio-economic class (e.g. Macaulay 2002b, 2005), gender (e.g. Holmes 1995; Levey 2006; Macaulay 2005) and age (e.g. Andersen 2001; Erman 2001; Levey 2006; Macaulay 2002b, 2005), as well as with speaker role (e.g. Fuller 2003) and communities of practice (e.g. Moore and Podesva 2008). More recently, sociolinguists have also been concerned with cross-dialectal comparisons in the social and functional distribution of DMs and with the dissemination of innovative forms across geographical space (e.g. Buchstaller 2008; Buchstaller and D’Arcy fc.; Cheshire et al. 2005).

Yet despite the fact that DM analysis is “a growth market in linguistics” (Fraser 1998: 301), there have been few investigations into the variation in the formal encoding of DMs. The limited number of studies which have appeared demonstrate that social factors, notably age (e.g. Ito and Tagliamonte 2003), gender (e.g. Tagliamonte and D’Arcy 2007), socio-economic class (e.g. Cheshire 2007) and education (e.g. Ferrara 1997), as well as discourse function (e.g. Lindemann and Mauranen 2001; Stenström 1998; Scheibman 2000;) impact on the
distribution of lexical and phonetic variants of DMs. While variation in the social patterning of discourse variants has been attributed largely to “shifting norms and practices in the community” (Ito and Tagliamonte 2003: 257), function-specific patterning of discourse variants has been attributed to changes associated with grammaticalization, which involve semantic bleaching, pragmatic strengthening and phonetic reduction (Hopper and Traugott 2003). Lindemann and Mauranen (2001), Stenström (1998) and Scheibman (2000) found that phonetically full forms of JUST, BECAUSE and I DON’T KNOW correlate with older layers of meaning (“exactly” for JUST, grammatical subordination for BECAUSE, lack of knowledge for I DON’T KNOW), while phonetically reduced forms correlate with newer, grammaticalized layers of meaning (mitigation for JUST, continuation signalling for BECAUSE, face-saving and turn-yielding for I DON’T KNOW).³ The studies listed above, then, highlight that formal discourse variation is not random but displays systematic patterning along social factors and discourse function.

While previous studies of formal discourse variation provide interesting insights into the distribution of non-localised and supra-local discourse variants, i.e., variants that have a wide geographical distribution, they leave open the question of whether localised discourse variants, i.e., variants whose distribution is associated with a given locality, display similar patterns of distribution across social factors and discourse function. This paper seeks to address the current dearth of information regarding the distribution of localised discourse variants. By combining qualitative methods from conversation analysis and quantitative methods from Labovian sociolinguistics, I will explore the differential distribution of non-localised and localised variants of the expressions I DON’T KNOW and I DON’T THINK in the variety of English spoken in Berwick-upon-Tweed, a town in the far north-east of England.

The extracts in (1) and (2) set out the formal variation of the selected DMs.¹ The nature of the formal variation, which affects the phonetic and grammatical encoding of negative auxiliary
DO, is detailed in Section 3.2. The variation is here and throughout represented by means of variation in orthography, with each orthographic form corresponding to a different degree of reduction (I don’t know, I dono, I dunno; I don’t think, I doØ think) or a different grammatical form of negative DO (I divn’t knaa; I divn’t think).

(1) a. Keith: So (...) I divn’t knaa. I’m pretty I I would say I’m English. But I dono.

   b. HP: Why not.

      Natalie: @ I don’t know? (.) I dunno? I just don’t like being called a Berwicker.

(2) a. HP: When you’re married.

      Ryan: I don’t think I’ve got one for when you’re married. I doØ think there is one sort of like.

   b. HP: Do you think that you have nothing in common with the rest of the North-East?

      Or

      Jane: We haven’t really.

      Helen (h) Well, [I] divn’t think we have.

      HP: [yeah]

      Jane: No. I divn’t think we have.

In what follows I present evidence suggesting that the distribution of the non-localised and localised variants of I DON’T KNOW and I DON’T THINK illustrated in (1) and (2) is conditioned by different parameters: while non-localised discourse variants display function-specific patterning, localised discourse variants, like localised phonological variants, exhibit orderly social heterogeneity. I demonstrate that the same variants of negative DO do not carry the same functional or social meanings across the selected DMs. The differential meanings are attributed to different outcomes of the processes involved in the DMs’ grammaticalization and to interactional constraints. Finally, I posit that DMs should be viewed as integral components of the grammar of a language (Traugott 1995). The evidence of my data points to the need for
a firmer integration of qualitative methods and discourse variables in the study of language variation and change.

The paper is structured as follows. I begin with a brief review of the literature on I DON’T KNOW and I DON’T THINK in Section 2. Section 3 provides details on the data and methods of analysis. This section also introduces the variables and their variants and circumscribes the variable contexts. In Section 4, I provide a brief overview of the main functions performed by I DON’T KNOW and I DON’T THINK in the present dataset. The results of the quantitative analysis in terms of the variants’ overall as well as functional and social distribution are presented in Sections 5 to 7. Section 8 discusses the findings and finally, Section 9 is the conclusion to the paper.

2. The expressions I DON’T KNOW and I DON’T THINK

Discourse and conversation analysts downplay the importance of I DON’T KNOW and I DON’T THINK as verbal representations of a cognitive state or process. They highlight the expressions’ interactional importance for signalling epistemicity and mitigation, and contributing to the construction of coherent discourse. Because the expressions’ functional repertoires are discussed and illustrated at length in Section 4, I will only briefly review the qualitative literature here.

Most studies of I DON’T KNOW have described its use in terms of politeness, i.e., as a device that serves to save speakers’ and hearers’ face (Beach and Metzger 1997; Potter 2004; Tsui 1991; Weatherall 2007; Wooffitt 2005). Textual uses of I DON’T KNOW as regards turn-exchange or topic-development have been noted but mostly in passing (Beach and Metzger 1997; Ford and Thompson 1996; Hutchby 2002; Östman 1981; Schegloff 1996; Weatherall 2007). The discourse role of I DON’T THINK has been described solely in terms of speaker
attitudes (Bublitz 1992; Nuyts 1990; Simon-Vandenbergen 1998). Where the negative in I DON’T THINK has scope over the proposition that follows (rather than over think), the expression serves to signal speakers’ uncertainty vis-à-vis the validity of this proposition and to mitigate potential face-threats. No mention has been made in the literature of textual functions served by I DON’T THINK.

Even without reviewing the qualitative literature in great detail, it is evident that I DON’T THINK and particularly I DON’T KNOW exhibit multiple discourse functions. Like other DMs, they are polypragmatic. Bybee and Scheibman (1999) posit that the variables’ multifunctionality reflects their status as grammaticalized discourse units. They argue that as a result of their frequency in discourse, I DON’T KNOW and I DON’T THINK have lost their internal structure, become automated as single processing units, and weakened their association with their composite parts (I, don’t, know/think); in the process, they have generalized their meanings, developed new discourse functions, and undergone phonetic reduction and, in the case of I DON’T KNOW, fusion.

Bybee and Scheibman’s analysis of these expressions in American English conversational data focused on their formal and functional distribution. Depending on the realization of the vowel in don’t as either full, [o], or reduced, [ǝ], Bybee and Scheibman classified the 37 tokens of I DON’T KNOW and 19 tokens of I DON’T THINK in their data into two categories: full vowel variants and reduced vowel variants. In the case of I DON’T KNOW, the distributional analysis revealed form-function correlations that were later summarized by Scheibman (2000: 120) as follows: “[a]ll variants of don’t in I don’t know convey the expression’s lexical meaning of ‘not knowing’, but, with one exception, only reduced vowel forms occur in contexts of the collocation’s interactive, face-saving functions. Moreover, only reduced vowel variants participate in […] signalling a speaker change.” Their analysis of I DON’T KNOW, then, revealed a distribution of phonological variants according to function: full variants are
associated with referential uses (reflecting their being composed of and processed as three lexical items); reduced variants are linked with pragmatic uses (reflecting their being stored and processed as single units).

I DON’T THINK was never used referentially in Bybee & Scheibman’s data (‘I do not hold the opinion that’) but functioned exclusively to soften disagreements or to signal reduced epistemicity. Contrary to Bybee and Scheibman’s expectations, the variable was not consistently reduced: “though the meaning of I don’t think is compositionally unanalyzable in conversation — indicating a more grammaticized unit — there is no consistent formal reduction concomitant with this functional shift […] (i.e., we find in the data both full and reduced variants of don’t in I don’t think)” (Bybee and Scheibman 1999: 588). Bybee and Scheibman’s findings suggest that the grammaticalization of I DON’T THINK is not congruent with phonetic reduction. These results contrast with those obtained for I DON’T KNOW. Yet they correspond to the more general observation whereby attrition may occur in the process of DMs’ grammaticalization but is not a necessary concomitant of it (Brinton 2006: 308).

Bybee and Scheibman’s results provide valuable insights into the formal variation of I DON’T KNOW and I DON’T THINK. Their conclusions are necessarily tentative because the analyses are based on a limited number of tokens. The limited data also precluded investigation of the variants’ sociolinguistic distributions. Nonetheless, Bybee and Scheibman’s study provides an important impetus for my larger-scale investigation into the formal variation of I DON’T KNOW and I DON’T THINK in Berwick English (henceforth BwE). Their findings raise the question of whether in the present corpus discourse function affects the variation in the formal encoding of I DON’T KNOW and I DON’T THINK as a result of the variables’ potential grammaticalization, and if so, whether it affects the occurrence of both non-localised and localised variants. Further, the literature discussed in Section 1 which demonstrates that formal discourse variation evinces orderly social heterogeneity raises the question of whether social factors are
also implicated in the formal variation of I DON’T KNOW and I DON’T THINK shown in (1) and (2) above.

This paper, then, seeks to further current insights into the formal properties of I DON’T KNOW and I DON’T THINK. To address the research questions set out above, I conduct a detailed analysis of the expressions’ pragmatic uses in the data and quantify their distribution across form, function and the broad social dimensions of age and gender. Before turning to the results of the qualitative and quantitative analyses, Section 3 provides details about the data and methods.

3. Data and methods

3.1. The data and sample

The analysis presented below is based on data recorded as part of a research project on language and identity in Berwick-upon-Tweed (Llamas et al. 2006), a town located five kilometres south of the Scottish-English border. The data were collected using the method designed for the Survey of Regional English (SuRE), i.e., semi-structured sociolinguistic interviews (Llamas 2007a). The interviews were conducted by the author (HP), a non-native speaker of English and outsider to the community. Various fieldwork techniques were utilized to mitigate potential effects of the Observer’s Paradox (Labov 1972: 209), including the conducting of interviews with self-selected pairs of speakers in informants’ homes (Labov 1972: 210).

As shown in Table 1, the speaker sample includes 36 speakers. They were located on the basis of two predetermined social criteria, age and gender. The three age groups were emically defined to reflect shared life stages (Eckert 1997). Socio-economic class has not been
included as an independent variable; the speaker sample represents a socially homogeneous group of working-class speakers.

**INSERT TABLE 1 HERE**

All speakers in the sample participated on a voluntary basis. They were informed of the aims of the project, which was broadly defined as research into local dialect words, and were aware that their speech was recorded. The audio-recordings were made with a SONY portable mini-disc recorder and a condenser microphone. The corpus, amounting to approximately 35 hours of speech, has been fully transcribed and consists of approximately 260,000 words (including filled pauses and minimal response particles but excluding the interviewer’s contributions).

3.2. *The variables and the variable contexts*

For the ensuing analysis, all tokens of present tense negative periphrastic *DO* in collocation with the predicates *know* and *think* and the first person singular pronominal subject *I*, or a zero subject that is co-referential with *I*, were extracted from the data. Some tokens of *I DON’T KNOW* and *I DON’T THINK* in the dataset were removed from the variable context due to their comparatively limited functional versatility, their limited productivity in the corpus, or their not constituting discourse variables.

As demonstrated below, the data contain unbound tokens of *I DON’T KNOW* that are syntactically unintegrated and not followed by an overt complementation (3a), and bound tokens that are followed by a complement in object position (3b).  

(3)  

a. *Oh, I dono. I’d rather just be a Berwicker.*
b. But I feel sort of intimidated wi Muslims, cos *I divn’t knaa* their religion.

*I divn’t knaa* what else I would call the rain.

*And I don’t know* if they have a Romany connection.

While unbound tokens perform a range of functions in the ideational, interpersonal and textual discourse domains (see Section 4), bound tokens of *I don’t know* function solely to signal insufficient knowledge and reduced epistemicity. Because of their comparatively limited functional versatility, these tokens’ distribution cannot easily be compared with that of unbound tokens. They were hence removed from consideration in the present paper.⁶

The data contain tokens of *I don’t think* that are followed by non-finite complements, as shown in (4).

(4)  

*a. Na. Don’t really think about the border.*

*b. I don’t think like that so much now.*

In (4), *I don’t think* is followed by non-finite phrasal complements, and the predicate *think* refers to a cognitive process. These tokens do not constitute examples of transferred negation and are not subject to the same functional versatility as tokens of *I don’t think* that denote ‘hold an opinion’ and are followed by the pro-form *so* (5a), or a finite clausal complement with ‘that’ (5b) or ‘zero’ (5c). They were therefore removed from consideration in the present paper.

(5)  

*a. I divn’t think so. They’re all just grouped together.*

*b. Now they would still be from Berwick because I don’t think that you talk about people being from Tweedmouth or Spittal.*

*c. Because certainly Alan Beith, our MP, I don’t think Ø he has much power with the Labour government in London.*

Tokens of *I don’t think* that have retroactive scope over a preceding negative proposition, as in (6), rather than a following positive polarity complement, as in (5), were not included in the
variable context either. Post-positioned tokens of I DON’T THINK have a limited functional versatility compared to pre-positioned tokens and are too negligibly instantiated in the corpus to warrant quantitative analysis.

(6) He wouldn’t call him fither to his face, I divn’t think.

Tokens of I DON’T KNOW and I DON’T THINK where DO is negated with the non-standard clitic particle -nae, yielding the form dinnae, were also too negligible to merit consideration in the quantitative analysis. Eight tokens of I DON’T KNOW and three tokens of I DON’T THINK whose form could not be unambiguously determined had to be removed from the variable context. Tokens which occur in quoted speech were also removed because it was not possible to establish whether the variant in the quote was part of the speaker’s or the quoted person’s repertoire. Finally, tokens that are followed by aborted or undecipherable talk were excluded from consideration in the analysis of pragmatic function since an unambiguous utterance interpretation was not possible.

A total of 380 tokens of I DON’T KNOW and 256 tokens of I DON’T THINK were retained in the database. In addition to speaker age and gender, these were coded for the function they perform (see Section 4) and for their variant.

3.3. The variants of I DON’T KNOW and I DON’T THINK in BwE

For the quantitative analysis, all instances of the expressions I DON’T KNOW and I DON’T THINK in the data were subjected to auditory analysis and classified in terms of their surface realization.

In the “full variants” of I DON’T KNOW (spelled I don’t know), a conspicuous morpheme boundary, mostly in the form of [ʔ], occurs between the [n] of don’t and the [n] of know. The
first vowel is usually realized as [ɔ] or [ʊ]. In the “semi-reduced variants” (spelled I dono), the morpheme boundary between the [n] of don’t and the [n] of know is either absent or constituted by a geminate nasal; the first vowel is usually produced with lip-rounding and is similar in quality to the ones used in the full variants. In the “reduced variants” (spelled I dunno), there is never a morpheme boundary between the nasals; in contrast to the “semi-reduced” variants, the first vowel is reduced to [ǝ]. In the variant orthographically represented as I divn’t knaa, negative periphrastic DO contains a KIT-vowel and some degree of friction. Also, with this variant, the lexical item know is usually, but not always, replaced with knaa [nɐ].

The negative auxiliary in “full variants” of I DON’T THINK (spelled I don’t think), is produced with a full vowel, [o] or [u], followed by the nasal [n] of the negative particle -n’t and optionally a [ʔ]. In the “reduced variants” (spelled I doØ think), which are produced with either a full or a reduced vowel, [o], [u] or [ǝ], the nasal of the negative clitic -n’t is omitted. The boundary between negative periphrastic DO and think is sometimes marked by [ʔ]. With both but particularly reduced variants, the initial fricative of think is occasionally realised as [f] or [fi], or is omitted altogether. In the variant orthographically represented as I divn’t think, negative periphrastic DO contains a KIT-vowel and some degree of friction.

I have alluded in the introduction to the different geographical spreads of these variants of I DON’T KNOW and I DON’T THINK. In use since at least the early nineteenth century (Wright 1902), the form divn’t constitutes a well-established feature of northern English. Its use (and that of knaa) has been reported in the historical and contemporary literature for the northern-most counties of England (Beal 2004; Upton et al. 1994; Wright 1902). Because of divn’t’s
limited geographical spread, the variants *I divn’t knaa* and *I divn’t think* will be referred to here as localised variants. The full and reduced variants of the variables exhibit a wider geographical dispersion. *I don’t know* and *I don’t think* are the standard variants and hence non-localisable; *dunno*, which was first attested in 1842 (OED: 2008), is cited in the OED without a geographical label; loss of the nasal in *I doØ think* can be considered as regionally unmarked because it was the erosion of the negative particle that led to the introduction of DO-support (Labov 1994).7 Because these variants are widely distributed throughout England and the English-speaking world, they will be subsumed under the label non-localised. As for *dono*, the vowel quality in the first syllable evokes an association with northern English. Nonetheless, I would argue that *I dono* can be categorised along *I don’t know* and *I dunno* as non-localised. The variant represents a stage in the reduction from non-reduced to fully reduced (loss of morpheme boundary > vowel reduction) which can be assumed to be widespread.

3.4. *Methods of data analysis*

The qualitative analysis draws heavily on the theories and methodology of conversation analysis (e.g. Hutchby and Wooffitt 1998; Ten Have 1999). In the allocation of functions, systematic attention is paid to the sequential context of an utterance, the temporal development of the interaction, as well as the ambient linguistic context, including other DMs, filled and unfilled pauses, and the prosodic encodings of the variables. These features are interactionally significant and serve to disambiguate the variables’ functions in a given context (Aijmer 2002: 30; Du Bois et al. 1993: 49–73; Heritage and Atkinson 1984: 5).8 Also, utterance interpretation is guided by next speakers’ treatment of preceding talk (Heritage and Atkinson 1984: 9; Sacks et al. 1974: 728–729). Following Brinton (1996), I broadly
distinguish between the ideational, interpersonal and textual domains of discourse. Unlike Holmes (1984b, 1995), I include multifunctionality as a parameter in the analysis.

Quantitative methods from sociolinguistic variation theory (Labov 1972) are employed to establish the contribution of function, age and gender to the occurrence of non-localised and localised discourse variants in the data. A distributional analysis gives the frequency of all variants by all factors. Where token numbers permit it, these patterns are subjected to the Pearson chi-square test to establish their statistical significance.

4. The discourse functions of I DON’T KNOW and I DON’T THINK

A pragmatic function was allocated to all tokens of I DON’T KNOW and I DON’T THINK in the data. As pointed out above, which function the variables perform on any one occasion is dependent on their prosodic encoding, their sequential placement in the interaction and the surrounding linguistic material. Depending upon their relevance for the ensuing quantitative analysis and discussion, some functions will be discussed in more detail than others. For an in-depth description of the variables’ functional repertoires, readers are referred to Pichler (2008).

As pointed out above, the variables operate in the interpersonal and textual domains of discourse. The former is concerned with the expression of speakers’ attitudes towards their propositions and their interlocutors (Brinton 1996: 38). In this domain, I DON’T KNOW and I DON’T THINK broadly speaking function to qualify speakers’ commitment towards their propositions and to mitigate face-threatening acts, i.e., acts that run contrary to addressees’ and speakers’ negative face wants (their desire to be unimpeded by others) and positive face wants (their desire for approval) (Brown and Levinson 1987; Fraser 1980).
In extracts (7a) and (7b) from the data, the variables serve to signal the speakers’ uncertainty regarding the validity of their propositions. As shown in the examples, the DMs can occur after or before the propositions over which they have scope. The disclaimer allows speakers to withdraw from their propositions if challenged, and thus serves to protect their face.

(7)  

a. Keith: They were I-Irish. They were [probably] from up the road. They were probably HP: [yeah]
   Keith: = [gipsies] or something like that. I dunno? And they were (.). They HP: [mhm] [mhm]
   Adam: [ @]
   Keith: = said “O:h. Where are from.” You know, straight away.

b. HP: What about males and females. Are there differences?
   Shannon: “I don’t think there are really.”

This effect is also evident in (8a) and (8b). The variables here serve to mitigate the unwelcome effect of potentially offensive or easily contested statements, thus working towards saving not only the speakers’ but also the hearers’ face.

(8)  

a. Luke: For the kids that are on drugs I blame the parents me. @
   HP: Why.
   Luke: I divn’t knaa? I think they’re just (?) Aye. They’re no looking after their kids properly, or they just (.). just letting them get away wi it?

b. Luke: But (.). I wouldn’t like to sound like a Cockney or [(h)] a liver- (.). like you
   HP: [Yeah.]
   Luke: = know Liverpool, like a Scouser. Divn’t think I’d like to sound like
   HP: Yeah.
   Luke: = them. @
In the exchanges in (9), the face-threat of Lori’s and Charlene’s disaligning assessments is softened by prefacing and delaying them with a DM (Pomerantz 1984). The speakers thus strive to maintain good social rapport with their interlocutors.

(9) a. ((Talking about the Isle of Skye.))

Godfrey: “Yeah,” it was it’s a really nice place.

HP: [“Yeah.”]

Lori: [I dunno,] because I got

Godfrey: == I had a great holiday up there.

HP: Yeah?

Lori: I got bored and sunburnt.

b. HP: You have your [you’ve] a telephone voice. Yeah.

Charlene: [But I]

I do- have I? I doØ think I’ve got a telephone voice.

This interational effect is also produced by I DON’T THINK when it serves to soften the abruptness of a denial or rejection, as exemplified in (10):

(10)HP: Do you think Berwick is a fashionable place to be?

@

Godfrey: <@ No::: @ I don’t think so somehow. @>

The epistemic and mitigating effect of I DON’T KNOW and I DON’T THINK in (9) to (11) is produced by the prosodic encoding of the DMs and the utterances in which they occur. Fall-rise or rising intonation, high pitch and reduced volume are in themselves markers of doubt and uncertainty (Brown and Levinson 1987: 172; Cruttenden 1997). Also note that when these DMs function to soften or mitigate speech acts, they frequently occur with other
linguistic items signalling reduced epistemicity: something like that (7a), I think (8a) or would (8b).

While the literature describes I DON’T THINK solely in terms of tentativeness and mitigation, as shown in (7) to (10), the BwE corpus also contains tokens of the variable that convey personal conviction. Immediately prior to the turn in (11), Matthew’s view that it is possible to recognize the Berwick accent on the television had been challenged. While the conditional clause slightly qualifies his earlier view, I DON’T THINK serves to emphasize Matthew’s commitment to his earlier position. Stress on don’t, tonic stress on the negative polarity items in the complement (any, at all), the utterance-final descending intonation contour associated with definiteness (Cruttenden 1997: 91), and the emphasis on “any difficulty at all” through reduced tempo support the interpretation of this token as a boosting device.

(11)Matthew: If they were speaking broad enough, I don’t think there’s >any difficulty at all<.

Whether I DON’T THINK functions as a hedge or as a booster is determined primarily by prosodic features and the surrounding linguistic context.10

Another usage of I DON’T THINK not discussed in the literature is its potential to affect interactional alignment. In response to Jane’s negatively framed proposition in (12), Helen’s “I divn’t think so” provides an assertive confirmation of Jane’s preceding assessment. The assertive tone is created through marked descending intonation indicative of definiteness (Cruttenden 1997: 91) and the lower than usual pitch typical of assertiveness (Holmes 1984a).

In the textual domain, which is concerned with the structural organization of discourse, DMs typically function to mark boundaries, to initiate or prevent turn-exchange, and to signal repair (Brinton 1996: 38–39). The turn-exchange uses of I DON’T KNOW (turn-taking, -holding and -yielding) are illustrated in (13).

(13)a. Luke:  I despise <@ Geordies. @>

        HP:       Why. Wh-what is [wrong with them.]

        Luke:   [I dunno:::] every time I go down to Newcastle th- all the

Geordies are always really cheeky.

b. Keith: <I’d only been away once before.> >And that was< with her. [And] eh

        HP:       mhm              [mhm]

        Keith:   = (.) I dunno (.). just like obviously I wanted to go Australia and that.

c. Leah:  (.h) >Has no effect on Berwick cos you never, < (..) I dunno?

As demonstrated in (14), when speakers abort and recast their utterances, I DON’T KNOW is sometimes inserted in discourse to bridge the gap between the repaired and repairing segments of a turn.

(14)Daniel:  No, well, (.). I s- I dunno, I keep saying, (..) if if it’s so good where you came

        from, why don’t you go back?

In addition, I DON’T KNOW is used by speakers to affect the topical development of the interview. Gabriel’s dunno in (15) serves to decline HP’s topic proffer by disavowing access to it (Pomerantz 1984: 57–58). The closure-implicative effect of dunno is achieved by the DM’s levelled prosodic delivery, which signals finality and definiteness (Cruttenden 1997). The topic-declining effect is attested in HP’s reaction: instead of probing Gabriel’s minimal response, she acknowledges the complexity of her question and moves on almost immediately to the next item on the interview agenda.
Would you say that younger people, older people use more non-standard grammar than younger ones?

Yeah.

Yeah?
Why do you think.

Dunno

It’s hard to say.

Mhm.

== Yeah. (..) Ok. (h) And do you think there’s a difference between girls and boys, or males and females?

Some tokens of *I DON’T KNOW* operate concurrently in the interpersonal and textual discourse domains. In the extracts in (16), *I DON’T KNOW* functions as a repair and turn-yielding device while at the same time qualifying the speakers’ forthcoming or preceding propositions. To avoid a subjective interpretation of these tokens as either interpersonal or textual tokens, which would ignore their intrinsic multifunctionality, these and similar uses of the variable were categorized as interpersonal-textual tokens.

But with like a proper Geordie from Newcastle they talk really, (h) like like a proper Geordie, they’re (. ) mo:re *I dunno*? Their words are more (. ) shorter.

Frae like a Berwick person.

What accent would you say you had and do you like it?

Em. It’s a mixture of probably Scottish and Geordie. But °*I dunno*°.

While at all times conveying a subjective epistemic stance, *I DON’T THINK* can in certain sequential environments simultaneously perform textual functions. These multifunctional uses have not previously been identified. Because details of these functions are not relevant to the results discussed in this paper, it is sufficient here to note that the subjectivity inherent in *I DON’T THINK* can be exploited by interlocutors to affect topic-closures or -shifts and to mark
boundaries in discourse (for details, see Pichler 2008). In (17), for example, the tentativeness expressed by I don’t think so serves to implement effectual topic closure.

(17)((After a lengthy discussion of dialect words for ‘beard’.)

    HP: Any others?

    Godfrey: I don’t think so. I think we kind of exhausted that one.

Depending on its sequential positioning, I DON’T THINK can serve to indicate transitions within larger discourse units in addition to signalling a subjective epistemic stance. It can mark the starting point of an elaboration or the transition to a conclusion that is inferred from a preceding subjective validation. In (18), the DM marks the boundary between the provision of factual information and the speaker’s subjective opinion about the issue discussed.

(18)Theodore: They [wanted] Home Rule for [Berwick], for Scotland, Berwick for [Scotland].

    HP [mhm] [mhm] [mhm]

    Theodore: = They was trying to get it into Scotland for donkeys’ years. (..)

    Guy: mhm

    HP: mhm

    Theodore: = But I doØ think it’ll ever happen like, you know.

In addition, the variable serves to structure longer opinion sequences and enumerations of pro- and contra-arguments. This is shown in (19), where Gabriel provides his opinion on the then England football manager Sven Göran Eriksson.

(19)Gabriel: I think he’s took the team as far as he can go. I think (..) he: picks his his

    HP: yeah

    Gabriel: = favourites. I doØ think he’s one of the managers that says, “Right, you’re not playing well. In bri- (..) we’re gonna bring in youngsters.”
Finally, I DON’T KNOW and I DON’T THINK have the potential to function in the ideational domain of discourse, which is concerned with the communication of propositional content (Brinton 1996: 38). The data contain no tokens of I DON’T THINK that signal its referential meaning ‘I do not hold the opinion’. Tokens of I DON’T KNOW that communicate its referential meaning of insufficient knowledge, do occur, however, and are illustrated in (20).

(20)Jane: Well, I was a telephonist for years and a lot of people thought, you know frae further down the country thought I was Welsh.
   HP: Why.
   Jane: I dunno? We divn’t knaa the connection?

Referential tokens of I DON’T KNOW differ from pragmatic ones in that they usually carry primary stress on don’t or know and are not generally accompanied by perceptible prosodic variation or by other DMs. Unlike pragmatic tokens, referential tokens of I DON’T KNOW are always followed by an attempt to account for the lack of knowledge but never by the provision of the requested assessment. Referential tokens are included in the quantitative analysis to test whether they favour different variants than pragmatic tokens.

Having established the functions performed by each variable in the interview data, I present the overall distribution of variants before turning to a discussion of the variants’ distribution in terms of function, age and gender.

5. Overall distribution of variants

Table 2 shows the overall distribution of variants of I DON’T KNOW in the data. I dunno is the dominant variant, occurring more often than all the other variants combined. I divn’t knaa is more frequent than I don’t know. I dono is only negligibly attested in the data. It will therefore not be subjected to statistical tests and will feature only marginally in the discussion of the
results. According to Bybee and Scheibman (1999), the phonetic reduction of I DON’T KNOW is a result the variable’s grammaticalization: through frequent use, the expression’s components have come to be stored and processed as a single prefabricated phrase, leading to pragmatic strengthening and phonetic attrition (see Section 2). I consider below whether the predominance of I dunno in the BwE data is indicative of the variable’s grammaticalized status as a DM and how I divn’t knaa fits into the overall picture.

**INSERT TABLE 2 HERE**

Table 3 shows the overall distribution of variants of I DON’T THINK in the present dataset. Roughly half of all tokens of the variable in the data are realized with the full variant, I don’t think. The reduced variant, I doØ think, constitutes slightly less than one third of all instances of the variable. As pointed out above, all tokens of the variable in the data perform pragmatic functions. The distribution of variants suggests that unlike the case of I DON’T KNOW, the grammaticalization of I DON’T THINK is not accompanied by consistent phonetic reduction. This finding dovetails with that obtained by Bybee & Scheibman (1999) in their analysis of I DON’T THINK in American English (see Section 2). The localised variant I divn’t think is less frequent than either of the non-localised variants.

**INSERT TABLE 3 HERE**

The following sections seek to establish whether function and the social factors age and gender are relevant to the distribution of the variants of I DON’T KNOW and I DON’T THINK.
6. The functional reality of non-localised and localised variants of I DON’T KNOW and I DON’T THINK

In this section which focuses on function and also in Section 7 which focuses on social factors, the distributional analysis for I DON’T KNOW proceeds in two steps. First, an analysis is conducted into the distribution of the non-localized variants. This analysis is then followed by an analysis of the localized variant I divn’t knaa. The decision to proceed in this way is necessitated by the fact that ten out of the 36 speakers in the sample do not employ the variant I divn’t knaa. These speakers are not considered in the analysis of I divn’t knaa because their inclusion in the quantification might skew the results and conceal important constraints on the occurrence of this variant (Tagliamonte 2006: 86–88). Since the differential distribution of the non-localized variants of I DON’T KNOW across speakers who do and speakers who do not use I divn’t knaa does not reach statistical significance ($\chi^2 = 5.22$, df = 2, p < 0.07) and since it can be assumed that constraints on the occurrence of non-localized variants are likely to operate across speakers who use and those who do not use the localized variant I divn’t knaa, all speakers are included in the quantification of the non-localized variants. For reasons outlined in Section 3.2, some tokens of I dunno (n = 4) and I divn’t knaa (n = 2) could not be coded for function. This explains the differential overall token numbers presented in this section compared to those presented in Sections 5 and 7.

Figure 1 gives a breakdown of the functional distribution of the non-localised variants of I DON’T KNOW. It tracks the percentage and raw scores of each of the four broad functional categories (referential, interpersonal, interpersonal-textual, textual) across the three non-localized variants (I don’t know, I dono, I dunno). (The figures for the localized variant I divn’t knaa are not shown in Figure 1. I divn’t knaa was, however, included as a non-application value in the tabulations, which explains why the percentages for each of the
The patterning of variants reported in Bybee and Scheibman (1999) is clearly visible. The reduced variant *I dunno* exhibits the greatest propensity to occur in all pragmatic categories (interpersonal, interpersonal-textual, textual), whereas the occurrence of the full variant *I don't know* is strongly correlated with referential uses. The occurrence of the semi-reduced variant *I dono* does not seem to be conditioned by function, which might reflect its intermediate status between full and reduced variants. The results suggest that the distribution of *I don't know* and *I dunno* is functionally conditioned as a result of the variable’s grammaticalization.

**INSERT FIGURE 1 HERE**

Figure 2 examines the functional distribution of *I divn’t knaa* amongst the 20 *I divn’t knaa*-users in the sample. It indicates the percentage of all tokens of the variable in each of the four functional categories that were realized as *I divn’t knaa*. *I divn’t knaa* does not occupy a functional niche in the dataset. It appears with roughly equal frequencies (33-37%) across all functional categories. While the more frequent non-localised variants carry a functional load, *I divn't knaa* seems to be a functional “wild card.”

**INSERT FIGURE 2 HERE**

As regards *I DON’T THINK*, it has been pointed out above that only ten out of the 36 speakers in the sample use the localised variant *I divn’t think*. In order to avoid skewing the results, speakers who do not use *I divn’t think* are therefore not considered in the analysis of this variant. Because speakers who employ the variant *I divn’t think* (henceforth referred to as
divn’t think-users) have significantly different rates of the non-localized variants, *I don’t think* and *I doØ think*, than those speakers who do not use the variant *I divn’t think* (henceforth referred to as non-*divn’t think*-users) ($\chi^2 = 4.23, \text{df} = 1, p = 0.04$), Figure 3 below and Table 3 in Section 7 present the distribution of variants separately for non-*divn’t think*-users and *divn’t think*-users. Since some tokens of *I don’t think* (n = 14), *I doØ think* (n = 4) and *I divn’t think* (n = 3) could not be coded for function (cf. Section 3.2), the overall token numbers presented in the figures in this section differ from those presented in Sections 5 and 7.

Figure 3 investigates the functional distribution of the variants of *I DON’T THINK* in non-*divn’t think*-users and in *divn’t think*-users. The figure tracks the percentages of both functional categories (interpersonal, interpersonal-textual) across the variants (*I don’t think*, *I doØ think*, *I divn’t think*) within the two user groups. The variants display no marked function-specific patterning amongst speakers who do not employ the variant *I divn’t think* nor amongst those who do. Figure 3, however, obscures more fine-grained form-function correlations evident in the data. These are illustrated in Figure 4, where the tokens of the non-localized variants produced by non-*divn’t think*- and *divn’t think*-users are combined on the basis of exhibiting similarities in their form-function distribution. The functions surrounded by a border are to be noted. They demonstrate a correspondence of reduced variants with disalignments and of full variants with alignments and boosters. Yaeger-Dror (1985, 1997) argues that in interactional registers a Cognitive Prominence Principle (henceforth CPP) predicts prominence of negatives that carry important focal information, while a Social Agreement Principle (henceforth SAP) predicts prominence of negatives that show support of co-conversationalists, and non-prominence of negatives that are non-supportive of co-conversationalists. Figure 4 suggests that in some of its functions, the phonetic realization of *I DON’T THINK* is affected by the CPP and SAP: the negative particle in boosters is always fully enunciated because it is the bearer of important focal information, i.e., that something is not
the case; the face-threat of I DON’T THINK in disaligning turns is mitigated by means of attrition of the negative clitic; conversely, the clitic in affiliation markers tends to be fully enunciated because the negative is supportive of interlocutors. My suggestion that the variation between full and reduced variants of I DON’T THINK is conditioned by Yaeger-Dror’s CPP and SAP is necessarily tentative due to restricted data. Also, the operation of the CPP and SAP fails to explain the predominance of full over reduced variants in the vast majority of other functions. Nonetheless, it seems that non-localised variants of I DON’T THINK display some function-specific patterning on the micro-level of analysis.

INSERT FIGURE 3 HERE

INSERT FIGURE 4 HERE

Figure 5 shows the percentage of tokens realized as I don’t think in each of the functions displayed. The functional distribution of the localized variant I divn’t think remains rather inconclusive due to severely limited data. It appears, however, that the use of I divn’t think is not associated with specific functions.

INSERT FIGURE 5 HERE

The results presented in this Section suggest that the distribution of frequently occurring non-localised variants, I don’t know, I dunno, I don’t think and I doØ think, is, at least to a degree, functionally conditioned. The localised variants, I divn’t knaa and I divn’t think, by contrast,
do not seem to carry a functional load. In order to fully grasp the meaning of non-localised and localised discourse variants, however, it is not sufficient to look at their functional distribution. The impact of social factors also needs to be considered.

7. The social reality of non-localised and localised variants of I DON’T KNOW and I DON’T THINK

Figure 6 examines the sociolinguistic distribution of the non-localized variants of I DON’T KNOW with respect to age and gender. The figure shows the percentage of each of the three non-localized variants of I DON’T KNOW (I don’t know, I dono, I dunno) in each of the six social cohorts (old female, middle female, young female; old male, middle male, young male). (The figures for the localized variant I divn’t knaa are not shown in Figure 6. The localized variant was, however, included as a non-application value in the tabulations, which explains why the percentages of variants in each cohort do not add up to 100%). In all social cohorts, except old males, I dunno is the most frequent variant. Older males’ lower than average rate of I dunno is compensated by their significantly higher use of I don’t know. Cross-tabulations with function reveal that older males’ noticeably higher rate of I don’t know is due to their frequent use of this variant in pragmatic uses. In contrast to other social cohorts, older males do not fully participate in the functional split between full and reduced variants pointed out above. The overall higher incidence of I don’t know and I dunno amongst women are due to higher rates of alternative variants, i.e., I dono and I divn’t knaa, amongst men from the middle and young age-groups (cf. Figures 6 and 7).

INSERT FIGURE 6 HERE
The social distribution of *I divn’t knaa* amongst the 20 *divn’t knaa*-users in the sample is displayed in Figure 7. It shows the percentage of tokens realized as *I divn’t knaa* in each of the six social cohorts. The distribution reveals that social factors are strongly implicated in the occurrence of *I divn’t knaa*. Young males show by far the greatest propensity to use the localized variant. This distribution reflects a pattern commonly reported for phonological variables, whereby men have higher rates than women of localized variants (Foulkes and Docherty 1999; Milroy et al. 1994; Watt and Milroy 1999).

**INSERT FIGURE 7 HERE**

Finally, Table 4 compares the sociolinguistic distribution of the variants of *I DON’T THINK* across non-*divn’t think*- and *divn’t think*-users. As pointed out in Section 6, the differential distribution of the non-localised variants *I don’t think* and *I doØ think* across non-*divn’t think*- and *divn’t think*-users is statistically significant. The markedly lower rate of *I doØ think* amongst the *divn’t think*-speakers seems to be due not only to the effect of *I divn’t think*, but also due to male *divn’t think*-users’ frequent use of *I don’t think*. When the figures for both groups of speakers (non-*divn’t think*- and *divn’t think*-users) are combined, the occurrence of *I doØ think* appears not to be affected by social factors. As regards *I don’t think*, however, male speakers show significantly higher rates of this variant than females. This is due to women’s marked predominance in the use of *I divn’t think*, rather than to women’s leading the variable’s phonetic reduction. This pattern contrasts with that found for *I divn’t knaa*, where it was young males who, in line with the trends reported in the literature for phonological variables, displayed the highest propensity to use the localised variant. Because the patterns for *I divn’t think* are based on only ten speakers, they cannot be meaningfully interpreted in
terms of age. Nonetheless, the data are strongly suggestive of social factors making an important contribution to the occurrence of the localised variant *I divn’t think*.

**INSERT TABLE 4 HERE**

The results suggest that while the occurrence of non-localised variants of *I DON’T KNOW* and *I DON’T THINK* is only minimally conditioned by social parameters, the use of localised variants is socially diagnostic.

**8. Discussion**

The results presented in Sections 6 and 7 demonstrate that discourse variables conform to the fundamental tenet of sociolinguistics: that linguistic variation is not random but systematically constrained by contextual factors (Weinreich et al. 1968: 99–100). An important new insight gained by the present analysis is that the occurrence of non-localised and localised discourse variants is conditioned by different parameters. Unlike localised variants, non-localised variants of *I DON’T KNOW* and *I DON’T THINK* show relatively little sensitivity to the social factors age and gender. Close scrutiny of the data revealed that the absence of social conditioning also pertains to the functional compartmentalisation of these variants: men and women and different age groups do not use the non-localised variants of *I DON’T KNOW* and *I DON’T THINK* for significantly different pragmatic functions. This further highlights the relative lack of social meaning attached to non-localised variants.

The results show that full and reduced variants of *I DON’T KNOW* and *I DON’T THINK* display at least some function-specific patterning. With *I DON’T KNOW*, the pronounced functional split between full and reduced variants is indicative of the variable’s grammaticalized status. In
addition to divergence, the distribution of I DON’T KNOW exhibits the following diagnostics of grammaticalization (Hopper 1991): persistence, layering and decategorialization. The fact that pragmatic tokens are virtually never (0.8%) modified through intervening adverbs (e.g. really, even) highlights the variable’s formulaic nature as a single processing unit.

In Pichler (2008), I propose that pragmatic uses of I DON’T KNOW developed from a matrix-complement construction in which the variable communicates speakers’ lack of knowledge vis-à-vis the proposition contained in the finite complement. In the process of grammaticalization, the matrix clause I DON’T KNOW came to be used as an epistemic stance formula towards the following erstwhile complement. The bonds between I DON’T KNOW and the finite clause over which it had scope weakened, leading in some contexts to the expression’s reanalysis as an independent element with scope over preceding or following propositions. Consequently, I DON’T KNOW gained positional mobility, widened its scope over discourse and acquired additional functions. In the present dataset, the variable occurs virtually categorically outside the main constituents of a clause, occasionally constituting the sole component of a turn (see examples in Section 4). Only one token occurs between obligatory constituents of a clause, with local scope over the following lexical items:

(21) Rebecca: They just (...) I dunno? >have more s:ense< of what’s going on (. ) really, if they were in the North-East.

The production of this token by the youngest informant in the sample leads me to hypothesize that the positional mobility of I DON’T KNOW might still be increasing and that the variable might continue to develop new discourse functions.

With I DON’T THINK, the variation between full and reduced variants reflects subtle pragmatic differences that tally with Yaeger-Dror’s (1985, 1997) Cognitive Prominence and Social Agreement Principles. The exploitation of full and reduced variants for micro-level functional
distinctions might to some degree account for the lack of consistent reduction in the realization of the DM I DON’T THINK. Another factor that prevents consistent attrition in I DON’T THINK is pitch prominence on don’t in some non-boosting contexts. Raised pitch on negative DO functions to signal high degrees of doubt. It could therefore be argued that the signalling of fine-grained nuances of meaning through prosody affects the linguistic variation in the data. A more thorough investigation of I DON’T THINK’s prosodic realization and of the degrees of uncertainty conveyed by the variable than was possible within the constraints of this paper is required to confirm this hypothesis. Finally, different changes associated with grammaticalization do not always occur simultaneously (Cheshire 2007: 166, 173). The meaning generalization of I DON’T THINK might be predating its phonetic reshaping.

Despite the lack of persistent phonetic reduction, I endorse the view expressed elsewhere that I DON’T THINK is a highly routinized discourse formula. Pragmatic uses of pre-positioned I DON’T THINK derived through semantic bleaching and conventionalization of implicatures from its propositional reading ‘I do not hold the opinion that’. In other words, the pragmatic reading of I DON’T THINK is a grammaticalized reading of negative raising (Horn 1978). The high rate (99%) in the data of ‘zero’-complementizers after I DON’T THINK is a strong concomitant of its syntactic and informational reanalysis from a complement-taking referential matrix clause to an epistemic stance marker. The variable’s formulaic status is also reflected in its decategorialization and ossification: it is never modified by an adverb; it is frozen in the first person singular; and it is virtually categorically formed with think (in the present data, only 1% of collocations of negative DO with I and an opinion predicate are formed with opinion predicates other than think). The prevalence in the data of the construction ‘I don’t think p’ over ‘I think not-p’ (ratio 21:1) is further testimony to the fact that the modification of negative propositions is highly routinized.
In the literature, sociolinguistic variation in discourse is generally described in terms of “discourse styles” (Macaulay 2005) and “conversational practices” (Coates 2004: 85–110), i.e., differential frequencies and differential functional uses of DMs across social groups. Cross-tabulations showed no significant social variation in the pragmatic usage of the variables discussed here. However, as displayed in Table 4, social differences are evident in the variables’ overall frequencies. Older speakers have markedly lower rates of the DMs I DON’T KNOW and I DON’T THINK than speakers from the middle and particularly young age groups. The frequency of I DON’T THINK is also affected by gender. Importantly, the analysis has shown that sociolinguistic variation in discourse extends to the surface realization of DMs. In contrast to the non-localised variants, the localised variants I divn’t knaa and I divn’t think do not carry a functional load but they have social meaning: the use of I divn’t knaa is strongly associated with young males; the use of I divn’t think is strongly associated with females. These distributions highlight that the orderly social heterogeneity evident in phonological, morpho-syntactic and lexical variation also exists on the discourse level. It is not only the overall frequency of discourse variables or the choice of one discourse variable over another that is socially diagnostic but also the choice of discourse variant.

**INSERT TABLE 4 HERE**

The results presented in Sections 6 and 7 further reveal that the same variant of negative auxiliary DO does not carry the same functional or social meaning across different discourse variables. With the expression I DON’T KNOW, phonetically full and reduced variants of don’t are strongly associated with referential and pragmatic uses respectively; with the expression I DON’T THINK, on the other hand, they reflect fine-grained pragmatic differences on the micro-
level of analysis. The differential meanings of full and reduced variants of don’t across the two DMs can be attributed to different outcomes of the processes involved in grammaticalization. The differential social distribution of don’t in I DON’T KNOW and I DON’T THINK can be attributed to differential discourse styles and differential conversational practices, which cause male and female speakers to employ the two variables with differential frequencies. Due to these interactional constraints (localised) discourse variants ought not to be expected to correspond fully to the patterns of distribution of (localised) phonological variants.

It has been argued that localised phonological variants can be interpreted as indexations of local identity (Llamas 2007b; Watt 2002). I would like to endorse Cheshire’s (2007) proposition that discourse variants might equally play a role in speakers’ linguistic performance of their social identities (Coupland 2001; Eckert 2000). I don’t knaa and I don’t think, which are strongly associated with northern varieties of English (see Section 3.3), might be performing a dual role in discourse: they fulfil a range of interpersonal and textual functions, while at the same time serving to index speakers’ local and/or regional identities. By contrast, exclusive use of the non-localised variants of the DMs I DON’T KNOW and I DON’T THINK might be signalling speakers’ orientations towards a pan-British or global culture. This might potentially explain why the use of I don’t knaa and I don’t think is restricted to a subsample of speakers in the data.

9. Conclusion

This paper has investigated the formal variation of the expressions I DON’T KNOW and I DON’T THINK. The analysis revealed that localised and non-localised variants are conditioned by different parameters. The former carry a functional load, the latter have social meaning. This
study thus makes an important contribution to the current state of knowledge about discourse variation. It demonstrates that discourse variation extends to the formal level of language use and that formal variation in discourse is functionally and socially meaningful. These findings could only be produced by combining qualitative with quantitative methods of data analysis. They therefore support the case for an integration of qualitative approaches in the study of language variation (Cheshire et al. 2005; Schneider and Barron 2008). Macaulay (2002a, 2005) has repeatedly criticised the failure of sociolinguistic studies that focus on phonological and morphological variation to fully investigate “what Milroy (1979: 91) calls ‘the sociolinguistic complexity’ of the communities” (2005: 4–5). My results argue that discourse variables deserve a more prominent place on the sociolinguistic research agenda than they have hitherto been accorded.

Appendix

Transcription conventions

[   ] overlap

== latching

= continuation of speaker turn

- truncated words

(h), (.h) inbreath, outbreath

@ laughter

<@  @> produced with laughter

< > increased tempo
> < reduced tempo

___ emphatic stress

○ ○ soft speech

superscript font higher than usual pitch

subscript font lower than usual pitch

: syllable lengthening

(., (..), (…) short, medium and long pause

. final intonation contour

, continuing intonation contour

? rising intonation contour

(?) undecipherable words

((  )) extralinguistic information

_________________________

Notes
1 I gratefully acknowledge the generous support of the Carnegie Trust for the Universities of Scotland which enabled me to carry out the necessary fieldwork. I would like to thank Stephen Levey and an anonymous reviewer for valuable comments on earlier drafts of this paper. Thanks also to Carmen Llamas and Dominic Watt for their input on the phonetic categorisation of the variables and Mercedes Durham for her advice on the statistical analysis. All remaining errors are, of course, my own.

2 My definition of discourse markers is broader than that provided, for example, by Fraser (1998). Fraser draws a distinction between “discourse markers,” which perform a structural role, and “pragmatic particles,” which express speaker attitudes. In my view, this distinction is not tenable because many discourse features perform both functions, sometimes simultaneously (see Section 4).
Small capitals are used throughout this paper to refer to discourse variables. Lower case italics are used to refer to their phonetic or grammatical variants.

All informant names in this paper are pseudonyms. HP is the interviewer.

Thompson (2002) argues that utterances with I DON’T KNOW and I DON’T THINK, as in (3b) and (5), should not be analyzed in terms of matrix and complement clauses but as monoclusal constructions with an epistemic frame. While I fully endorse this view, I employ the term “complement” here for lack of a better term.

Suffice here to say that the non-localised variants of the 168 bound tokens of I DON’T KNOW in the data display similar patterns of distribution across function as unbound tokens: phonetically full variants are favoured for referential uses; reduced variants are favoured for pragmatic, i.e., epistemic, uses. See further Pichler (2008).

This view contradicts that expressed by Petyt (1983), who argues that loss of the nasal in don’t, a phenomenon he calls “secondary contraction,” is regionally restricted within Britain.

Fleischman (1999) maintains that co-occurring DMs each fulfil a different pragmatic function. I share the view expressed elsewhere (Stubbe and Holmes 1995; Aijmer 2002; Macaulay 2005) that DMs do not function in isolation from their linguistic context but that they interact with and mutually reinforce each other.

Although Brown and Levinson (1987) and Cruttenden (1997) deal with southern British English and American English, which differ in their intonation from BWE, the intonation patterns I observed were broadly as expected from those studies, given the discourse functions of the tokens I analysed.

My category of “boosters” differs from what is in the literature referred to as I DON’T THINK’s propositional meaning (Nuyts 1990: 572; Simon-Vandenbergen 1998: 313–315). In (11), I DON’T THINK does not serve to deny the suggestion that Matthew holds the opinion expressed in the complement (‘It is not the case that I think p’) but to forcefully deny the proposition expressed in the complement (‘p is not the case’). My epistemic reading of I DON’T THINK in (11) is supported by the fact that the DM can be omitted from the utterance without changing its propositional content (‘There’s no difficulty in spotting them at all.’) or can be paraphrased with an epistemic adverbial (‘There’s absolutely no difficulty in spotting them at all.’).
References


Moore, Emma & Robert Podesva. 2008. ‘We do say in’ he, don’t we?’ Tag questions, social meaning, and style. Paper presented at Sociolinguistics Symposium 17, Amsterdam, 4 April.


Weatherall, Ann. 2007. ‘I don’t know, I really don’t know’: Indexing knowledge (un)certainty in talk in interaction. Poster presented at the 10th International Pragmatics Conference, Gothenburg, 9 July.


Tables and Figures

Table 1. Speaker sample

<table>
<thead>
<tr>
<th></th>
<th>young (17-23)</th>
<th>middle (27-48)</th>
<th>old (60-81)</th>
</tr>
</thead>
<tbody>
<tr>
<td>male</td>
<td>6</td>
<td>6</td>
<td>6</td>
</tr>
<tr>
<td>female</td>
<td>6</td>
<td>6</td>
<td>6</td>
</tr>
</tbody>
</table>

Table 2. Overall distribution of variants for I DON’T KNOW

<table>
<thead>
<tr>
<th>VARIANT</th>
<th>%</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>dunnno</td>
<td>57</td>
<td>217</td>
</tr>
<tr>
<td>divn’t knaa</td>
<td>21</td>
<td>81</td>
</tr>
<tr>
<td>don’t know</td>
<td>15</td>
<td>58</td>
</tr>
<tr>
<td>dono</td>
<td>7</td>
<td>24</td>
</tr>
<tr>
<td>TOTAL N</td>
<td></td>
<td>380</td>
</tr>
</tbody>
</table>

Table 3. Overall distribution of variants for I DON’T THINK

<table>
<thead>
<tr>
<th>VARIANT</th>
<th>%</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>don’t think</td>
<td>52</td>
<td>133</td>
</tr>
<tr>
<td>do Ø think</td>
<td>31</td>
<td>80</td>
</tr>
<tr>
<td>divn’t think</td>
<td>17</td>
<td>43</td>
</tr>
<tr>
<td>TOTAL N</td>
<td></td>
<td>256</td>
</tr>
</tbody>
</table>
Figure 1. Functional distribution of non-localised variants of I DON’T KNOW (with I don’t knaa included as a non-application value) (The differential distribution of I don’t know and I dunno across referential uses and all pragmatic uses combined is highly significant, $\chi^2 = 113.34$, df = 1, p < 0.001.)

Figure 2. Functional distribution of the localised variant I divn’t knaa
Figure 3. Functional distribution of variants across non-d’ve think- and d’ve think-users (The differential functional distribution of variants amongst d’ve think-users is not significant, $\chi^2 = 1.03$, df = 2, p = 0.60.)

Figure 4. Distribution of non-localised variants across individual functions
Figure 5. Distribution of *I divn’t think* across individual functions

Figure 6. Social distribution of non-localised variants of *I DON’T KNOW* (with *I divn’t knaa* included as a non-application value) (Differences in the distribution of *I don’t know* and *I dunno* across age are highly significant amongst males, $\chi^2 = 39.35$, df = 2, $p < 0.001$, but not amongst females, $\chi^2 = 1.69$, df = 2, $p = 0.43$. When all age groups are combined, the differences in the distribution of *I don’t know* and *I dunno* across gender are not significant, $\chi^2 = 3.43$, df = 1, $p = 0.06$.)
Figure 7. Social distribution of the localised variant *I divn’t knaa*

Table 3. Social distribution of variants of *I DON’T THINK* across non-*divn’t think-* and *divn’t think-*users (When the tokens of non-*divn’t think-* and *divn’t think-*users are combined, the differential social distributions of non-localised variants are statistically significant due to middle males’ infrequent use of *I doØ think*, $\chi^2 = 15.94$, df = 5, p < 0.01.)

<table>
<thead>
<tr>
<th></th>
<th>non-*divn’t think-*users</th>
<th>divn’t think-users</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td><em>don’t think</em></td>
<td><em>doØ think</em></td>
</tr>
<tr>
<td>female</td>
<td></td>
<td></td>
</tr>
<tr>
<td>old</td>
<td>75 6 25 2</td>
<td></td>
</tr>
<tr>
<td>middle</td>
<td>53 16 47 14</td>
<td></td>
</tr>
<tr>
<td>young</td>
<td>65 26 35 14</td>
<td></td>
</tr>
<tr>
<td>male</td>
<td></td>
<td></td>
</tr>
<tr>
<td>old</td>
<td>35 7 65 13</td>
<td></td>
</tr>
<tr>
<td>middle</td>
<td>76 25 24 8</td>
<td></td>
</tr>
<tr>
<td>young</td>
<td>41 9 59 13</td>
<td></td>
</tr>
<tr>
<td><strong>TOTAL</strong></td>
<td>80 64</td>
<td>44 16</td>
</tr>
</tbody>
</table>
Table 4. Frequency in 10,000 words of the DMs I DON’T KNOW and I DON’T THINK across age and gender

<table>
<thead>
<tr>
<th></th>
<th>I DON’T KNOW</th>
<th>I DON’T THINK</th>
</tr>
</thead>
<tbody>
<tr>
<td>old</td>
<td>1.7</td>
<td>7</td>
</tr>
<tr>
<td>middle</td>
<td>4.1</td>
<td>12.6</td>
</tr>
<tr>
<td>young</td>
<td>26.7</td>
<td>12.7</td>
</tr>
<tr>
<td>male</td>
<td>12.4</td>
<td>8.3</td>
</tr>
<tr>
<td>female</td>
<td>12.2</td>
<td>13.7</td>
</tr>
</tbody>
</table>