FORM-FUNCTION RELATIONS IN DISCOURSE: THE CASE OF 'I DON'T KNOW'*

HEIKE PICHLER

Abstract

In this paper qualitative methods from conversation analysis and quantitative methods from sociolinguistics are combined to describe the expression I DON'T KNOW in terms of its functions and to determine to what extent variation in its surface form correlates with pragmatic function and the extra-linguistic variables of age and gender. The qualitative analysis reveals that I DON'T KNOW performs a range of functions on the subjective and textual levels of discourse. These are briefly described and exemplified. The quantitative analysis indicates that the variants I don't know and I dunno are functionally divergent: I dunno, the most frequent variant in the corpus, is strongly disfavoured for the declaration of insufficient knowledge and favoured for subjective-textual uses; the variant I don't know, on the other hand, is strongly favoured for referential and subjective uses and disfavoured in textual contexts. The increase in the use of I divn't knaa amongst a subsample of young male speakers is accounted for by an extension in the use of this variant amongst these speakers.

1. Introduction

The extracts from the data in example (1) illustrate two intrinsic features of the expression I DON'T KNOW: it can occur repeatedly within individual speaker turns and it is highly variable in its surface form, represented here by means of variation in orthography (I don't know, I dono, I dunno, I divn't knaa).

(1) a. Luke: I dunno? It's just something ab- I dunno? It's just only I think only Berwick people can tell. I divn't knaa what it is. Just just as if you h- hear
   HP: yeah
   HP: yeah uh-huh
   Luke: = Don't have a clue. It's weird. @

b. Jerry: I I dunno if if the percentage is maybe more one way than the other, I dono.

c. HP: Why not.
   Natalie: @ I don't know? I dunno? I just don't like being called a Berwicker.
   HP: @

* 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 the audience at the 'First Newcastle Postgraduate Conference of Theoretical and Applied Linguistics' for their comments, particularly Isabelle Buchstaller and Karen Corrigan. I wish to thank Carmen Llamas, Mercedes Durham and an anonymous reviewer for their helpful comments on earlier drafts of this paper. I would also like to thank Dominic Watt for his help with all things phonological and Mercedes Durham for her patient assistance with Goldvarb. All remaining errors are, of course, my own.

1 A key to the transcription conventions is provided in the appendix.

2 All informant names in this paper are pseudonyms.
Recent studies suggest that variable surface realizations of syntactic and discourse features are motivated by discourse function (Ford 1993, Stenström 1998, Scheibman 2000, Tao 2001). The current paper aims to contribute to this research by conducting a qualitative and quantitative analysis of the variable I Don't Know in a corpus of interview data recorded in Berwick-upon-Tweed, a border town in the far North-East of England. The paper begins with a brief literature review in section 2. Section 3 describes the data and method. Section 4 presents a brief summary of the qualitative analysis of the expression's pragmatic functions, followed by a quantitative analysis of pragmatic and extra-linguistic constraints on the linguistic variation. Finally, section 5 discusses the findings.

2. Contemporary research on I Don't Know

The pragmatic functions of I Don't Know have received considerable research interest. What researchers examining I Don't Know share is a belief that the expression performs a variety of functions other than a declaration of insufficient knowledge or lack of certainty. Yet as regards its pragmatic usages, analysts have focused their attention on different levels of discourse.

Tsui's (1991), Potter's (2004) and Wooffitt's (2005) descriptions of I Don't Know focus on the expression's interpersonal, i.e. face-saving, functions. Potter (2004) and Wooffitt (2005) argue that I Don't Know is used by conversationalists to protect their own positive face-wants: the risk of their utterances being interpreted unfavourably is reduced by prefacing them with a disclaimer; at the same time, the disclaimer averts potential contradictions from interlocutors by denying the relevance of immediately preceding or following propositions. Tsui's (1991) analysis convincingly illustrates that I Don't Know is used by speakers not only to protect their own but also their interlocutors' face-wants. This, according to Tsui, is the case when the expression is employed to preface disagreements, to minimize impolite beliefs and to avoid making assessments, explicit disagreements, or commitments. Kärkkäinen (2003: 25) justifiably criticises Tsui's analysis for its preoccupation with the notion of face at the expense of discussing textual functions. These have been emphasized by conversation analysts and include topic-closure (Ford & Thompson 1996), topic-curtailment (Beach & Metzger 1997), turn-taking (Schegloff 1996) and turn-yielding (Östman 1981, Schegloff 1996, Scheibman 2000). The fact that Schegloff (1996: 62) refers to the turn-taking function of I Don't Know as 'prefatory epistemic disclaimer' suggests that instances of I Don't Know can function on different levels of discourse simultaneously, i.e. they can initiate a turn and at the same time hedge its content. I will return to this point below.

Scheibman's (2000) study of I Don't Know in American English conversational data discusses the relation between the expression's functions and its phonetic realizations. Scheibman analysed 38 tokens of full and reduced realizations of don't in I Don't Know in the speech of six adult Americans, aged 27 to 52, and concluded that:

All 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. (Scheibman 2000: 120)
Scheibman’s analysis thus reveals a distribution of phonological variants according to function: while reduced forms of periphrastic negative do are used in her data to express a lack of knowledge and also to perform interpersonal and textual functions, full variants are, with one exception, used only for the declaration of insufficient knowledge.

The review of the literature thus raises the following questions: (1) Which pragmatic functions does I DON’T KNOW perform in the corpus of Berwick English analysed for this paper? (2) Is the functional trend between full and reduced variants of I DON’T KNOW observed by Scheibman for American English present in British English? (3) More specifically, does the occurrence of non-standard localised variants of the expression, i.e. I divn’t knaa and I dinnae ken, correlate with function as well?

3. Data and methods
3.1. The community

The data for the present investigation were collected in Berwick-upon-Tweed, England’s northernmost town, lying only three miles (five kilometres) south of the Scottish-English border. With a population of 13,040, Berwick-upon-Tweed is the largest town on the North Sea coast between Newcastle to the south and Edinburgh to the north-west. 3

3.2. The informants

As shown in Table 1, the data are taken from a sample of 36 speakers from Berwick-upon-Tweed. The sample is equally stratified by age and includes three generations of speakers to allow for apparent time analysis. Socio-economic class has not been included as an independent variable.

<table>
<thead>
<tr>
<th>Table 1: Speaker sample</th>
<th>young (17-23)</th>
<th>middle (27-48)</th>
<th>old (60-81)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>male 6 female 6</td>
<td>male 6 female 6</td>
<td>male 6 female 6</td>
</tr>
</tbody>
</table>

3.3. The data

The data were collected using the method designed for the Survey of Regional English (SuRE), i.e., semi-structured sociolinguistic interviews 4 (see Llamas 1999 for a detailed illustration of the method). For the present analysis only one part of the interviews, namely the discussions of the Identification Questionnaire (IdQ), was analysed. The data, amounting to approximately 20 hours of speech, have been fully transcribed and constitute approximately 120,000 words.

---

3 See Watt & Ingham (2000) for a discussion of the effects of the town’s geographical/cultural location on its dialect.

4 Unlike Levinson (1983: 284) and Scheibman (2002: 17–18), who strongly argue in favour of basing the study of pragmatic features on conversation, I would argue that interview data, too, lend themselves to the analysis of such features (see, for example, Schiffrin 1987, Macaulay 2005), as long as it is acknowledged that the conclusions drawn from studying interview data might not apply elsewhere.
3.4. Variants of I DON'T KNOW in Berwick English

For the quantitative analysis, all instances of the expression I DON'T KNOW in the data have been classified into five categories. In the 'full variants' (orthographically represented as 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 [ɐ] or [Y]. In the 'semi-reduced variants' (represented in writing as 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 two [n]s; in contrast to the 'semi-reduced' variants, the first vowel is reduced to [ɪː]. In the variant orthographically represented as I divn't knaa, a form commonly associated with Newcastle/Tyneside (Beal 1993), negative periphrastic 'do' contains a KIT-vowel and some degree of friction. Also, with this variant, the lexical item 'know' is usually, but by no means always, replaced with 'knaa' [ strengthen]. Finally, with the variant represented as I dinnae ken, a form usually associated with Scotland (Macafee 1992), periphrastic 'do' is negated with the clitic negative particle '-nae'. In this variant, 'know' is always replaced with 'ken'.

3.5. Circumscribing the variable context

The data contain unbound tokens of I DON'T KNOW, i.e., stand-alone tokens that do not take a complement, as in example (2a), and bound tokens that are followed by a dependent element, as in examples (2b) and (2c).

(2) a. Gabriel: <I dunno? It's just> (. ) maybe cos you're nervous and that, it just comes out.
   b. Daniel: But I feel sort of intimidated wi Muslims, cos I divn't knaa their religion.
   c. Gabriel: But (. ) I dunno why we play in the Scottish leagues.

This paper will only discuss unbound instances of I DON'T KNOW. As illustrated in section 4.1. below, unbound I DON'T KNOWs are used by speakers to express a lack of knowledge and to perform various pragmatic functions. Tokens that fulfil a pragmatic function can be referred to as discourse markers (henceforth DMs). Unlike tokens used for the declaration of insufficient knowledge, instances of the DM I DON'T KNOW are not employed to express propositional meaning, but are used first and foremost to express speaker attitudes and to create cohesion in discourse (see, for example, Brinton 1996: 29–40, Fraser 1999, or Andersen 2001: 38–81 for detailed discussions of the characteristics of DMs).

3.6. Extraction and Coding

---

5 For a discussion of bound tokens, see Pichler (in preparation).
Every context of unbound I DON'T KNOW was extracted from the data. Twelve tokens had to be excluded from the analysis because their form could not be determined with certainty, they occurred in quoted speech, or because interruptions by other informants prevented unambiguous utterance interpretation. A total of 239 instances of the expression were retained in the database. Each token was coded for function, age and gender.

3.7. Methods of data analysis

Drawing heavily on the theories and research methodology of conversation analysis (see, for example, Hutchby & Wooffitt 1998 or ten Have 1999 for comprehensive outlines), a function is attributed to every occurrence of the variable in the corpus. Systematic attention is here paid to the sequential context of an utterance, the temporal development of the interaction, as well as paralinguistic and, importantly, prosodic features. These features have been argued to be interactionally significant and to contribute to utterance interpretation (Heritage & Atkinson 1984: 5, du Bois et al. 1993: 49–73). Following Brinton (1996) and Andersen (2001), I will broadly distinguish between the subjective (expressing speaker and hearer attitude) and textual (contributing to and expressing coherence relations) levels of discourse. Unlike Holmes (1984a, 1984b), I include multifunctionality as a parameter in the analysis (as did, for example, Schiffrin 1987 and Andersen 2001). Hence tokens that operate simultaneously on both levels of discourse will be categorised as subjective-textual.

Employing quantitative methods of sociolinguistic variation theory (Labov 1972), distributional patterns of linguistic variation are then quantified across age, gender and function. Where feasible, a multivariate analysis of the contribution of each factor group to the occurrence of different variants is conducted with the aid of Goldvarb X (Sankoff, Tagliamonte & Smith 2005) (for details on variable rule analysis see, for example, Guy 1993, Tagliamonte 2006). The quantitative analysis will reveal the underlying mechanisms constraining the use of different variants.

4. Results

4.1. Qualitative analysis of unbound I DON'T KNOW

This section provides a brief overview of the different uses of I DON'T KNOW in the data. It thus provides the backdrop for the quantitative analysis.

4.1.1. Subjective functions of DM I DON'T KNOW

Subjective functions of DMs belong to the interpersonal level of language and are concerned with attitudes to propositions and interlocutors (Coates 1996: 156). Similar to many other DMs, I DON'T KNOW broadly speaking functions here to attenuate the strength of propositions and to mitigate face-threatening acts, i.e., acts that run contrary to addressees' and/or speakers' negative face wants (their desire to be unimpeded by others) and positive face wants (their desire to be approved of) (Brown & Levinson 1987). The following extracts from the data provide examples of I DON'T KNOW functioning subjectively:

---

6 Due to space considerations, it is not possible to provide examples for the whole inventory of pragmatic functions here. Readers are referred to Pichler (in preparation) for a more detailed discussion and illustration of functions.
(3) a. ((Barbara has just asserted that older people use more non-standard grammar than younger people.))
   HP: Why do you think that is.
   Barbara: I dunno? Maybe just just e:h education at schools.

b. 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 (?) they're no looking after their kids properly or they just (?) just letting them get away wi it.

In example (3a), Barbara hedges her utterance with I dunno, thus signalling to the interviewer that her reply might not be reliable. This allows her to withdraw from her utterance if challenged. The same is true for example (3b). I divn't knaa facilitates Luke's expression of a potentially controversial view by mitigating it, i.e., by reducing the 'unwelcome effects which a speech act has on the hearer' (Fraser 1980: 342). Note that when I DON'T KNOW functions subjectively as in the examples above, it frequently co-occurs with other epistemic markers (maybe, just, I think). This further reinforces the overall tentativeness of the utterances. It is also worth noting that subjective tokens of I DON'T KNOW are usually uttered with fall-rise or rise intonations, which in themselves are markers of doubt and tentativeness (Cruttenden 1986: 106, Brown & Levinson 1987: 172, Coates 1987: 115). Finally, when used as an epistemic marker as in (3), I DON'T KNOW occurs in turn-initial, -medial and -final positions and both before and after the propositions that it modifies. Turn-initial pre-positioning, however, is preferred.

With a more local scope, I DON'T KNOW can signal that the immediately following expression only loosely communicates speakers' thoughts and/or that speakers are unsure of their choice of wording. Other subjective tokens of I DON'T KNOW function as softeners to preface disagreements, thus attending to the hearer's positive face (see also Tsui 1991), or as avoidance strategies to mitigate interactive conflict.

4.1.2. Textual functions of I DON'T KNOW

Brinton (1996: 38) defines the textual level of discourse as that where 'the speaker structures meaning as text, creating cohesive passages of discourse'. Textual functions of the DM I DON'T KNOW in the corpus include repair, hesitation and turn-exchange devices, exemplified in this order in (4):

(4) a. Daniel: No well (.) I s- I dunno. I keep saying if it's so good where you came from, why don't you go back. That's what I say, you know.

b. HP: So would you consider Berwick to be in the larger north-eastern part of England or a larger borders area of Scotland?
   Albert: + (h) E::h (..) I divn't knaa? (..) $ I oh I consider the (.h) (..) oh I s-- I consider Berwick to be in the larger north-eastern part of England.
   HP: mhm
   HP: uh-huh

c. Rebecca: °So I dunno. ((to her interview partner)) What do you think?°
Informants' use of I DON'T KNOW can also affect the topical development of the interview. The following extract serves as an illustration:

(5) HP: Would you say that younger people (.) older people use more non-standard grammar than younger ones?
Gabriel: Yeah.
HP: Yeah? Why do you think.
Gabriel: Dunno.

Gabriel flatly denying knowledge of the reasons behind age-related language variation conveys his reluctance and/or inability to participate in a discussion of this topic (Pomerantz 1984: 57–8). His minimal response thus curtails the topic proffered by the interviewer. This curtailing effect is heightened by the falling intonation on dunno which implies finality, completeness and definiteness (Cruttenden 1986: 100), suggesting that there is nothing more to follow. The interviewer, whose positive face wants have been attended to by the provision of a minimal response, is thus prompted to move on to the next item on the interview agenda.

I DON'T KNOW is also used by interviewees to indicate their desire to close a topic and, less frequently, to discard interruptions by other interlocutors in order to pursue the original topic (see also Ford & Thompson 1996: 169–170 and Beach & Metzger 1997: 571–574). Textual instances of the variable thus include the following categories: repair, hesitation, turn-exchange and topic development. A number of turn-final instances of I DON'T KNOW were ambiguous, i.e., it was not possible to establish with certainty if speakers closed a turn with the intention of yielding it or the intention of closing the topic. To account for the ambiguity of these tokens and to avoid a subjective categorisation of them as either turn-exchange devices or topic management devices, a further category, i.e., turn-closing, was added to the inventory of functions. It includes the ambiguous tokens.

4.1.3. Subjective-textual functions of DM I DON'T KNOW

As indicated above, some tokens of I DON'T KNOW are multifunctional, simultaneously expressing a subjective and textual function. The extracts in (6) illustrate these multifunctional occurrences of I DON'T KNOW:

(6) a. HP: What accent would you say you had and do you like it?
Leah: Em it's a mixture of probably Scottish and Geordie. But I dunno.
b. HP: What's so good about Radio Borders.
Alicia: It's just (..) I don't know you can relate to (.) more of the things that they talk about sometimes because it's about the area.

In the first example, (6a), Leah's turn-final hedge simultaneously works to yield the turn to her interlocutor. In example (6b), the hesitation marker simultaneously attenuates the force of the assertion made by Alicia. As well as containing multi-functional epistemic and hesitation markers the corpus also contains instances of repair and turn-taking I DON'T KNOWS that introduce an element of tentativeness to the utterance. These tokens were categorised as subjective-textual as well.
4.1.4. Referential meaning

Roughly one fifth of the unbound tokens of I DON’T KNOW in the data convey the expression's referential meaning of 'not knowing'. The following examples are offered as illustrations thereof:

(7) a. HP: And would you rather have a different accent or dialect?
  Ryan: I don't know. Never really thought about it.

b. 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 don't know? We divn't knaa the connection.

The interpretation of these tokens of I DON’T KNOW as expressing lack of knowledge is supported by the immediately following utterances ('never really thought about it', 'we divn't knaa the connection'). Further, these instances of I DON’T KNOW lack the prosodic features usually accompanying DM usages, such as variation in speech rate, loudness or pitch range. Nor do they co-occur with other DMs, filled or unfilled pauses, which is characteristic of the DM usages of I DON’T KNOW.

In order to test whether pragmatic and referential uses of I DON’T KNOW prefer different variants, referential tokens of the variable are included in the quantitative analysis.

4.1.5. Categorisation of functions

As we have seen, textual tokens of the variable can carry elements of interpersonal and referential meaning (see (5) above). Similarly, referential tokens can carry some interpersonal and/or textual meaning. This is because instances of I DON’T KNOW operate on a continuum from more to less referential and more to less pragmatic. Whilst acknowledging this intrinsic feature of DMs, the preceding analysis describes instances of I DON’T KNOW in terms of their most salient effect on the interaction and its participants in a given context. Such categorisations are indispensable for the quantification of the data.

4.2. Quantitative analysis of unbound I DON’T KNOW

I will now turn to the quantitative analysis to reveal the conditioning factors constraining the occurrence of different surface realizations of the variable. Table 2 shows the overall linguistic distribution of the 239 tokens of unbound I DON’T KNOW in the corpus.

<table>
<thead>
<tr>
<th>don't know</th>
<th>dono</th>
<th>dunno</th>
<th>divn't knaa</th>
<th>dinnae ken</th>
<th>TOTAL N</th>
</tr>
</thead>
<tbody>
<tr>
<td>%</td>
<td>%</td>
<td>%</td>
<td>%</td>
<td>%</td>
<td></td>
</tr>
<tr>
<td>17</td>
<td>41</td>
<td>8</td>
<td>20</td>
<td>52</td>
<td>123</td>
</tr>
</tbody>
</table>

Table 2: Overall distribution of variants of I DON’T KNOW

The distribution reveals that I dunno is the most frequent variant, accounting for more than half of all tokens. I divn't knaa is the second most frequent variant, constituting more than a fifth of all tokens, closely followed by I don't know. I dono is the second least preferred variant. The variant I dinnae ken is exceptionally rare and is therefore not included in the
analysis that follows. Interestingly, when referential usages of the expression are separated from its DM usages, a different hierarchy of variants is revealed for the two uses:

- I don’t know (45%) > I dunno, I divn’t knaa (22% each) > I dono (9%) > I dinnae ken (2%) for referential uses.
- I dunno (58%) > I divn’t knaa (22%) > I don’t know (11%) > I dono (8%) > I dinnae ken (1%) for all pragmatic uses.

Further distributions not reproduced here show different hierarchies of variants on different levels of discourse (subjective, subjective-textual, textual). In fact, as revealed by the multivariate analysis presented in Table 3, function is the most significant factor group for I don’t know and I dunno.

Table 3: Variable rule analysis by function, gender and age

<table>
<thead>
<tr>
<th>Function</th>
<th>I don’t know</th>
<th>I dunno</th>
</tr>
</thead>
<tbody>
<tr>
<td>Referential</td>
<td>.83</td>
<td>.68</td>
</tr>
<tr>
<td>Subjective</td>
<td>.77</td>
<td>.50</td>
</tr>
<tr>
<td>Subjective-textual</td>
<td>.44</td>
<td>.49</td>
</tr>
<tr>
<td>Textual</td>
<td>.28</td>
<td>.58</td>
</tr>
<tr>
<td>Range</td>
<td>.55</td>
<td>.49</td>
</tr>
</tbody>
</table>

The functionally-conditioned distribution of the two variants can be summarized as follows:

- I don’t know strongly favours referential usages at .83. Subjective functions also favour this variant at .77. Subjective-textual usages have no effect on the occurrence of this variant. Finally, I don’t know is strongly disfavoured for textual functions at .28.

The data includes only two tokens of I dinnae ken in the corpus, one functioning as a turn-holder, the other as declaration of insufficient knowledge. They occurred in the speech of the oldest female and second-oldest male informant in the sample, aged 78 and 79 respectively. This suggests that I dinnae ken is on its way out in the speech community, provided that it has ever been frequent in the first place.

The multivariate analysis reveals the relative importance and significance of each factor group. Factor groups (gender, age, function) are the independent variables hypothesised to influence the occurrence of particular variants. The decimal numbers under the heading ‘factor weight’ indicate the probable incidence of individual variants with individual factors (e.g. male or female). Factor weights above .5 favour a variant, factor weights below .5 disfavour it and factor weights around .5 have no effect on the occurrence of a variant. Factor weights in square brackets do not make a statistically significant contribution to the variation. The range indicates the strength of a particular factor group’s contribution to the linguistic variation.

The variants I dono and I divn’t knaa have been included as non-application values in these runs. Because of its marginal occurrence in the corpus (8%, N=20) the results for I dono are not reproduced here. I divnt knaa will be dealt with below.
• *I dunno* favours subjective-textual uses with a factor weight of .68. Subjective and textual usages have no effect on its occurrence. Referential usages strongly disfavour *I dunno* at .19.

Gender also exerts a statistically significant effect on the variation: females favour both variants. Age does not make a significant contribution to the occurrence of either variant.\(^\text{10}\)

In order to reveal further trends in the variation, cross-tabulations are carried out for both variants. Because of the uneven distribution of tokens across speaker cells (see Appendix 2 for a breakdown of tokens according to social variables), however, these cross-tabulations have to be interpreted with caution.

Cross-tabulations of the social variables with function reveal that the preference for *I don’t know* in referential contexts and its comparatively rare occurrence in textual uses remain stable over time and across age. With regard to its subjective and subjective-textual uses, variations across age and gender cannot be interpreted meaningfully due to small token numbers.

*I dunno* is the preferred variant for all social cells apart from young males. Cross-tabulations of age and gender with function do not reveal any marked social differences in the use of *I dunno* for different functions. An apparent increase in the use of *I dunno* for referential uses with decreasing age cannot be interpreted meaningfully because the pattern is based on only ten tokens of the variant.

The most striking social differences in the data occur in the use of *I d'n't knaaa* (see Appendix 2). While marginal in female speech (8%, N=8), *I d'n't knaaa* constitutes roughly a third of all tokens of the variable in male speech (34%, N=45). This pattern is in line with many other studies showing that male speakers use a higher proportion of non-standard variants. The data further reveal a strong interaction of gender with age: *I d'n't knaaa* is used considerably more often by young males compared to their middle and older counterparts. This variation in apparent time might be suggestive of young males being in the vanguard of change in progress, a pattern frequently found in sociolinguistic studies.

Figure 1 illustrates the uneven distribution of tokens of *I d'n't knaaa* in the data which makes it unfeasible to conduct a multivariate analysis of this variant.

**Figure 1: Distribution of I d'n't knaaa across speaker cells**

\(^{10}\) A cross-tabulation of the data reveals a decline in the use of *I don’t know* with decreasing age for male speakers, while the variation remains stable among females. This explains why age was not chosen as making a significant contribution to the occurrence of *I don’t know* despite the differences in factor weights.
Two young males, referred to here as Adam and Luke, are responsible for almost three quarters of all occurrences of the variant in the corpus. Middle speakers are responsible for less than a fifth of all tokens of I divn't knaa. Older speakers use the variant even less often than middle ones and other young speakers virtually reject it.

The question that arises from this uneven distribution of tokens is: What causes a subsample of young male speakers to use a non-standard localised variant considerably more often than other speakers in the sample? A cross-tabulation of speaker groups with function provides some insight.

Figure 2: Functional distribution of I divn't knaa by speaker groups

Figure 2 reveals that, unlike all the other speakers in the sample, Adam and Luke use the variant I divn't knaa not only for referential and textual uses but also for subjective-textual and subjective functions. What we witness here then is an extension in the use of I divn't knaa, i.e., the use of the variant in new contexts (Heine 2003:579–580), amongst a subsample of speakers.

5. Discussion

In her analysis of I DON'T KNOW in American English conversational data, Scheibman (2000) found that full and reduced variants are used for the expression of lack of knowledge, but only reduced forms are used for DM functions. Although not as clear-cut as Scheibman's, the findings of the present analysis also suggest a functional trend. I dunno, the most frequent variant in the corpus, is strongly disfavoured for the declaration of insufficient knowledge and favoured for subjective-textual usages; the variant I don't know is strongly favoured for referential and subjective uses and disfavoured in textual contexts. The variation between the full and reduced forms of the expression I DON'T KNOW thus involves functional divergence. Due to small token numbers, it is not possible to assess the stability of this trend across age and gender with any degree of certainty.

As for the non-standard localised variant I divn't knaa, the analysis shows that two young male speakers are responsible for almost three quarters of all tokens of the variant in the data. Cross-tabulations further reveal that these two speakers use the variant for a wider range of functions than other speaker groups in the sample. The analysis therefore suggests that one factor contributing to the sudden increase in the use of I divn't knaa amongst a subsample of young male speakers is an extension in the use of this variant to new levels of discourse, i.e., subjective and subjective-textual, amongst these speakers.
The analysis thus suggests that grammaticalization and subjectification processes lead
to a functional divergence of the standard variants and an increase in the use of the non-
standard localised variant. These processes are further explored in Pichler (in preparation).

6. Conclusion

The present study has shown that the variation between standard variants of the
variable I DON’T KNOW can be explained in terms of its functions in discourse. Further, the
analysis revealed that the increase in the use of a non-standard localised variant can be
explained, at least in part, by an extension in use. This paper thus highlights the advantages of
combining qualitative with quantitative methods in the study of sociolinguistic variation.

Appendix 1: Transcription conventions

The conventions are largely borrowed from du Bois et al. (1993) and Sacks, Schegloff

[   ] overlap
= continuation of a turn
___ emphasis
- truncated words
< > allegro speech
@ laughter
(h), (.h) inbreath, oubreath
+ tisking
$ audible swallowing

Appendix 2: Distribution of variants of I DON’T KNOW by gender and age

<table>
<thead>
<tr>
<th></th>
<th>Male</th>
<th>Female</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>don’t know</td>
<td>dono</td>
</tr>
<tr>
<td>old</td>
<td>%</td>
<td>%</td>
</tr>
<tr>
<td>middle</td>
<td>41</td>
<td>9</td>
</tr>
<tr>
<td>young</td>
<td>12</td>
<td>6</td>
</tr>
<tr>
<td>TOTAL</td>
<td>10</td>
<td>13</td>
</tr>
</tbody>
</table>

References

Benjamins.

Cambridge University Press.


