Pichler H, Hesson A. Discourse-pragmatic variation across situations, varieties, ages: I DON'T KNOW in sociolinguistic and medical interviews. *Language & Communication* 2016, 49, 1-18.

Copyright:

© 2016. This manuscript version is made available under the [CC-BY-NC-ND 4.0 license](#)

DOI link to article:

<http://dx.doi.org/10.1016/j.langcom.2016.04.002>

Date deposited:

30/06/2016

Embargo release date:

14 May 2018

This work is licensed under a
[Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International licence](#)

Discourse-pragmatic variation across situations, varieties, ages:

I DON'T KNOW in sociolinguistic and medical interviews

Heike Pichler^{a,*}, Ashley Hesson^b

^a School of English Language, Literature and Linguistics, Newcastle University, Percy Building,
Newcastle upon Tyne NE1 7RU, UK

^b College of Human Medicine, Michigan State University, 1355 Bogue Street, East Lansing, MI
48824, USA

ABSTRACT

Our analysis of I DON'T KNOW (IDK) variability in four sub-corpora of sociolinguistic and mental health interviews clarifies the context-dependency of discourse-pragmatic variation, and demonstrates the translational potential of socially responsible linguistic research. Systematic and accountable analysis of over 700 tokens of IDK in these data reveals that while IDK variants, functions and syntactic configurations are differentially distributed across situation, variety and age, form-function correlations in IDK use are robust across these parameters. These results demonstrate that discourse-pragmatic variables such as IDK retain some probabilistic relationships that aid in consistent interpretation across contexts of use, while other aspects of variability are used in stylistically distinct ways across e.g. situations and social groups. Moreover, our analysis of IDK in mental health interviews shows how the variationist analysis of

* Corresponding author: Tel. +44 (0)191 208 3519

E-mail addresses: heike.pichler@ncl.ac.uk (H. Pichler), bartell6@msu.edu (A. Hesson)

site-particular communicative practices can uncover language variation patterns that have the potential to affect practice, research and teaching outside of the sociolinguistics community.

Key words: I DON'T KNOW, discourse-pragmatic variation, situational context, doctor-patient communication, British and American English

1. Introduction

The sociolinguistic interview is a mainstay of variationist methodology and continues to constitute the primary data source for studies of language variation and change.² However, some variationists have successfully utilized other data sources and demonstrated how their analysis allows investigation of contextual effects on language use not readily observable in sociolinguistic interview data. Coupland (1980), for example, recorded interactions at a travel agency to characterize the effect on linguistic variation of situational factors that are invariable in (one-to-one) sociolinguistic interviews: participant and channel. Other scholars ventured to playgrounds and schools (Cheshire, 1982) or barbecues and hospitals (Podesva, 2007) to observe the impact of situational context on language use; moreover, recent research has made use of diverse corpora to compare language use across workplace and classroom settings (Holmes and Schnurr, 2006; Schlee, 2008a; see also Chaemsaihong [2012] for a theoretical discussion of situational linguistic variation). In this paper, we support both the extension of variationist analyses beyond the sociolinguistic interview and the cross-situational analysis of language variation patterns (e.g. Finegan and Biber, 1994, 2001) while also attending to the effects of more widely studied social predictors such as variety and age.

Our investigation focuses on the construction I DON'T KNOW (IDK) in two situationally distinct yet structurally comparable settings:³ sociolinguistic and mental health interviews. It is motivated by our aims to: (i) clarify the role of situational factors in linguistic variation and advance our understanding of the nature of discourse-pragmatic variability (see Pichler, 2010, p.

² Becker (2013) draws attention to the inconsistent use of the term 'sociolinguistic interview': some use it broadly to refer to any face-to-face interaction between members of a speech community and a linguistic researcher that is recorded for the analysis of sociolinguistic variation and change; others use it more narrowly to refer to the techniques developed by Labov (1966, 1984) for the systematic elicitation of different contextual speech styles from individual speakers. In this paper, we apply the label 'sociolinguistic interview data' to any naturalistic speech data obtained for variation analysis in an informal interview setting and involving a linguistic fieldworker, regardless of whether the interview protocol is structured to elicit different contextual styles.

³ The term 'construction' is used here and throughout in a pre-theoretical way to refer to syntagmatic strings.

586); (ii) explore new impact strategies for variationist research that, if developed in close collaboration with medical practitioners, may benefit mental health patients (PTs) of different dialects and ages. To meet these objectives, our data are differentiated not just by situational context; the medical interview data are further differentiated by variety (British vs. American English) and broadly construed age groups (adolescent vs. young adult vs. adult). Our variationist analysis of the variable functional, morpho-phonological and syntactic properties of IDK in these data allows us to explore both inter- and intra-situational differences in IDK use, and establish to what extent different dimensions of discourse-pragmatic variability are generalizable beyond the immediate context, variety and/or social group in/by which they were produced. Moreover, the analysis of mental health data affords us the opportunity to identify situation-specific patterns of IDK use and propose ways in which health-care practitioners' (DRs') knowledge of these patterns could affect diagnosis and treatment decisions.

Previous (largely non-variationist) studies of IDK in non-medical discourse, including media, courtroom and sociolinguistic interviews as well as casual conversations, have shown that the function, form and syntax of IDK are highly variable.⁴ IDK functions as a cognitive claim of insufficient knowledge and a non-cognitive interactional device with interpersonal, face-saving functions as well as textual, turn-exchange or topic-development functions (e.g. Beach and Metzger, 1997; Pichler, 2009, 2013; Potter, 1996; Tsui, 1991; Weatherall, 2011; Wooffitt, 2005). The construction's form is variable in terms of the vowel quality in *don't* (full vs. reduced) and the morpheme boundary between *don't* and *know* (audible vs. non-audible), locating variants on a continuum of articulatory reduction (Bybee and Scheibman, 1999; Pichler, 2009, 2013;

⁴ Throughout this paper, we follow the terminological conventions established in Pichler (2013) and use 'form' to refer to the variable morpho-phonology of IDK as described below and in more detail in Section 2.2. The 'form-function' correlations discussed in the paper thus refer to correlations between the function of IDK and its variable degree of articulatory reduction. We do not use 'form' to refer to other variable linguistic properties of IDK such as its syntactic configuration.

Scheibman, 2000). In terms of syntax, IDK variably occurs without overt complementation, with a dependent WH-word, or with a dependent phrasal or clausal complement (Pichler, 2013, p. 77–78). Example (1) from our mental health interview data shows that the variability of IDK with respect to its function, form and syntax extends to talk in this situational context.⁵ When questioned about his recent mental wellbeing, the adolescent PT in (1) uses IDK with non-cognitive (*I dunno*-1) and cognitive functions (*I don't know*-2, *I don't know*-3), in its reduced and full forms (*I dunno*, *I don't know*) as well as without (*I dunno*-1) and with dependent complements (*I don't know*-2, *I don't know*-3).

(1) DR: So how has the depression been?

(..)

PT: ***I dunno***-1.

(..)

DR: Like, um (..) say on a scale of zero to ten. Zero is

PT: == I don't know when (.) ***I don't know***-2 when I feel depressed. ***I don't know***-3 the feeling of depressed.

⁵ All examples are reproduced *verbatim* from the Verilogue or Berwick English corpora introduced in Section 2.1. The variable orthographic presentations of IDK (*I don't know*, *I dono*, *I dunno*) represent different variants whose phonetic, phonological and morphemic properties are introduced in Section 2.2. In examples (1)-(4) and (6)-(18), the following transcription conventions are used:

[], [[]]	overlap	., ::, :::	syllable lengthening
==	latching	(.), (..), (...)	short, medium, long pause
=	turn continuation	.	final intonation contour
-	false start, truncation	,	continuing intonation contour
(h)	inbreath	?	rising intonation contour
@	laughter	(text)	uncertain transcription
> <	reduced tempo	(?)	indecipherable words
CAPITALS	louder than surrounding talk	[text]	extra-linguistic information
<u>underlining</u>	emphatic stress	[...]	text omitted
° °	soft speech	bold type	highlighting of the variable or otherwise important text
superscript font	higher than usual pitch		

In the examples in (5), underlining is used to highlight distinctions made in data coding.

(Verilogue 34993, AmE youth PT aged ≤ 17)

In this paper, we build on previous studies of IDK to explore whether and to what degree the construction's linguistic variability illustrated above is constrained by the extra-linguistic factors situation, variety and age. A wealth of studies have documented that the frequency, function, form and positioning of discourse-pragmatic variables is affected by: (i) a range of social factors, including: age, sex, ethnicity, socio-economic background, community of practice membership, locality (e.g. Andersen, 2001; Cheshire, 1981, 2007; D'Arcy, 2005; Drager, 2016; Erman, 1992, 2001; Ferrara, 1997; Fuchs and Gut, 2016; Holmes, 1995; Macaulay, 2005; Moore and Podesva 2009; Pichler, 2016; Stubbe and Holmes, 1995; Waters, 2009); and (ii) diverse situational and interactional parameters, including: discourse type and activity context; topic, purpose and attitudes to the interaction; speaker roles and relationships; and communicative channel (e.g. Cameron et al., 1988; Escalera, 2009; Freed and Greenwood, 1996; Fuller, 2003; Jucker and Smith, 1998; Kyratzis and Ervin-Tripp, 1999; Lam, 2009; Mauraanen, 2004; Redeker, 1990; Schleef, 2008b; Stubbe and Holmes, 1995; Verdonik et al., 2009). However, we hypothesise that not *all* dimensions of discourse-pragmatic variation are affected by social and situational factors. Bybee and Scheibman's (1999) and Pichler's (2013) analyses of IDK in American English (AmE) casual conversations and British English (BrE) sociolinguistic interviews reveal identical form-function correlations in the construction's use: in both datasets, phonetically full IDK variants are strongly associated with cognitive uses while phonetically reduced variants are associated with non-cognitive uses. Notwithstanding the fact that Bybee and Scheibman's and Pichler's results are not reliably comparable,⁶ they suggest that when variation

⁶ In contrast to Bybee and Scheibman (1999) who identified two variants of IDK based on the variable vowel quality in *don't*, Pichler (2013) identified three non-localized variants of IDK because she also considered whether the

in form and function are correlated, discourse-pragmatic variation patterns may be consistent across situational contexts and, possibly, varieties. Furthermore, Pichler's results suggest that these correlation patterns are only marginally affected by social factors.

Our following variationist analysis sets out to reliably test the hypothesis that while the distribution of IDK forms, functions and syntactic configurations is inconsistent across extra-linguistic contexts, form-function correlations in IDK use are maintained across these contexts. We therefore extend Pichler's (2013) variationist analysis of IDK in sociolinguistic interviews to BrE and AmE child/adolescent and adult mental health interviews, i.e., (semi-)structured interviews conducted by DRs to assess their PTs' psychopathology and/or discuss their treatment plans and effects. The results confirm our hypothesis: while some but not all IDK variation patterns are robust, none of the patterns explored are straightforward indicators of social group membership or situational context.

Discussion of our medical interview results will also show how the insights from our variationist analysis of IDK might help overcome potential communicative barriers to effective healthcare provision, and improve DRs' diagnostic accuracy and therapeutic effectiveness when treating PTs with mental health disorders. In our mental health interview data, IDK occurs in a variety of contexts, including: answers to information requests (2); discussions of treatment options (3); and co-constructions of mutually satisfactory DR-PT relationships (4).

(2) DR: Um, would you like to try a different antidepressant. Or were you thinking maybe talking therapy or perhaps even a combination of both maybe?

morpheme boundary between *don't* and *know* was audible. While Bybee and Scheibman's qualitative analysis of IDK is implicitly bottom-up, usage-based and frequency-driven, Pichler's bottom-up approach to qualitative IDK analysis is first and foremost informed by insights from conversation analysis. Moreover, the generalizability of the results reported above is constrained by the fact that Bybee and Scheibman's results are based on low token numbers (N=37).

PT: I don't, *I don't know*. *I* really *don't know* what the answer is.

DR: Mhm.

(Verilogue 99560, BrE adult PT aged ≥ 35)

(3) PT: And to be honest, I think (..) the the kind of tiredness I feel, I think it, you know,

it it feels like, it (.) you know, I do need to kind of perk things up in there. So, but I

don't, but, obviously *I don't know* until I've tried these things. (h) You

DR: Hum.

PT: = know, what (.) what (.) what they kind of feel like. You know what, I, I, you know,

I don't even *know* (.) I mean I, *I don't know* if it will change my personality. I sort of hope not.

(Verilogue 89504, BrE adult PT aged ≥ 35)

(4) DR: Any effect from that? That was for the concentration and [attention problem]?

PT: [E::m. It seemed] like

(.) I think it was a little difference, but (.) you know, not enough. *I don't know* if the dosage is maybe

DR: == We may have to go a little higher.

(Verilogue 90519, AmE adult PT aged ≥ 35)

IDK, then, regularly occurs in tasks that are central to the medical interview and to the provision of PT-centred care (Frankel, 2002, p. 85; Heritage and Maynard, 2006; McWhinney, 1989; Ong et al., 1995, p. 903–904), a view supported by Hutchby's (2002) work on IDK in

naturally occurring child counselling sessions. Hutchby closely analysed one six-year-old's recurrent use of IDK in response to counsellor questions about therapeutic topics, i.e., the child's supposedly problematic home environment (his parents' strained relationship, their conflicting explanations for their actions). The analysis revealed that the child strategically used IDK to frustrate the counsellor's lines of questioning and to resist talking on counselling-relevant topics. Crucially, the counsellor understood the child's particular IDK usage: at key points of the interaction, the child's IDK did not act as a non-cognitive strategic device but as an actual cognitive claim. This understanding enabled the counsellor to develop a therapeutically fruitful way of overcoming the child's resistance. Instead of re-stating questions or simply giving up, the counsellor validated the child in his resistance to talking by reassuring him that it was ok to 'not know' considering his young age and the difficulty of his situation. Hutchby's study thus illustrates the importance of DRs' ability to differentiate cognitive and non-cognitive uses of IDK in order to make successful therapeutic interventions. However, as Hutchby (2002, p.164) points out:

it is not always immediately clear when children are in fact resisting, or even *that* they are resisting. The counsellor here is faced with the task of judging, in the course of the counselling session itself, what it is that the child's 'Don't know's are really doing.

Assuming that Hutchby's assessment of the therapeutic importance of correct IDK interpretation extends across mental health contexts, our variationist investigation of the context-, variety- and age-specific nature of IDK usage has the potential to facilitate successful DR-PT communication in this clinical context. It describes patterns of IDK usage specific to this context that, if deemed

useful and relevant by practitioners, can be translated into empirical guidelines for DRs' interpretation of PTs' IDK function.⁷ The research presented here thus explores the possibilities for future cross-disciplinary collaboration between variationist linguists and mental health providers to promote effective health communication.

The remainder of this paper is structured as follows: Section 2 introduces our data and methodology. Sections 3 and 4 present the results of the qualitative and quantitative data analyses. The theoretical implications and potential practical applications of the results are discussed in Section 5. Finally, Section 6 is the conclusion.

2. Data and methods

2.1 Corpora and data samples

In order to address the research aims set out in Section 1, we conduct a cross-situational, -variety and -age analysis of data drawn from: Pichler's (2013) private, 260,000-word sociolinguistic interview corpus collected in Berwick-upon-Tweed, northeast England; and the corporately owned, individually licensed Verilogue, Inc. database of audio-recorded DR-PT interactions (Kozloff and Barnett, 2006). Pichler's (2013) Berwick English (BwE) interview data were collected between 2003 and 2005 using the interview protocol designed for the Survey of Regional English to solicit information about informants' attitudes towards their locality and dialect and their use of local dialect words (Llamas, 2007). The original speaker sample was composed of 36 white, working-class speakers aged between 17 and 81, and was stratified by speaker sex and age. The Verilogue corpus contains a wide range of healthcare conversations,

⁷ The value of our work lies in the fact that any such guideline would be derived from the analysis of *mental health interviews*. As shown by Tay (2011), psychotherapy and mental health counselling in general are specialized contexts for discourse-pragmatic variation. Our findings support this view: clinical recommendations about IDK cannot be derived from existing analyses of IDK in other contexts (see further Sections 4 and 5).

including routine, outpatient mental health consultations between paid psychiatrists practising in the UK or USA and their unpaid, consenting PTs (and their caregivers). The de-identified recordings and transcriptions, together with information about both DR (sex, years in practice) and PT (demographic information, diagnosis), are uploaded on secure servers in a fully searchable format. Our focus on mental health interviews (as opposed to other healthcare interactions) recorded in the UK and USA was motivated by our research aims set out in Section 1 and to facilitate comparison with qualitative literature on IDK use in this setting (Hutchby 2002).

Though not solicited specifically for this purpose, the Verilogue medical interview data lend themselves to the analysis of sociolinguistic variation (see also Hesson, 2014) and to comparison with Pichler's (2013) sociolinguistic interview data. Although talk produced in this context may not be representative of PTs' most vernacular speech style (see Labov, 1972, p. 85), the medical data nonetheless contain abundant examples of vernacular features characteristic of casual speech styles (see, for example, the double negative in extract (5) or the use of discourse *like* and *cos* in extract (17)). Furthermore, the mental health interviews were conducted and recorded in quiet settings, and DRs recorded PTs' basic demographic information. This facilitates the reliable coding of IDK variants (see Section 2.2) and social factors.

Notwithstanding minor differences between individual interviews, Pichler's (2013, p. 22–25) sociolinguistic interviews and the Verilogue mental health interviews share these characteristics: they follow a fairly structured question-answer format; they involve dyadic or triadic interactions between relative strangers whose roles are clearly defined as DR/interviewer and PT/interviewee; and they are primarily informational and only secondarily interpersonal. The structural, social and psychological similarity of the mental health interviews promotes their

internal comparability as well as reliable comparison with Pichler’s sociolinguistic interviews. The main difference between the medical and sociolinguistic interviews that is likely to impact participants’ linguistic behaviour is in their respective physical settings and interactional goals: the former were conducted in institutional settings to gather information required for the diagnosis and treatment of mental health disorders; the latter were conducted in non-institutional settings to collect speech data for the analysis of linguistic variation. Thus, comparison of IDK variation patterns across the sociolinguistic interviews and the different sub-corpora of mental health interviews introduced below allows us to reliably assess the effect of situational context, variety and age on IDK variation.

The final sample composition was determined by our aims to investigate the effect on IDK variation of multiple contextual factors as well as by the recordings and demographic information available in the Verilogue database. To this end, we included in the analysis all BwE sociolinguistic interviews with adults aged ≥ 18 ($N_{\text{speakers}}=27$, $N_{\text{tokens}}=353$)⁸ as well as random sub-samples of the following Verilogue mental health interviews: interviews recorded in the UK with adults aged ≥ 18 ($N_{\text{speakers}}=14$, $N_{\text{tokens}}=91$); interviews recorded in the USA with adults aged ≥ 18 ($N_{\text{speakers}}=33$, $N_{\text{tokens}}=120$); and interviews recorded in the USA with youth aged ≤ 17 ($N_{\text{speakers}}=33$, $N_{\text{tokens}}=150$). The speaker samples in all sub-corpora are stratified by sex (male vs. female); the adult sub-corpora are further stratified by broad age groups (young adults aged 18-34 vs. adults aged 35+).⁹ The age groupings in our data (youth = ≤ 17 -year-olds; young adults = 18-34-year-olds; adults = 35+-year-olds) are dictated by a combination of the following:

⁸ To ensure reliable and valid results, we excluded from Pichler’s original analysis of the BwE sociolinguistic interview corpus the following tokens: those extracted from interviews with a small number of 17-year-olds; those realized as *I divn’t knaa*, *I dinnae ken*, i.e., variants that do not occur in any of the other sub-corpora; and tokens that had been coded as performing functions across multiple interactional domains (see further Section 3).

⁹ The medical sub-corpora contain interviews with patients of different racial backgrounds and with different primary mental health diagnoses (e.g. ADHD, depression, schizophrenia). Due to the small number of non-white PTs and the uneven representation of primary diagnoses, race and diagnosis are not included as factor groups in the multivariate analyses in Section 4.

variationist insights, i.e., the well-documented stabilization of speakers' vernaculars at age 17 (Labov, 2001, p. 447; Tagliamonte and D'Arcy, 2009, p. 66); traditional divisions in medical practice, i.e., that PTs aged ≤ 17 are seen by youth psychiatrists and those aged ≥ 18 by adult psychiatrists; and Verilogue's conventions of recording metadata, i.e., the fact that for adult PTs Verilogue provides age-brackets rather than discrete ages. Inclusion in our analysis of the combination of sub-corpora listed above allows for comparison of the effect on IDK variation of: situational context (BwE sociolinguistic vs. BrE mental health interviews); variety (BrE vs. AmE adult mental health interviews); and age (AmE adult vs. youth mental health interviews). Unlike the BwE sociolinguistic interview data, the mental health interview data were collected in a variety of locations across the UK and the USA; any results gauged from these data are therefore representative of a variety of speakers and regions rather than one particular speech community.

2.2 *The dependent and independent variables: data coding*

Careful inspection of the sub-corpora introduced in Section 2.1 yielded a coding file containing 714 tokens of the variable IDK, defined here as all forms 'derived from the same linear string of components schematized as: (first person singular pronominal subject) + (negative periphrastic DO) + (predicate *know*)' (Pichler, 2013, p. 31).¹⁰ To test the hypotheses set out in Section 1, each IDK token had to be coded for: its variant form, syntactic configuration, and discourse functionality; the sub-corpus in which it occurs; and the sex and age of its speaker. By coding each token of the variable for its linguistic characteristics (variant form, syntax, function), we allow for cross-corpora consideration of the independent distribution of these variable properties as well as the form-function relationship reported in the literature (see Section 1), thus catering to

¹⁰ While our overall token number may be considered modest by some standards, all cell counts were meticulously reviewed during the analytic process to ensure that, in accordance with variationist norms (Levon, 2010, p. 78), all factors included in the multivariate analysis of the data (see Table 3) contained cell counts of >5 .

the multi-dimensional nature of discourse-pragmatic variation. Even if correlations between morpho-phonological form and discourse-pragmatic function were robust across situational context, variety and age (as determined by coding each IDK token for its sub-corpus), the independent proportional distributions of variant forms, functional categories and syntactic configurations of IDK may differ between sub-corpora *irrespective* of any form-function correlations found (see Section 1). By coding each IDK token for social predictors, we seek to establish the social robustness of IDK variation within corpora.¹¹

The sociolinguistic sub-corpus had already been coded for the operation of the above constraints in Pichler's (2009, 2013) previous analysis of these data. Close inspection of a subset of the AmE youth medical interviews established that Pichler's (2013, p. 77-81) coding protocol for form and syntax could be extended without modification to our medical interview data. Thus, in terms of variant form, we initially divided IDK tokens into three form categories distinguished by their lack or degree of reduction. In full variants of IDK, orthographically represented as *I don't know*, the vowel in *don't* is realized with lip-rounding and a conspicuous morpheme boundary occurs between the nasals of *-n't* and *know*. In semi-reduced variants, orthographically represented as *I dono*, the first vowel is produced with lip-rounding but in contrast to the full variants there is no morpheme boundary between the two nasals. Finally, in reduced variants, orthographically represented as *I dunno*, the vowel in *don't* is reduced to a schwa and there is no audible morpheme boundary between the nasals. For our multivariate analysis in Section 4, we consolidate the three-way variant distinction into a two-way distinction of full vs. (semi-)reduced variants. In line with a step-wise theory of grammaticalization (Traugott and Trousdale, 2010), we suggest that the absence of an audible morpheme boundary between the nasals of *-n't* and

¹¹ Because the focus of our analysis is on uncovering the situational and social robustness of IDK form, function and syntax variability and because Wagner et al.'s (2016) analysis of general extenders has shown that frequency is not a reliable indicator of register variation, we will not investigate in this paper cross-data differences in IDK frequency.

know is the result of fusion and rebracketing, where ‘[f]usion [...] is a characteristic of the right-hand side of a continuum at the left of which are discrete morphs and at the right end a single morph’ (Hopper and Traugott, 2003, p. 58). Unlike full variant forms, both the semi-reduced and reduced variant forms have undergone this step in the grammaticalization process, thus warranting their re-categorization as one variant form for quantitative purposes.

With regard to syntactic configuration, we differentiated between: unbound tokens of IDK without an overt complement (5a) versus bound tokens with a dependent WH-word (5b) or with a phrasal/clausal complement (5c).

- (5) a. Um, I guess it’s been a little bit better this week. ***I dunno***. It’s, it’s really a different day today. (Verilogue 52315, AmE youth PT aged ≤ 17)
- b. Noone likes taking the medication. ***I don’t know*** why. But none of it was good. (Verilogue 95445, BrE adult PT ≥ 35)
- c. ***I don’t know*** no president. (Verilogue 57924, AmE youth PT ≤ 17)
- I dono*** if it’s anxiety. ***I dono*** what it is. (Verilogue 58886, AmE youth PT ≤ 17)

Because the functionality of discourse-pragmatic variables is strongly context-dependent (see Section 1), we could not uncritically apply to our mental health interview data the functional taxonomy developed by Pichler (2013) to describe the use of IDK in the BwE sociolinguistic interview data. Such an approach would have risked yielding inadequate descriptions of IDK usage in medical interactions. Therefore, to develop a functional taxonomy of IDK use in mental health interviews, we adopted a bottom-up approach which involved close examination of every occurrence of IDK in these data and was informed by a detailed understanding of the

interactional dynamics of medical interviews (Bergmann, 1992; Heritage and Maynard, 2006).

The methods used to establish the functionality of IDK tokens are introduced in more detail in Section 2.3 below; the schema developed for coding medical IDK tokens for function is detailed in Section 3, which presents the results of the qualitative analysis.

The basic procedures used to reduce the subjectivity of the functional analysis of IDK in the sociolinguistic interview data are described in Pichler (2013: 48). To further avoid biasing the analysis of the medical data, we adopted the following practices for linguistic predictors. First, we independently coded the form, syntax and function of all medical IDK tokens. To the extent possible, the second author focused exclusively on the morpho-phonological form and syntactic configuration of medical IDK tokens while the first author focused exclusively on their discourse-pragmatic function. Throughout, we were blinded to each others' codings. Next, we swapped coding roles to determine inter-rater reliability. The first author blindly coded a random sub-sample of 20% of IDK tokens in each medical sub-corpus for function while the first author coded the same tokens for form and syntax. This process revealed the following percent-match levels of agreement between coders: 87% for form; 98% for syntax; and 78% for function, indicating acceptable levels of inter-rater reliability across linguistic predictors (see further Wagner et al. [2015] for expected levels of inter-reliability in discourse-pragmatic function coding).¹² Tokens whose form, syntax or function categorization proved contentious or otherwise problematic were brought to collaborative attention and interrogated until inter-coder agreement was reached.

2.3 Methods of data analysis

¹² Percent agreement was selected, as opposed to Cohen's kappa or other weighted measures of inter-rater reliability, based on data structure and the type of ratings being tested, i.e., principled, scheme-based assessments (Hugh 2012).

As stated above, the IDK tokens in the sociolinguistic interviews had already been coded for function by Pichler (2013). To develop a taxonomy of IDK functions in the mental health interviews, we adopted a bottom-up approach. Like Pichler (2013), we heavily drew on key insights from conversation analysis (Atkinson and Heritage, 1984; Schegloff, 2007) to inform our interpretation of the data, making our method comparable to hers yet sensitive to context-specific IDK use. In the identification and allocation of IDK functions, we thus paid systematic attention to the sequential context of the turns containing IDK, the temporal development of the interaction, as well as to the ambient linguistic context, including other discourse-pragmatic variables, (un)filled pauses, and the prosodic encodings of IDK. These features are interactionally significant and served to disambiguate the construction's functions in a given context. Also, utterance interpretation was guided by next speakers' treatment of preceding talk.

In order to identify and account for patterns of IDK variation in the data, we first compared the overall form, function and syntax distribution of IDK across the four sub-corpora, employing the Pearson χ^2 -test to assess statistical levels of significance in variation patterns. We then performed mixed-effect logistic regression analyses (in Rbrul; Johnson, 2009) for each of the sub-corpora, predicting the full variant *I don't know* as opposed to a combined category of (semi-)reduced variants (*I dono*, *I dunno*). Multiple logistic regression analysis was selected as our primary quantitative method because it allowed for the identification of statistically significant correlational patterns in the distribution of IDK forms, functions and social co-variants. Identical cross-corpora analyses were conducted and compared to: (i) uncover any cross-corpora differences in the variants' underlying grammar; and (ii) test our hypothesis that form-function correlations are robust across situational contexts, varieties and age groups. To mitigate adverse effects of badly distributed data and meet the frequency thresholds for viable

statistical modelling (Guy, 1980), we configured function along the three domains referential, interpersonal, textual (see Section 3.7). Additionally, because of its interaction with function, syntax was not included as a predictor in the multivariate analyses. Finally, we included individual speaker as a random effect in the regressions to reduce the risk of providing inaccurate significance estimates of social factors.

3. Qualitative analysis: functions of IDK in mental health and sociolinguistic interviews

In this section, we provide an overview of the functional repertoire of IDK in our data which we derived through application of the methods outlined in Section 2.3. The novelty and contribution of what follows lies in: (i) developing Hutchby's (2002) case study of IDK into a large-scale qualitative review of IDK in medical consultations to provide a more comprehensive description of the construction's functional spectrum in this situational context; (ii) demonstrating that the functional diversity identified for IDK in other situational contexts (see Section 1 for details) extends to medical interviews; and (iii) illustrating that the construction's interactional effects at times strongly depend on its situational context of use. Our comprehensive description of IDK functions in our mental health interview data enables us to illustrate the importance of differentiating the multitude of functions IDK performs in medical practice (see Section 5.2). For these reasons and because a detailed overview of IDK functions in sociolinguistic interviews is provided in Pichler (2013, p. 81-99), the focus below will be on describing IDK functions in mental health interviews. We briefly outline how we arrived at the functional categorizations of tokens (see further Pichler, 2013, p. 81-99; Hesson and Pichler, MS), and provide information

about contextual features that typify individual functional categories and contribute to the disambiguation of IDK meanings.¹³

For the medical interview data, we initially developed a micro-taxonomy of IDK functions reflecting the construction's high functional versatility as well as individual tokens' intrinsic polyfunctionality. However, the lack of robust mutual exclusivity between the resulting categories made us question the objectivity and practicality of differentiating subtle nuances of IDK meaning, especially if the taxonomy is to be: (a) quantified to test our hypotheses about IDK variability; and (b) applied in clinical settings where DRs would have to interpret the functionality of IDK on-line in interaction and, in most cases, without our specialist knowledge of discourse functionality. We therefore reconfigured our initial micro-taxonomy of IDK into a macro-taxonomy of six functional categories (see Table 1 in Section 3.7). These categories are sufficiently distinct to capture the construction's functional versatility in the data without sacrificing their heuristic value for identifying patterns of variation in IDK use or jeopardising their applicative potential for interpreting PTs' IDK use in real time. Yet at the same time, they are respectful of the construction's intrinsic multifunctionality. They incorporate tokens that are subtly different in meaning but perform comparable and overlapping interactional functions (see further Sections 3.1-3.6). Finally, the application of this taxonomy across the medical sub-corpora led to levels of inter-rater agreement comparable to those reported for other discourse-pragmatic variables (Wagner et al., 2015; see further Section 2.2), making the analysis (and its possible future application in medicine) more reliable.

Below we proceed as follows. We first distinguish in Sections 3.1 to 3.6 six broad categories of IDK use (cognitive claim, epistemic marker, mitigation device,

¹³ For considerations of space, we are unable to thoroughly catalogue associations between individual IDK functions and different sequential environments. This is not to deny the potential value of such an exploration for linguistics and/or medical practice.

avoidance/resistance strategy, turn-exchange device, topic development device). We then classify the six functional categories into the three domains: referential, interpersonal and textual (see Section 3.7); this will highlight the context-sensitivity of IDK use and facilitate the quantification of the data in Section 4. Where possible, we illustrate each of the six functional categories with one example from the sociolinguistic interview data and at least one from the mental health interview data. The examples are presented in their larger sequential and interactional context of occurrence, and the typographical means detailed in footnote 4 are employed to replicate accompanying paralinguistic and prosodic features. As shown below, these details informed data interpretation. Where individual data extracts contain multiple instances of IDK, it is the bolded token that is the focus of the preceding commentary.

3.1 Cognitive claim of insufficient knowledge

Sociolinguistic interviewees (IVEs) and mental health PTs regularly use IDK to claim a cognitive state, i.e., lack of knowledge. As illustrated in (6) and (7), cognitive tokens of IDK most often occur in direct response to interviewer (IVR)/DR information questions where they signal IVEs'/PTs' cognitive inability to supply the requested information. The interpretation of these tokens as knowledge disavowals is supported by the following discourse: the IVE in (6) follows IDK with another, emphatic declaration of his lack of knowledge (*I've na idea*); the PT in (7) follows his initial emphatic knowledge disavowal (*I don't even know*) with an elaborate account for the declared knowledge gap: because his father is not an accessible source of information, the PT is unable to provide the DR's requested information about his mental health. Unlike, for example, epistemic or turn-holding IDK tokens (see Sections 3.2 and 3.5), cognitive

IDKs are not generally produced with rising intonation and/or accompanied by other discourse-pragmatic features, filled or unfilled pauses.

(6) HP: Why is drugs so bad in Eyemouth?

Luke: == *I dunno* what it is. I've nay idea.

(Sociolinguistic interviews, BwE young adult IVE aged 18-34)

(7) DR: Well, you know, you have a strong family history. Your biological father, [there is]

PT: [mhm]

DR: = a (.) strong history of what? Bipolar and (..) schizophrenia too, right?

PT: *I don't even know* anything about him, because he left when I was two weeks old. (.)

I had just been born two weeks, and he, he left. And I only saw him one time when I was five years old, and that was for a couple hours, and I didn't like him, I never went back. Haven't seen him since.

(Verilogue 10888, AmE young adult PT aged 18-34)

In mental health interviews, cognitive IDK tokens also occur in assessment sequences about PTs' symptoms, anxieties and behaviour patterns. In the example in (8), where the knowledge disavowing interpretation of IDK is supported by its modification with *even*, the PT uses IDK to communicate his lack of understanding with respect to concerning experiences: he does not know or understand whether his feelings are clinically relevant, and is unable to provide contextualizing information that may account for his mood swings. In this particular context, IDK simultaneously functions as an appeal for help and clarification.

(9) DR: What are the three words I asked you to remember.

(.)

PT: Bro::wn. (..) U:::h @

[...]

DR: That word has (.) same meaning as being truthful, telling the truth all the time. But
(it's) a different word.

(..)

PT: Um? >***I don't know*** that word<?

DR: You don't know that word?

PT: No.

DR: (Very good.) Honesty.

PT: Eh:

DR: == And eh the last one was not lip but it's tulip. (.) What is tulip?

(.)

PT: ***I don't know.***

DR: IT'S A FLOWER.

CG: Tulip's a flower.

PT: Well, ***I don't know.***

DR: == You never heard of a tulip flower?

PT: No.

DR: Oh, okay. Now you did. (.) He learned something.

(Verilogue 58130, AmE youth PT aged ≤ 17)

3.2 Epistemic marker

As predicted by previous research (e.g. Beach and Metzger, 1997; Tsui, 1991), IDK is not solely used in our data to signal a cognitive state. In both sociolinguistic and medical interviews, IDK regularly functions as an epistemic marker. In (10) and (11), the speakers follow/preface their propositions with IDK to qualify the reliability of their assessments and to mitigate their accountability for their accuracy. The attenuating effect of IDK is reinforced by the fall-rise intonation contour (example 10) and/or its co-occurrence with other hedging devices (*probably*, *or something like that* in (10); *maybe*, *about* in (11)). Its use in (11) warns the DR that his PT may in fact be abusing drugs/alcohol more regularly than he is prepared to admit.

(10) Keith: They were I-Irish. They were [probably] from up the road. They were

HP: [yeah]

Keith: = probably [gipsies] or something like that. ***I dunno?***

HP: [mhm]

(Sociolinguistic interviews, BwE young adult IVE aged 18-34)

(11) DR: Uh, how many days, (.) have you counted how many days you're clean and sober?

PT: ***I dunno.*** Maybe about (...) a month?

(Verilogue 97032, AmE young adult PT aged 18-34)

3.3 Mitigation device

Speakers sometimes exploit the uncertainty inherent in IDK to mitigate the potentially face-threatening effect of disagreements, as in (12), or to balance social asymmetries, as in (13). DR-

PT relationships are inherently asymmetrical due to ‘differences in status and domain specific knowledge’ (Tates and Meeuwesen, 2000, p. 152). By couching his indirect medication recommendation in uncertainty (*I don’t know if I should (.) we should bump it up*), the PT in (13) puts himself into a (discursively) subordinate and less knowledgeable position to that of the DR, thus avoiding to undermine the DR’s expertise and authority. (Also note the change in pronoun from *I* to *we* to involve the DR in the decision-making process.)

(12) 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.

(Sociolinguistic interviews, BwE young adult IVE aged 18-34)

(13) PT: So I’ve been off that that for probably a month. And when I was on it. I

DR: Okay.

PT: = didn’t really notice any (.) you know, it wasn’t really working as a like

DR: Okay.

PT: = it used to. So *I don’t know* if I should (.) if we should bump it up or,

DR: Okay.

PT: = or something like that because I think I (.) grew tolerant to the, the 10 mg.

And plus, I’m a bigger guy anyway, so I don’t know.

DR: Mhm.

(Verilogue 49648, AmE young adult PT aged 18-34)

3.4 Avoidance/resistance strategy

An IDK function that is not represented in our sociolinguistic interview data but that is very prominent in our (youth) mental health interview data is that highlighted in Hutchby (2002): IDK as an avoidance/resistance strategy. When the DR in (14) raises issues related to the PT's problematic home environment, the PT responds with IDK. As indicated by the DR's reaction (*I'll (..) probably talk to you in a minute. Without your father*), he interprets the PT's use of IDK as an avoidance device; he delays pursuing of the therapeutically relevant topic until the caregiver has left the room, at which point the PT willingly engages with the DR's topic (not all reproduced below). The PT's final turn in (14) confirms the DR's interpretation of his earlier IDK as an avoidance strategy: in the presence of his father, the PT was reluctant to engage with a topic that his mother had explicitly told him not to discuss with the DR.

- (14) DR: I mean, i- i- is there stuff going on at home that's causing you to not (.) talk to your parents, or (.) you not talking to them is (.) putting stress on them and then there's trouble at home. (.) Which, I dunno which way: (.)

PT: ***I dunno?*** @

DR: E:m. (.) I guess ***I'll (..) probably talk to you in a minute. Without your father***, okay, for a second. but

CG1: (?) [I came in to just to it was just (?)]

DR: [I know. I know. I know. I know.]

[...]

[CG1 has just left consultation room.]

DR: Um. (..) **What the hell is going on in your house?**

[...]

[PT and DR talk about CG2 who, according to PT, puts on a smiley façade on consultation visits.]

PT: **Yeah, she tells me not to say anything.**

(Verilogue 5233, AmE youth PT aged ≤ 17)

In (15), the PT makes a point of letting the DR know that she is not attending the consultation out of her own volition (*Except for when you take me out of school and take me here. Then I'm not as happy of a person.*). She reacts to the most of the DR's questions with minimal responses that she only rarely elaborates on (e.g. by providing an account for her non-answers): *no/nope*, *sure* and *IDK*. The systematic non-cooperation signalled by the PT's persistent use of these response devices frustrate the DR's attempts to elicit talk on all topics associated with therapeutic matters. Moreover, they serve as a face-saving strategy, for example when the PT admits her non-adherence to medication.

(15) DR: Can you tell me what was going on at the time?

PT: **>I dunno<?**

CG: She was depressed about school.

DR: Okay.

PT: == NO:.. (?)

DR: Were you depressed about school? [Or] was there something that was going

PT: [NO.]

DR: = on (*h*) e::h with you

PT: == I'm a very happy person.

DR: Okay.

CG: @ (?)

PT: I am. **Except for when you take me out of school and take me here.**

DR: Aha.

PT: = **Then I'm not as happy of a person.**

[...]

DR: And eh and how does that make you feel.

PT: ***I dono?*** I forget when I wake up. [@]

DR: [Mhm.] You forget your dreams when you
wake up. (.) [(Alright.)]

PT: [Not really] because I go right back to sleep. So (enough of that).

DR: Okay.

CG: Sometimes you don't get to sleep till late.

DR: Do you take any medications?

PT: >***I dono.***<

DR: Okay. E::h

PT: == I don't take, I refuse to.

DR: You refuse to take medication. But you have taken medications before.

PT: Sure.

DR: Okay. And uh (*h*) when you took it, uh, did they help you?

PT: Nope.

DR: Okay. And uh (.) did they give you any side effects?

(.)

PT: *I dunno.*

DR: You don't know.

CG: You'd know if they gave you side effects.

(Verilogue 9524, AmE youth PT aged ≤ 17)

Unlike cognitive or epistemic IDK tokens, avoidance/resistance tokens of IDK usually constitute the sole element of a turn. Moreover, they are not generally accompanied by the provision of a hedged response or an account for the declared knowledge gap.

3.5 Turn-exchange device

In addition to the interpersonal uses of IDK outlined above, IDK is also used in the textual domain to affect and prevent turn-transfer between participants. Beyond using IDK in turn-initial and -final position to launch and yield turns (see Pichler, 2013, p. 87-91 for examples), IVEs and PTs use IDK turn-medially to secure their hold on the floor while they are planning the continuation or re-formulation of their multi-unit turns.¹⁵ In the illustrative examples in (16) and (17), IDK is used to bridge the gap between aborted and continued, recast or reformulated utterances. As illustrated, these IDK tokens tend to co-occur with filled and unfilled pauses and other discourse-pragmatic features, particularly those signalling turn continuation (*and*, *but*).

¹⁵ Unlike Pichler (2013, p. 85, 89), we do not differentiate for the current investigation repair and turn-holding tokens of IDK whose interactional effect is essentially the same, i.e., to signal speakers' communicative presence while they plan the re-formulation or continuation of their turn.

- (16) Keith: I'd only been away once before. And that was with her. [And] eh (..) **I**
 HP: [mhm]
 Keith: = **dunno** (.) just like obviously I wanted to go Australia [[and that. I
 HP: [[mhm mhm]]
 Keith: = wanted]] to go away, like, for a year.

(Sociolinguistic interviews, BwE young adult IVE aged 18-34)

- (17) PT: Like, it was it was just like obvious cos my mind was just like (.) constantly
 like (..) **I don't know**, it was just like (..) I just couldn't like (.) fall asleep (for
 more than), **I dunno**. But like (..) um, you know, when it wa- when it wears
 off or whatever and, you know, I'm able to sleep, like (there's) other things
 kind of get in the way of me sleeping.

(Verilogue 20127, AmE young adult PT aged 18-34)

3.6 Topic-development device

Finally, IDK is used by IVEs and PTs in our data to affect the topic development of interviews. In (18), the IVE's use of IDK in response to the IVR's follow-up question serves to decline the proffered topic by disavowing access to it. The closure-implicative effect of IDK is attested in the IVR reaction which is to almost immediately move on to the next interview question. The PT's non-bolded, turn-final token of IDK in (8) in Section 3.1 illustrates this function in the mental health interview data. The PT struggles to explain his feelings and symptoms; he uses IDK to successfully terminate his attempt to do so (see the DR's immediate up-take of the floor).

(See also the non-bolded, turn-final token of IDK in (13) in Section 3.3.) The terminating effect of these IDK tokens is reinforced by their terminating intonation, i.e., a falling boundary tone.

(18) 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***

HP: It's hard to say.

Gabriel: Mhm.

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

(Sociolinguistic interviews, BwE young adult IVE aged 18-34)

3.7 Summary: IDK functions in sociolinguistic and mental health interviews

Table 1 compares the functional taxonomies developed from the analyses of IDK in the sociolinguistic and mental health interviews. To highlight cross-situational similarities and differences in IDK use, it classifies functional categories according to domains of use: referential, interpersonal (concerned with the expression of speaker attitudes and co-ordination of speaker-hearer relationships), and textual (relating to the development and organization of discourse) (see Brinton, 1996; Halliday, 1979). Thus, while IDK is used in both interview contexts to signal a cognitive state, hedge or mitigate propositions, prevent turn transfer, and affect topic development, the medical interview taxonomy features more interpersonal functions

than the sociolinguistic interview taxonomy. The presence in the medical interview data of the interpersonal avoidance/resistance category is due to situation-specific social roles and interactional demands. In both the medical and sociolinguistic interviews, IDK frequently occurs as the (near-)sole turn component in second pair parts of question-answer adjacency pairs. In the medical interviews, where the DR IVR occupies a more powerful role than the PT IVE (Pilnick and Dingwall, 2011; ten Have, 1991), such tokens tend to be used in response to counsellor questions on mental health-related issues, and serve to disrupt conversation related to assessment or treatment, i.e., issues with potentially deleterious implications for PTs' desired progression of the consultation as well as their healthcare management (see (14) above). As such, these tokens perform important face-saving functions, warranting categorization in the interpersonal domain. By contrast, in the sociolinguistic interviews, where the IVE has little personal stake in the interview progression, such tokens tend to occur in response to IVR questions about non-personal matters and function as role-congruent displays of the IVE's relative control over topic selection (see (17) above). As such, these tokens affect the topical development of interviews, warranting categorization in the textual domain. Thus, our analysis of IDK use in medical and sociolinguistic interview data demonstrates that IDK tokens in identical sequential contexts can have vastly differential interactional effects across situational contexts of use.

Table 1 IDK functions in sociolinguistic and mental health interviews

	sociolinguistic interviews	medical interviews
Referential	cognitive state claim	cognitive state claim
Interpersonal	epistemic marker	epistemic marker
	mitigation device	mitigation device

		avoidance/resistance strategy
Textual	turn-exchange device	turn-exchange device
	topic-development device	topic-development device

4. Quantitative analysis: IDK variation patterns

In this section, we examine the distribution of form, function and syntax for IDK across a range of contextual factors, and test our hypothesis that form-function correlations in IDK use are robust across situational contexts, varieties and social groups.

Table 2 shows the distribution of IDK variants, functions and syntactic configurations across the four sub-corpora studied. A comparison of variation patterns across the BwE sociolinguistic interviews and the BrE medical interviews (columns 3 and 4) allows us to assess the effect of situational context on IDK variability. The comparison confirms previous reports that the functionality of discourse-pragmatic variables is significantly affected by situation ($\chi^2=26.942$, $p<0.0001$): in sociolinguistic interviews, textual IDK uses outnumber referential and interpersonal uses; in the BrE (and AmE) *adult* medical interviews, referential tokens constitute the majority of IDK uses. At the same time, the comparison of the BwE sociolinguistic and BrE medical interviews reveals a previously undocumented effect of situational context on discourse-pragmatic variation, i.e., its effect on variant distribution: full variants are statistically less frequent in the sociolinguistic interviews than in the (BrE adult) medical interviews ($\chi^2=40.305$, $p<0.0001$). Finally, unlike form and function, the syntactic configuration of IDK as bound or unbound seems to be unaffected by situational context ($\chi^2=0.345$, $p=0.56$).

Table 2. Distribution of variants and functional categories across four sub-corpora

		BwE adult socioling. interviews (N=353)	BrE adult medical interviews (N=91)	AmE adult medical interviews (N=120)	AmE youth medical interviews (N=150)
Variant	full	24% (85)	60% (55)	53% (63)	51% (77)
	(semi-)reduced	76% (268)	40% (36)	47% (57)	49% (73)
	<i>significance (context)</i>	$\chi^2 = 40.305$, df = 1, p < 0.0001			
	<i>significance (variety)</i>	$\chi^2 = 1.3235$, df = 1, p = 0.25			
	<i>significance (age)</i>			$\chi^2 = 0.0363$, df = 1, p = 0.8488	
Function	referential	37% (131)	61% (55)	62% (74)	39% (59)
	interpersonal	21% (74)	26% (24)	14% (17)	45% (68)
	textual	42% (148)	13% (12)	24% (29)	16% (23)
	<i>significance (context)</i>	$\chi^2 = 26.942$, df = 2, p < 0.0001			
	<i>significance (variety)</i>	$\chi^2 = 7.1924$, df = 2, p = 0.0274			
	<i>significance (age)</i>			$\chi^2 = 30.021$, df = 2, p < 0.0001	
Syntax	unbound	58% (206)	55% (50)	53% (63)	70% (105)
	bound	42% (147)	45% (41)	47% (57)	30% (45)

	<i>significance</i> (<i>context</i>)	$\chi^2 = 0.345$, $df = 1$, $p = 0.557$	
	<i>significance</i> (<i>variety</i>)	$\chi^2 = 0.1244$, $df = 1$, $p = 0.7243$	
	<i>significance</i> (<i>age</i>)	$\chi^2 = 8.6857$, $df = 1$, $p = 0.0032$	

The contribution to IDK variation of situational context stands in stark contrast to that of variety, as gauged from a comparison of variation patterns in the BrE and AmE adult medical interviews (columns 4 and 5 in Table 2). In this particular interview context *and age group*, neither the form ($\chi^2=1.3235$, $p=0.25$) nor syntactic configuration ($\chi^2=0.1244$, $p=0.72$) of IDK are significantly affected by variety: in both varieties, full variants outnumber (semi-)reduced variants; and unbound and bound tokens occur with roughly similar frequency. Function differentiates BrE and AmE adult PTs ($\chi^2=7.1924$, $p=0.03$). However, in both varieties, referential IDK uses outnumber the interpersonal and textual IDK uses combined.

Finally, a comparison of our AmE adult and youth medical interviews (columns 5 and 6 in Table 2) confirms the results reported for BwE sociolinguistic interviews (Pichler 2013): age only minimally affects IDK form variation ($\chi^2=0.0363$, $p=0.85$) but significantly affects IDK function variation ($\chi^2=30.021$, $p<0.0001$). In both the adult and youth medical interviews, full and (semi-)reduced variants of IDK occur with roughly similar frequency. However, while referential IDK uses dominate (BrE and AmE) adult medical interviews, interpersonal IDK use accounts for almost half of all IDK tokens in the youth medical interviews. The syntactic configuration of IDK also varies across age in medical interviews ($\chi^2=8.6857$, $p=0.0032$):

possibly because of the strong association of bound tokens with referential uses, bound IDK is significantly more frequent in the adult than the youth medical data. In sum, then, situational context significantly affects both the form and function of IDK use; variety minimally affects its (non-cognitive) functional distribution; age significantly affects IDK function and, to a lesser extent, IDK syntax.

To establish any form-function correlations in our data and to uncover the configuration of factors contributing to variant choice, we next conduct four independent, sample-specific logistic regression analyses of the contribution in each sub-corpus of function, age and speaker sex to the occurrence of the full variant *I don't know* (see Section 2.3). Table 3 shows the results. As indicated by the range values for significant predictors, function consistently makes the most important contribution to the occurrence of *I don't know*. Across the sub-corpora, the full variant is strongly favoured for referential IDK uses, and disfavoured for interpersonal and textual uses. Thus, although IDK variants and functional domains are differentially distributed across contextual factors (see Table 2), our hypothesis that form-function correlations in IDK use are robust across situational context, variety and age is confirmed. With the exception of age in the BrE adult medical interview data, social factors do not make a statistically significant contribution to variant choice. Note, however, that the full variant *I don't know* is non-significantly trending towards an association with adults and females (as opposed to young adults and males).

Table 3. Contribution of internal and external predictors to the probability of *I don't know* (with (semi-)reduced tokens, *I dono* and *I dunno*, as non-application values) [LO = log-odds; FW = factor weights]

<i>I don't know</i>																
BwE adult sociolinguistic interviews					BrE adult medical interviews				AmE adult medical interviews				AmE youth medical interviews			
input prob.	.158				.354				.495				.510			
total N	353				91				120				150			
deviance	300.845				95.4				142.541				181.686			
	LO	FW	%	N	LO	FW	%	N	LO	FW	%	N	LO	FW	%	N
function	p = 7.11e-10				p = 0.00222				p = 0.02				p = 1.21e-05			
referential	1.351	.794	45.0	131	1.244	.776	78.2	55	0.938	.719	60.8	74	1.145	.759	76.3	59
interspers.	-0.748	.321	10.1	74	-0.515	.374	33.3	24	-0.323	.420	41.2	17	-0.745	.322	33.8	68
textual	-0.603	.354	13.5	148	-0.729	.325	33.3	12	-0.614	.351	37.9	29	-0.400	.401	39.1	23
range	.473				.451				.368				.437			
age	Non-significant				p = 0.0233				Non-significant				Non-significant continuous			
adult	0.378	[.593]	30.1	196	0.926	.716	68.9	74	0.474	[.616]	50.0	60	(9-17 y.o.)			
yg. adult	-0.378	[.407]	15.9	157	-0.926	.284	23.5	17	-0.474	[.384]	55.0	60				
range					.432											
sex	Non-significant				Non-significant				Non-significant				Non-significant			
female	0.074	[.518]	26.5	166	0.306	[.576]	77.4	31	0.493	[.621]	64.3	56	0.111	[.528]	51.3	76
male	-0.074	[.482]	21.4	187	-0.306	[.424]	51.7	60	-0.493	[.379]	42.2	64	-0.111	[.472]	51.4	74

<i>range</i>				
speaker	random effect	random effect	random effect	random effect

5. Discussion

The aim of our analysis was to test the effect of situational context, variety and age on IDK variability while also exploring a new prospective application for variationist linguistic research. To this end, we conducted principled and accountable analyses of the construction's usage and distribution across four sub-corpora differentiated along relevant parameters. Our analysis produced important new insights into the nature and extent of discourse-pragmatic variation which we will summarize and discuss in Section 5.1. Moreover, we will suggest in Section 5.2 ways in which the results of our proof-of-concept analysis, if further developed in collaboration with mental health practitioners, might benefit DR-PT communication.

5.1 Theoretical and methodological implications

Our auditory, discourse-analytic and structural analyses of the IDK data reported above distinguish situationally labile patterns in IDK variation from those that are robust across contexts. They thus substantially clarify the IDK literature and set a new standard for multi-dimensional discourse-pragmatic variation analysis. Our analyses revealed that the morpho-phonological, functional and syntactic variability reported to characterize IDK use in other situational contexts (e.g. media, courtroom and sociolinguistic interviews, casual conversations; see Beach and Metzger, 1997; Bybee and Scheibman, 1999; Pichler, 2009, 2013; Potter, 1996; Scheibman, 2000) extends to both adult and youth mental health interviews. However, comparison of IDK use in the sociolinguistic and medical interview data also showed that despite considerable overlap, the functionality of IDK is not identical across interview contexts, confirming previous reports that the strategic use to which discourse-pragmatic variables are put

is often highly context-specific (see e.g. Fuller, 2003; Lam, 2009; Schleef, 2008b; Verdonik et al., 2009). The lack of complete functional overlap in IDK use across our data was shown to be due to the existence in the medical data of an interpersonal function not found in the sociolinguistic interviews: the signalling of avoidance/resistance. Thorough characterization of this function is important for the development of a comprehensive taxonomy of IDK use across situational contexts as well as potential efforts to translate these findings into clinical practice. Moreover, it has important methodological implications. On the surface, the avoidance/resistance tokens of IDK in the medical interviews (see example (14) above where the PT uses IDK to frustrate the DR's line of questioning) resemble very closely the topic-declining tokens in the sociolinguistic interviews (see example (18) above where the IVE uses IDK to decline the IVR's proffered topic): they share the same sequential positioning in second pair parts of question-answer adjacency pairs, tend to constitute the (near-)sole constituent of a turn, and are generally phonetically reduced and produced with level or falling intonation. However, owing to our in-depth understanding of both datasets, we were able to establish important differences in PTs' and IVEs' seemingly uncooperative response behaviour that led us to characterize PTs' use as interpersonal and IVEs' as textual: PTs used these IDK tokens in an attempt to save face; IVEs used them in an attempt to exert some control over the topical development of the interview. Thus, our analysis cautions against the uncritical application of coding and quantification schemas developed for analyses of one situational context to analyses of data collected in another context.

The multivariate analyses revealed consistent form-function correlations in IDK use. The strong association across our sub-corpora of full variants with referential uses and of (semi-)reduced variants with interpersonal and textual uses is in line with Hopper's (1991, p. 24)

Principle of Divergence whereby a construction becomes grammaticalized, i.e., undergoes morpho-phonological and semantic-pragmatic changes, in some environments but not others (see further Bybee and Scheibman, 1999; Pichler, 2013, p. 118–120). Our results suggest that once variant forms have thus become imbued with specific meanings, these associations are robust, at least across the parameters tested here: situation, variety, age. Pending further confirmation, we tentatively suggest that stable grammaticalization-induced form-function correlations identified in one dataset are likely generalizable to new datasets representing otherwise specialized speech settings.

The generalizability of discourse-pragmatic variation patterns does not apply to all dimensions of IDK use, however. Our distributional analysis established that in contrast to the robust form-function correlations discussed above, the distribution of variant forms, functional categories and syntactic configurations is not uniform across all the parameters differentiating our sub-corpora. This finding, i.e., that some aspects of IDK variation (i.e. form-function correlations) are robust across situational contexts while others (i.e. form, function and syntax distribution) are not, highlights the importance of analysing multiple dimensions of a single variable's use. This is particularly important if our aim is to yield accurate accounts of linguistic variation and improve our understanding of the nature of situation-conditioned discourse-pragmatic variation (see also Wagner et al. 2016).

Finally, inspection of the variation patterns in our adult sociolinguistic and medical interview data reveals that the form and function distribution of IDK varies across but is consistent within situational context: in the BwE adult sociolinguistic interviews, IDK is used predominantly in its (semi-)reduced form for non-referential uses; across the BrE and AmE *adult* mental health interviews, by contrast, IDK is predominantly used in its full form for referential

uses. It may be tempting to deduce from these patterns that IDK form and functionality are markers of situational context (see Traugott, 2001). Yet this hypothesis is refuted by the significantly different patterns of IDK use in the youth medical interview data. Here, IDK has a different functional profile than in the adult medical data, being used predominantly for non-referential (in particular interpersonal) functions. These results emphasize the importance of including social factors in analyses of situationally-conditioned language use.

Following Giles's (2001) and Traugott's (2001) proposals that speakers' subjective attitudes to and emotional assessments of situational contexts may affect their language use, and based on our in-depth understanding of our data, we tentatively suggest that adult and youth PTs' differential use of IDK in mental health consultations may reflect their differential construal of the speech situation. In our data, adult PTs tended to conceive of the consultation as an opportunity to learn more about their condition by detailing to the consultant what it is they do not know or understand about their symptoms and behaviours (see example (7) above). This is what may have triggered their high rate of cognitive IDK uses. Adolescents and children in our mental health interview data, by contrast, tended to conceive of the consultation as an unwelcome encounter that threatened their positive and negative face needs (see example (14) above). This is what may have triggered their high rate of avoidance/resistance IDK tokens. In short, we would like to posit that adults and youths may harbour differential subjective attitudes and assessments of mental health interviews, as reflected in their differential language use. This interpretation is supported by prior work suggesting that youths must be socialized into the interactional frame of medical consultations, a process that is largely encouraged by supportive parent and provider communicative practices (Tates et al., 2002). Future cross-situational analyses of discourse-pragmatic variation will therefore need to take account not just of social

variation patterns but also of the potentially non-homogeneous nature of speech produced in a given situational context.

5.2 Practical applications

In addition to clarifying the robustness of IDK use within and across interview settings, our analysis of IDK *in mental health interviews* was designed as a proof-of-concept analysis to explore the translational potential of studying discourse-pragmatic variation in psychiatrist-patient interaction. Because ‘research on professional-patient communication in the treatment of mental illness is in its infancy’ (Hassan et al., 2007, p. 150), our analysis is intentionally focused on a feature with sufficient prior linguistic investigation. This allow for a thorough and accountable analysis, despite the admittedly modest effect of a single construction on mental health consultations.¹⁶ Notwithstanding this limitation, our analysis demonstrates the clinical potential of describing site-specific patterns of language variation. If the insights described above were deemed relevant and applicable by practitioners, cross-disciplinary collaboration could translate them into awareness-raising probabilistic schemes and practical tools for in-office use (e.g., reference cards and conversation aids). Consider, for example our qualitative characterization and quantitative account of the cognitive and resistive categories; these in particular might offer DRs learnable cues to further develop their repertoire of interactional skills (Hesson and Pichler, MS).

Focusing first on cognitive uses of IDK, the fact that a PT cannot explain or does not understand their behaviour or symptoms, as expressed by PT use of IDK in (8) above, may be

¹⁶ To broaden the scope of the impact pathway piloted in the current work, subsequent extensions would need to replicate our variationist approach in a larger and more socially diverse dataset while also expanding the analysis to include additional linguistic and, potentially, non-linguistic devices used by PTs to signal e.g. insufficient knowledge (*I wouldn't know that, I've no idea, I haven't a clue*, etc.) or resistance (*yup, sure, whatever*, etc.).

therapeutically relevant to the DR's efforts to providing consistent relief from mood disorder symptomatology for which the DR needs to have a precise understanding of the PT's daily experiences. It is therefore important that the DR probes the PT on the quantity and duration of his behaviour and symptoms to determine whether his current treatment regimen is providing adequate symptom control. Moreover, when PTs use IDK because they sincerely do not know the answer to a question, as in (9) above, further probing may challenge the PT's competence and (especially among children) may create feelings of distrust or inadequacy that are unproductive to PTs' care and damaging to the DR-PT relationship. PTs are uniquely positioned to provide information about their symptoms and anxieties. Embarrassing PTs by highlighting their lack of knowledge may amplify the interactional face-threats posed by diagnostic evaluations for mental health disorders. In turn, interactional insensitivity may stimulate non-cooperation, a situation that could adversely affect PTs' quality of care. To this end, our finding that cognitive uses of IDK are probabilistically realized as *I don't know* (rather than *I dono* or *I dunno*) might act as an important signpost for cognitive IDK interpretation.

When PTs use IDK to avoid discussion of issues that contribute to an unstable or unhealthy psychosocial environment, as in (14) above, additional exploration of these may be ultimately necessary for PTs' psychosocial care. Furthermore, the realization that a PT strategically uses IDK (and other devices) to resist the incitement to talk about therapeutic matters in an attempt to avoid diagnosis and treatment (see example (15) above) potentially provides an opportunity for the DR to move to more effective interaction formats or to explicitly address the fact that the PT may be resisting questioning. Our analysis assists in these attempts by highlighting the fact that resistive uses are typically realized as *I dono* or *I dunno* (rather than *I don't know*), again offering interpretative guidance for DRs. Furthermore, our finding that the

functional profile of IDK varies across adult and youth mental health PTs could be applied by DRs: with young PTs in particular, DRs might attend to the fact that IDK (along with other linguistic and non-linguistic cues) is regularly used to frustrate questions on therapeutically relevant topics.

In sum, our analysis lays the foundation for the collaborative development of guidelines for the correct interpretation of a linguistic construction previously described as therapeutically important in mental health (see Hutchby 2002), a fundamentally interview-driven speciality (McCabe et al., 2002). Specifically, it offers: (i) a clinically relevant introduction to the multifunctionality of IDK in medical interviews; (ii) relatively objective linguistic indicators (i.e., form-function correlations) to augment DRs' native speaker intuitions and assist them with the difficult but therapeutically important task of probabilistically judging in real time which meaning a PT's IDK use signals.

6. Conclusion

The preceding variationist analysis of IDK in sociolinguistic and mental health interviews has enhanced discipline-specific knowledge about the nature of discourse-pragmatic variation, in particular about the effect of situational context, variety and age on IDK variability. At the same time, our extension of discourse variation analysis beyond the sociolinguistic interview has yielded findings of possible practical relevance to those regularly participating in mental health interviews: healthcare providers. Our study thus demonstrates the theoretical value and applicational potential of extending variationist analyses beyond sociolinguistic interview data.

The number of mental health condition diagnoses is increasing rapidly in the USA and UK (Collishaw et al., 2010; Olfson et al., 2014). Unlike strictly somatic conditions, such as

diabetes or cancer, there are no laboratory tests or imaging studies that aid in the diagnosis of mental health disorders. Thus, correct diagnosis and successful treatment of these conditions are crucially dependent on DRs' ability to develop effective, mutually intelligible dialogues with their PTs. Our analysis of IDK in mental health consultations has provided new insights into PTs' discursive and linguistic practices in this speech situation. If explored further with the help of medical practitioners (and if extended to a larger and more diverse dataset; see footnote 15), our linguistically grounded understanding of site-particular communicative practices and language variation patterns may improve diagnostic practice and therapeutic success. Analyses of sociolinguistic interviews have the potential to provide important hypotheses about the application potential of variationist research. However, our analysis of IDK variability across sociolinguistic and medical interviews suggests that truly successful impact work is crucially contingent on our commitment to analyse the very discourse produced by the anticipated beneficiaries of our research. With this in mind, we call for more variationist analyses of speech data produced in medical contexts and for closer cross-disciplinary collaboration between variationist linguists and medical practitioners.

Acknowledgments

We gratefully acknowledge the generous financial assistance of the *Newcastle Institute for Social Renewal*, *Spectrum Health* and MSU graduate school. Earlier versions of this paper were presented at DiPVaC 2 (Newcastle University, UK, April 2014) and NWAV 43 (Universities of Illinois, US, October 2014), and we thank the audience members for their insightful comments and questions. The constructive comments provided by the anonymous reviewers and the editor,

John Joseph, are also greatly appreciated. Finally, we would like to thank Jenny Cheshire and Suzanne Evans Wagner for their encouragement and support.

References

- Andersen, Gisle (2001). *Pragmatic markers and sociolinguistic variation: A relevance-theoretic approach to the language of adolescents*. Amsterdam: John Benjamins.
- Atkinson, Maxwell J., and John Heritage (1984). *Structures of social action: Studies in conversation analysis*. Cambridge: Cambridge University Press.
- Beach, Wayne A., and Terri R. Metzger (1997). Claiming insufficient knowledge. *Human Communication Research* 23:562–88.
- Becker, Kara (2013). The sociolinguistic interview. In Christine Mallinson, Becky Childs and Gerard van Herk (eds.), *Data collection in sociolinguistics: Methods and applications*, 91–100. London: Routledge.
- Bergmann, J. R. (1992). Veiled morality: Notes on discretion in psychiatry. In Paul Drew and John Heritage (eds.), *Talk at work: Interaction in institutional settings*, 137–62. Cambridge: Cambridge University Press.
- Brinton, Laurel J. (1996). *Pragmatic markers in English: Grammaticalization and discourse function*. Berlin: Mouton de Gruyter.
- Bybee, Joan, and Joanne Scheibman (1999). The effect of usage on degrees of constituency: The reduction of *don't* in English. *Linguistics* 37:575–96.
- Chaemsaihong, Krisda (2012). Performing self on the witness stand: Stance and relational work in expert witness testimony. *Discourse and Society* 23:465–86.
- Cameron, Deborah, Fiona McAlinden, and Kathy O'Leary (1988). Lakoff in context: The social and linguistic functions of tag questions. In Jennifer Coates and Deborah Cameron (eds.), *Women in their speech communities*, 74–93. London: Longman.

- Cheshire, Jenny (1981). Variation in the use of *ain't* in an urban British English dialect. *Language in Society* 10:365–81.
- Cheshire, Jenny (1982). *Variation in an English dialect: A sociolinguistic study*. Cambridge: Cambridge University Press.
- Cheshire, Jenny (2007). Grammaticalization and stuff like that. *Journal of Sociolinguistics* 11:155–93.
- Collishaw, Stephan, Barbara Maughan, Lucy Natarajan, and Andrew Pickles (2010). Trends in adolescent emotional problems in England: A comparison of two national cohorts twenty years apart. *Journal of Child Psychology and Psychiatry* 51:885–94.
- Coupland, Nikolas (1980). Style-shifting in a Cardiff work-setting. *Language in Society* 9:1–12.
- D'Arcy, Alexandra (2005). *Like: Syntax and development*. Unpublished PhD thesis, University of Toronto, Canada.
- Drager, Katie (2016). Constructing style: phonetic variation in quotative and discourse particle *like*. In Heike Pichler (ed.), *Discourse-pragmatic variation and change in English: New methods and insights*, 231–52. Cambridge: Cambridge University Press.
- Erman, Britt (1992). Female and male usage of pragmatic expressions in same-gender and mixed-gender interaction. *Language Variation and Change* 4:217–34.
- Erman, Britt (2001). Pragmatic markers revisited with a focus on *you know* in adult and adolescent talk. *Journal of Pragmatics* 32:1337–59.
- Escalera, Elena Andrea (2009). Gender differences in children's use of discourse markers: Separate worlds or different contexts? *Journal of Pragmatics* 41:2479–95.
- Ferrara, Kathleen W. (1997). Form and function of the discourse marker *anyway*: Implications for discourse analysis. *Linguistics* 35:343–78.
- Finegan, Edward, and Douglas Biber (1994). Register and social dialect variation: An integrated approach. In Douglas Biber and Edward Finegan (eds.), *Sociolinguistic perspectives on register*, 315–47. Cambridge: Cambridge University Press.

- Finegan, Edward, and Douglas Biber (2001). Register variation and social dialect variation: The register axiom. In Penelope Eckert and John R. Rickford (eds.), *Style and sociolinguistic variation*, 235–67. Cambridge: Cambridge University Press.
- Frankel, Richard M. (2002). The sociolinguistic turn in physician-patient communication research. In James E. Alatis, Heidi E. Hamilton and Ai-Hui Tan (eds.), *Georgetown university round table on languages and linguistics 2000. Linguistics, language, and the professions: Education, journalism, law, medicine, and technology*, 81–103. Washington, DC: Georgetown University Press.
- Freed, Alice F., and Alice Greenwood (1996). Women, men, and type of talk: What makes the difference. *Language in Society* 25:1–26.
- Fuchs, Robert, and Ulrike Gut (2016). Register variation in intensifier usage across Asian Englishes. In Heike Pichler (ed.), *Discourse-pragmatic variation and change in English: New methods and insights*, 185–210. Cambridge: Cambridge University Press.
- Fuller, Janet M. (2003). The influence of speaker roles on discourse marker use. *Journal of Pragmatics* 35:23–45.
- Giles, Howard (2001). Couplandia and beyond. In Penelope Eckert and John R. Rickford (eds.), *Style and sociolinguistic variation*, 211–19. Cambridge: Cambridge University Press.
- Guy, Gregory R. (1980). Variation in the group and the individual: the case of final stop deletion. In William Labov (ed.), *Locating language in time and space*, 1–36. New York: Academic Press.
- Halliday, M. A. K. (1979). Modes of meaning and modes of expression: types of grammatical structures and their determination by different semantic functions. In D.J. Allerton, Edward Carney and David Holdcroft (eds.), *Function and context in linguistic analysis: A festschrift for William Haas*, 57–79. Cambridge: Cambridge University Press.
- Hassan, Imren, Rosemarie McCabe, and Stefan Priebe (2007). Professional–patient communication in the treatment of mental illness: A review. *Communication and Medicine* 4:141–52.

- Have, ten Paul. 1991. Talk and institution: a reconsideration of the 'asymmetry' of doctor-patient interaction. In Deirde Boden and Don Zimmerman (eds.), *Talk and social structure*, 138–63. Berkeley, CA: University of California Press.
- Heritage, John, and Douglas W. Maynard (2006). Problems and prospects in the study of physician-patient interaction: 30 years of research. *Annual Review of Sociology* 32:351–74.
- Hesson, Ashley (2014). Medically speaking: Mandative adjective extraposition in physician speech. *Journal of Sociolinguistics* 18:289–318.
- Hesson, Ashley, and Heike Pichler (MS). Breaking down barriers: Interpreting patients' use of I DON'T KNOW in pediatric mental health consultations. Unpublished manuscript, Michigan State University, USA.
- Holmes, Janet (1995). *Women, men and politeness*. London: Longman.
- Holmes, Janet, and Stephanie Schnurr (2006). Doing femininity at work: more than just relational practice. *Journal of Sociolinguistics* 10:31–51.
- Hopper, Paul J. (1991). On some principles of grammaticization. In Elizabeth C. Traugott and Bernd Heine (eds.), *Approaches to grammaticalization: Focus on theoretical and methodological issues*. Vol. 1, 17–35. Amsterdam: John Benjamins.
- Hopper, Paul J., and Elizabeth C. Traugott (2003). *Grammaticalization*. 2nd ed. Cambridge: Cambridge University Press.
- Hugh, Mary L. (2012). Interrater reliability: The kappa statistic. *Biochemia Medica* 22:276–82.
- Hutchby, Ian (2002). Resisting the incitement to talk in child counselling: Aspects of the utterance 'I don't know'. *Discourse Studies* 4:147–68.
- Johnson, Daniel Ezra (2009). Getting off the GoldVarb standard: introducing Rbrul for mixed-effects variable rule analysis. *Language and Linguistics Compass* 31:359–83.
- Jucker, Andreas H., and Sara W. Smith (1998). And people just you know like 'wow': discourse markers as negotiating strategies. In Andreas H. Jucker and Yael Ziv (eds.), *Discourse markers: Descriptions and theory*, 171–201. Amsterdam: John Benjamin.

- Kozloff, Jeff, and Jamison Barnett (2006). Verilogue corpus of patient-provider interaction. Verilogue, Inc. Last accessed 21 November 2015 at: www.verilogue.com.
- Kyratzis, Amy, and Susan Ervin-Tripp (1999). The development of discourse markers in peer interaction. *Journal of Pragmatics* 31:1321–38.
- Labov, William (1966). *The social stratification of English in New York City*. Washington, DC: Center for Applied Linguistics.
- Labov, William (1972). *Sociolinguistic patterns*. Pennsylvania, PA: University of Pennsylvania Press.
- Labov, William (1984). Field methods of the project on linguistic change and variation. In J. Baugh and J. Sherzer (eds.), *Language in use: Readings in sociolinguistics*, 28–66. Englewood Cliffs, NJ: Prentice Hall.
- Labov, William (2001). *Principles of linguistic change. Vol. 2: Social factors*. Oxford: Blackwell.
- Lam, Phoenix W. Y. (2009). The effect of text type on the use of *so* as a discourse particle. *Discourse Studies* 11:353–72.
- Levon, Erez (2010). Organizing and processing your data: The nuts and bolts of quantitative analyses. In L. Litosseliti (ed.), *Research methods in linguistics*, 68–92. London: Continuum.
- Llamas, Carmen (2007). A new methodology: Data elicitation for regional and social language variation studies. *York Working Papers in Linguistics* 8:138–63.
- Macaulay, Ronald K. S. (2005). *Talk that counts: Age, gender, and social class differences in discourse*. Oxford: Oxford University Press.
- Mauranen, Anna (2004). ‘They’re a little bit different’...: Observations on hedges in academic talk. In Karin Aijmer (ed.), *Patterns in spoken and written corpora*, 173–97. Amsterdam: John Benjamins.
- McCabe, Rosemarie, Christian Heath, Tom Burns, and Stefan Priebe (2002). Engagement of patients with psychosis in the consultation: Conversation analytic study. *British Medical Journal* 325:1148–51.
- McWhinney, Ian (1989). The need for a transformed clinical method. In Moira A. Stewart and Debra Roter (eds.), *Communicating with medical patients*, 25–40. Newbury Park, CA: Sage.

- Moore, Emma, and Robert J. Podesva (2009). Style, indexicality and the social meaning of tag questions. *Language in Society* 38:447–85.
- Olfson, Mark, Carlos Blanco, Shuai Wang, Gonzalo Laja, and Christopher U. Correll (2014). National trends in the mental health care of children, adolescents, and adults by office-based physicians. *JAMA Psychiatry* 71:81–90.
- Ong, Lucille ML, Johanna CJM De Haes, Alaysia M. Hoos, and Frits B. Lammes. (1995). Doctor-patient communication: A review of the literature. *Social Science and Medicine* 407:903–18.
- Pichler, Heike (2009). The functional and social reality of discourse variants in a northern English dialect: I DON'T KNOW and I DON'T THINK compared. *Intercultural Pragmatics* 6:561–96.
- Pichler, Heike (2010). Methods in discourse variation analysis: Reflections on the way forward. *Journal of Sociolinguistics* 14:581–608.
- Pichler, Heike (2013). *The structure of discourse-pragmatic variation*. Amsterdam: John Benjamins.
- Pichler, Heike (2016). Uncovering discourse-pragmatic innovations: *innit* in Multicultural London English. In Heike Pichler (ed.), *Discourse-pragmatic variation and change in English: New methods and insights*, 59-85. Cambridge: Cambridge University Press.
- Pilnick, Alison, and Robert Dingwall (2011). On the remarkable persistence of asymmetry in doctor/patient interaction: A critical review. *Social Science and Medicine* 72:1374–82.
- Podesva, Robert J. (2007). Phonation type as a stylistic variable: The use of falsetto in constructing a persona. *Journal of Sociolinguistics* 11:478–504.
- Potter, Jonathan (1996). *Representing reality: Discourse, rhetoric and social construction*. London: Sage.
- Redeker, Gisela (1990). Ideational and pragmatic markers of discourse structure. *Journal of Pragmatics* 14:367–81.
- Schegloff, Emanuel A. (2007). *Sequence organization in interaction: A primer in conversation analysis*. Vol. 1. Cambridge: Cambridge University Press.
- Scheibman, Joanne (2000). *I dunno*: A usage-based account of the phonological reduction of *don't* in American English conversation. *Journal of Pragmatics* 32:105–24.

- Schleef, Erik. (2008a). Gender and academic discourse: Global restrictions and local possibilities. *Language in Society* 37:515–38.
- Schleef, Erik (2008b). The ‘lecturer’s ok’ revisited: Changing discourse conventions and the influence of academic division. *American Speech* 83:62–84.
- Stubbe, Maria, and Janet Holmes (1995). *You know, eh* and other exasperating expressions: An analysis of social and stylistic variation in the use of pragmatic particles in a sample of New Zealand English. *Language and Communication* 15:63–88.
- Tagliamonte, Sali and Alexandra D’Arcy. 2009. Peaks beyond phonology: Adolescence, incrementation, and language change. *Language* 85:58–108.
- Tates, Kiek, and Ludwien Meeuwesen (2000). ‘Let Mum have her say’: Turntaking in doctor-parent-child communication. *Patient Education and Counselling* 40:151–62.
- Tates, Kiek, Ludwien Meeuwesen, Ed Elbers, and Jozien Bensing. (2002). ‘I’ve come for his throat’: Roles and identities in doctor–parent–child communication. *Child: Care, Health and Development* 28:109–16.
- Tay, Dennis (2011). Discourse markers as metaphor signalling devices in psychotherapeutic talk. *Language and Communication* 31:310–7.
- Traugott, Elizabeth Closs (2001). Zeroing in on multifunctionality and style. In Penelope Eckert and John R. Rickford (eds.), *Style and sociolinguistic variation*, 127–36. Cambridge: Cambridge University Press.
- Traugott, Elizabeth Closs, and Graeme Trousdale (2010). Gradience, gradualness and grammaticalization: How do they intersect? In Elizabeth Closs Traugott and Graeme Trousdale (eds.), *Gradience, gradualness and grammaticalization*, 19–44. Amsterdam: John Benjamins.
- Tsui, Amy B. M. (1991). The pragmatic functions of *I don’t know*. *Text* 11:607–22.
- Verdonik, Darinka, Andrej Žgank, and Agnes Pisanski Peterlin (2009). The impact of context on discourse marker use in two conversational genres. *Discourse Studies* 10:759–75.

- Wagner, Suzanne Evans, Ashley Hesson, Kali Bybel, and Heidi Little (2015). Quantifying the referential function of general extenders in North American English. *Language in Society* 44:705–31.
- Wagner, Suzanne Evans, Ashley Hesson, and Heidi Little (2016). The use of referential general extenders across registers. In Heike Pichler (ed.), *Discourse-pragmatic variation and change in English: New methods and insights*, 211–31. Cambridge: Cambridge University Press.
- Waters, Cathleen (2013). Transatlantic variation in English adverb placement. *Language Variation and Change* 25:179–200.
- Weatherall, Ann (2011). *I don't know* as a prepositioned epistemic hedge. *Research on Language and Social Interaction* 44:317–37.
- Wooffitt, Robin (2005). *Conversation analysis and discourse analysis: A comparative and critical introduction*. London: Sage.

Vitae

HEIKE PICHLER is Lecturer in Sociolinguistics at Newcastle University, UK. Her main research interests are discourse-pragmatic variation and change, variationist methods and grammaticalisation. She is author of *The Structure of Discourse-Pragmatic Variation* (2013) and editor of *Discourse-Pragmatic Variation and Change in English* (2016).

ASHLEY HESSON is an MD candidate at Michigan State University, USA, with a PhD in linguistics. She studies perception and production of discourse-pragmatic variation in physician-patient interaction. Her work is featured in *American Speech*, *Journal of Sociolinguistics*, *Patient Education and Counseling* and *Journal of Language and Social Psychology*.

Table 1 IDK functions in sociolinguistic and mental health interviews

	sociolinguistic interviews	medical interviews
Referential	cognitive state claim	cognitive state claim
Interpersonal	epistemic marker	epistemic marker
	mitigation device	mitigation device
		avoidance/resistance strategy
Textual	turn-exchange device	turn-exchange device
	topic-development device	topic-development device

Table 2. Distribution of variants and functional categories across four sub-corpora

		BwE adult socioling. interviews (N=353)	BrE adult medical interviews (N=91)	AmE adult medical interviews (N=120)	AmE youth medical interviews (N=150)
Variant	full	24% (85)	60% (55)	53% (63)	51% (77)
	(semi-)reduced	76% (268)	40% (36)	47% (57)	49% (73)
	<i>significance (context)</i>	$\chi^2 = 40.305$, df = 1, p < 0.0001			
	<i>significance (variety)</i>	$\chi^2 = 1.3235$, df = 1 , p = 0.25			
	<i>significance (age)</i>	$\chi^2 = 0.0363$, df = 1 , p = 0.8488			
Function	referential	37% (131)	61% (55)	62% (74)	39% (59)
	interpersonal	21% (74)	26% (24)	14% (17)	45% (68)
	textual	42% (148)	13% (12)	24% (29)	16% (23)
	<i>significance (context)</i>	$\chi^2 = 26.942$, df = 2, p < 0.0001			
	<i>significance (variety)</i>	$\chi^2 = 7.1924$, df = 2, p = 0.0274			
	<i>significance (age)</i>	$\chi^2 = 30.021$, df = 2, p < 0.0001			
Syntax	unbound	58% (206)	55% (50)	53% (63)	70% (105)
	bound	42% (147)	45% (41)	47% (57)	30% (45)
	<i>significance (context)</i>	$\chi^2 = 0.345$, df = 1 , p = 0.557			
	<i>significance (variety)</i>	$\chi^2 = 0.1244$, df = 1 , p = 0.7243			
	<i>significance (age)</i>	$\chi^2 = 8.6857$, df = 1, p = 0.0032			

Table 3. Contribution of internal and external predictors to the probability of *I don't know* (with (semi-)reduced tokens, *I dono* and *I dunno*, as non-application values) [LO = log-odds; FW = factor weights]

<i>I don't know</i>																
BwE adult sociolinguistic interviews					BrE adult medical interviews				AmE adult medical interviews				AmE youth medical interviews			
input prob.	.158				.354				.495				.510			
total N	353				91				120				150			
deviance	300.845				95.4				142.541				181.686			
	LO	FW	%	N	LO	FW	%	N	LO	FW	%	N	LO	FW	%	N
function		p = 7.11e-10				p = 0.00222				p = 0.02				p = 1.21e-05		
referential	1.351	.794	45.0	131	1.244	.776	78.2	55	0.938	.719	60.8	74	1.145	.759	76.3	59
interspers.	-0.748	.321	10.1	74	-0.515	.374	33.3	24	-0.323	.420	41.2	17	-0.745	.322	33.8	68
textual	-0.603	.354	13.5	148	-0.729	.325	33.3	12	-0.614	.351	37.9	29	-0.400	.401	39.1	23
range		.473				.451				.368				.437		
age		Non-significant				p = 0.0233				Non-significant			Non-significant continuous (9-17 y.o.)			
adult	0.378	[.593]	30.1	196	0.926	.716	68.9	74	0.474	[.616]	50.0	60				
yg. adult	-0.378	[.407]	15.9	157	-0.926	.284	23.5	17	-0.474	[.384]	55.0	60				
range						.432										
sex		Non-significant				Non-significant				Non-significant			Non-significant			
female	0.074	[.518]	26.5	166	0.306	[.576]	77.4	31	0.493	[.621]	64.3	56	0.111	[.528]	51.3	76
male	-0.074	[.482]	21.4	187	-0.306	[.424]	51.7	60	-0.493	[.379]	42.2	64	-0.111	[.472]	51.4	74
range																
speaker	random effect				random effect				random effect				random effect			