In *Journal of Sociolinguistics* 9: 479-507 (2005)

Syntactic variation and beyond: gender and social class variation in the use of discourse-new markers[1]

Jenny Cheshire
Queen Mary, University of London

Syntactic variation and beyond: gender and social class variation in the use of discourse-new markers

ABSTRACT

The paper explores the relations between syntactic variation and the large-scale social dimensions of gender and social class. It argues on the basis of an analysis of the marking of discourse-new entities in interview speech that syntactic variants may frequently be involved in sociolinguistic variation, but indirectly, as just one of a broad set of choices that includes forms drawn from other components of language besides syntax. The analysis shows that although there is no sociolinguistic variation in the use of the strategies speakers use to mark discourse-new information, there are significant social class and gender differences in the use of Noun Phrases that are not marked. Whilst acknowledging the risks of generalizing on the basis of large-scale social categories, an interpretation of these differences is suggested in relation to findings from previous research that suggest differences in the interactive style of different gender and social class groups. The paper discusses some implications of the analysis for the fields of language variation and change, and pragmatics.

Key words: syntactic variation; pragmatic variation; information structure; gender; social class; discourse style

Running title: Syntactic variation and beyond

Number of words in main text: 9760

1. INTRODUCTION

In this paper I suggest a new approach to the sociolinguistic analysis of syntactic variation. I will argue that the social dimension of variation is fundamentally different for phonological forms and syntactic forms. Because the form-meaning relation is arbitrary for phonological variants, speakers can use them to index their membership in different social groups; but because speakers use syntactic forms in the construction of discourse including, crucially, the conveying and construction of propositional and attitudinal meanings, the social embedding of syntactic variation is often more complex. This makes it less easy to discern and, I suggest, conventional approaches to the analysis of variation may prevent us from discovering it at all. Syntactic forms may frequently be involved in sociolinguistic variation, but indirectly, as just one of a broad set of choices that includes forms drawn from other components of language besides syntax. Thus in order to see social variation involving syntactic forms it is helpful to adopt a broad brush approach that focuses on the choices speakers make from all components of their knowledge of language, not just the syntactic component. I will illustrate the approach with an analysis of the relationship between the large-scale social dimensions of gender and social class on the one hand and, on the other hand, an aspect of information structure: the realisation of discourse-new entities in spoken discourse. Like Rickford et al's (1995) study of syntactic variation, the analysis demonstrates the advantages of combining the approaches of different subfields within linguistics. I hope to show that in the specific case dealt with here there are insights for the fields of sociolinguistics (specifically, language variation) and pragmatics as well as, perhaps, for general linguistics.

Syntactic variation is often seen as inherently problematic for sociolinguistic analysis: it is difficult to collect enough tokens of the relevant variants for a quantitative analysis, there has been no agreement on whether the concept of the linguistic variable can be extended beyond phonology, and it is not clear to what extent analyses should be tied to a syntactic theory (nor, if so, which one). Furthermore, when researchers do focus on syntactic variation, their analyses tend to be more one-sided than those dealing with phonological variation, focusing more on language-internal constraints on variation than on the relationship between language and the social world (Milroy and Gordon 2003: 197). As a result, the main advances in our understanding of sociolinguistic variation continue to come from studies of phonological variation; and although our understanding of phonological variation is at a stage where generalisations can be made about the relationship between gender and social class on the basis of more than thirty years of research (see, for example, Labov 1990: 205), no such generalisations can be made for syntactic variation. It has been suggested that similar sociolinguistic patterns exist for syntactic variation as for phonological variation (for example, by Rickford et al 1995: 114-5), albeit with caveats concerning the need for further research, but unexpected patterns have also been found (for example, by Ferrara and Bell 1995), as has an absence of pattern (Meyerhoff in press). Like so many issues concerning variation in syntax (and discourse), the nature of the social embedding of syntactic variation remains unresolved.

Some researchers, in fact, have indicated that social factors are not much involved in syntactic variation. Some claim that syntactic variation is conditioned less by social factors than by language-internal, cognitive and situational constraints (see Rydén 1991, Scherre and Naro 1992). Others have suggested that syntactic variation might rarely, if ever, distinguish social groups in the way that 'classic' phonological (and, perhaps, morphosyntactic) markers do (Hudson 1996: 45,Winford 1996: 188). One reason for this could be the infrequency of syntactic forms relative to phonological or morphosyntactic variants: since syntactic variants are less frequently heard, they are less likely to become associated with a specific social group and to be socially evaluated in the way that is necessary for them to function as sociolinguistic indicators or markers (Bell 1984, 2000). By extension, they may be less likely to have a role in the marking and construction of social identities.

I argue in this paper that syntactic forms can be involved in distinguishing social groups and in the construction of social identities, but that they do not necessarily achieve their social function in the same way as phonological forms. As already mentioned, it may be necessary to look beyond syntax in order to understand their social role. Thus the illustrative analysis presented here began as an investigation of variation between English existential *there* clauses and canonical subject-verb clauses, but it ended as a broader analysis within an approach more usually associated with pragmatics.

2. DATA AND FRAMEWORK OF ANALYSIS

The corpus I analysed was compiled for a research project on dialect levelling (Cheshire, Kerswill and Williams 1999) in three English towns: Hull, Milton Keynes and Reading. Ann Williams recorded all the spoken interaction with the exception of six of the interviews recorded in Milton Keynes, for which Paul Kerswill was responsible. The project analysed the speech of 96 speakers aged between 14 and 15. There were 32 adolescents in each town, of whom 16 attended a school in a leafy middle-class area. The other 16 in each town attended a school in a more working-class area (with 'class' defined broadly in each case). In each school the fieldworkers recorded 8 male speakers and 8 female speakers, first in a one-to-one interview with the fieldworker and then in pairs, speaking more informally but again with the fieldworker. Group sessions were also recorded, and a number of language-related tasks carried out (see Cheshire, Kerswill and Williams 1999, Kerswill and Williams 1999). The research design, then, was set up so that many aspects of the situation were constant: the speakers were the same age, they were recorded in the same location (in a small room on school premises) and, except for the six interviews in Milton Keynes, with the same fieldworker. In the one–to-one sections of the interviews the interviewer put the same questions to all speakers in order to elicit information needed for our focus on dialect levelling: these were questions, therefore, about the social and regional origins of their parents, the speaker's attitude to school and their local area, plus details of their likes and dislikes and about how they spent their time when they were not at school. What differed in the research design were three of the large-scale social parameters conventionally used in sociolinguistic research: the speaker's gender, social class and geographical locality.

Variation between existential *there* constructions and canonical subject-verb clauses is illustrated in 1a below and the invented example 1b (with the exception of 1b, all the examples used in this paper are taken from the interviews)[2]:

(1a) there's a drug car in the village square . it's parked near the bus shelter

(1b) a drug car's in the village square . it's parked near the bus shelter

The long debate about whether the linguistic variable should be used to analyse syntactic variation as well as phonological variation seems to have resulted in a tacit consensus that the condition of strict semantic equivalence can be relaxed if variants can be shown to be equivalent in their discourse function (Dines 1980, Coupland 1983). The discourse function of existential constructions is usually said to be presentative: they are used by speakers to introduce new entities into their discourse. This is not their only function: for example, in the London-Lund corpus of British English they also introduce discourse-old entities, give prominence to an activity, and highlight an attribute or a scene-setting adverbial (Collins 2002; see also Schiffrin 1994:

275-278). In our corpus, however, the presentative function was the most frequent one; and I therefore identified all clauses in the corpus where the 96 adolescent speakers introduced discourse-new entities, with the intention of distinguishing those discourse-new entities occurring within an existential clause from those occurring in a canonical construction.

Although it was not the original intention of the project, the interviews turned out to be ideal for a study of information structure in natural discourse. The interviewer and interviewee had never met each other before the interview, so the questions asked during the first part of the interviews frequently required speakers to refer to items that were unfamiliar to the interviewer. I restricted the analysis to discourse entities represented by Noun Phrases, and coded each Noun Phrase as referring to an entity that was discourse-new, discourse-old or inferable from a trigger in the discourse, adopting the distinctions set out by Prince (1992) and following the procedures used by Arnold et al (2000). These categories of discourse entity are illustrated by the italicised Noun Phrases in extract 2.

Extract 2

The previous discussion had been about leisure time activities; Andrew had mentioned earlier in the interview that he played hockey, and the interviewer now asks about hockey as part of the continuing discussion about leisure activities.

I: and what about hockey?

Andrew: erm well it started well in my *primary school* really . cos erm in the primary when I was in *primary school* erm there was *a boy and his mum* that were into *hockey* . played for Westley so like they came down in *games lesson* and I chose to do hockey and I did it instead of playing *football* usually

The first time *primary school* occurs it is a discourse-new entity, never having been mentioned previously in the discourse. When it occurs a second time, however, it is discourse-old, since it has now been introduced into the discourse. *A boy and his mum is* another discourse-new entity, but *hockey* is discourse-old (it was mentioned in the interviewer's question). *Games lesson* is considered an inferable entity on the grounds that the discourse frame of sports played at school has now been opened by the triggers *hockey* and *primary school* (it is common knowledge that in our culture sports are played in games lessons at school). *Football* is considered an inferable entity for the same reason (football is not only a well known sport in our culture but it is well known to be played in games lessons at school).

As Arnold et al (2000:30) point out, Prince's characterisation of information structure has the advantage of offering a straightforward coding scheme that is useful for empirical studies. A further advantage is that the concept of assumed familiarity that underlies Prince's model (in other words, of the assumptions made by the speaker about the familiarity of the interlocutor with the discourse entity) relates several different interpretations of the traditional distinction between 'given-ness' and 'new-ness'. It sees as interdependent notions such as predictability (in the sense that speakers assume hearers can predict material that is recoverable from the discourse context), saliency (the idea that speakers assume that hearers have an entity in their consciousness at the time of speaking) and 'shared knowledge' (where speakers believe that listeners already know the information and believe it to be true, either because they have been explicitly told the information or because they can infer it from what is already known). This is discussed by Prince (1981: 225-233). The framework is more comprehensive than those concerned mainly with a single dimension such as, for example, frameworks derived from the notion of accessibility. Here forms are ranked hierarchically in terms of how readily they can be accessed from memory, so that, for example, unstressed or stressed pronouns are at the higher, most accessible, end of the hierarchy and full names are at the other end of the hierarchy (see, for example, Ariel 1991).

I decided to exclude a number of Noun Phrases from the analysis and to make some small adaptations to Prince's framework, as follows. First, Prince's framework distinguishes between entities that are new to the discourse as well as new to the hearer, and entities that are new to the discourse but not to the hearer (termed Brand New and Unused, respectively, in her 1981 paper). For example, in extract 3 the musician *Jimmy Hendrix* is new to the discourse but not, perhaps, to the hearer (and can be considered, therefore, an Unused entity rather than one that is Brand New). Andrew presumably assumes, in other words, that it will be clear to the interviewer that his friend introduced him to Jimmy Hendrix's music rather than to Jimmy Hendrix himself.

Extract 3

Andrew: I played piano since I was nine but then found it really really boring so I moved erm well in between that I got a guitar because er my friend like introduced me to Jimmy Hendrix and from that I went into more popular stuff

I wanted to avoid the analyst having to take on the role of the speaker in order to assume when information could be assumed to be known to the hearer, so I did not include in the analysis terms where the hearer-status was unclear. These were mainly proper nouns referring to towns or to people – sometimes famous people such as film stars or popular musicians such as Jimmy Hendrix, or school friends that the interviewer may or may not have met. Sometimes it was clear from the way the discourse proceeded that speakers had assumed an entity would be discourse-new but hearer-old when in fact it was not, as shown in extract 4[3], but since hearers may allow unfamiliar items to pass unchallenged in order not to disrupt the flow of talk it was not always possible to tell when such items were really hearer-new. However, items of this kind accounted for only a small number of NPs in the data.

Extract 4

I: do you read any magazines or anything?

Pete: erm..yeah I get Amazing X men

I: a. what?

Pete: Amazing X men it's a comic

I: oh is that a comic?

Pete: yeah

I: oh I didn't know is it American?

Pete: er yeah

Second, following Prince (1981), I did not include NPs occurring as adverbials, such as *last week*, nor NPs that were dummies or parts of an idiom. I also excluded indefinite *they*. Third, synonyms, partial or otherwise, were coded as discourse-old items. For example, I coded *mates* in extract 5 as a discourse-old item used as a synonym for *friends*.

Extract 5

James had explained that he hates coming to school

I: you don't even like coming to be with your friends?

James: well I'd rather stay at home and be with my mates

I confined the analysis to the first part of the interviews, where speakers were interviewed alone. When they were interviewed in pairs the speaker may have assumed that some of the entities to which they referred were new to the interviewer but familiar to the friend, making coding impossible. I anticipated, correctly, that there would be a large number of discourse-new items in the first sections of the interviews.

After excluding the NP tokens mentioned above a total of 970 NPs referring to discourse-new entities remained. They represented 9.75 per cent of the total number of NPs analysed (9953 Noun Phrases). A further 31.2 per cent (3105 NPs) referred to entities inferable from a trigger in the discourse; and the remaining 59.06 per cent (5878 NPs) referred to discourse-old entities. The adolescents referred to discourse-old entities more than any other kind, then, conforming to what Prince terms a General Conservation Principle:

> hearers do not like to make new entities when old ones will do… and speakers, if they are cooperative, form their utterances so as to enable their hearer to make maximal use of old entities (Prince 1981: 245-6).

3. DISCOURSE-NEW ENTITIES IN MARKED SYNTACTIC CONSTRUCTIONS

Of the 970 NPs referring to discourse-new entities, 65 (6.7 per cent) occurred in existential clauses. A further 79 occurred in other marked syntactic constructions. These included constructions with left dislocation, possessive HAVE (GOT) constructions and *it*-constructions, as in 6, 7 and 8:

6. and then who my uncle’s married to she comes from Somerset

7. they’ve got cameras up and they’ve got a police camera up as well

8. it's like too many people are going into business

Already, then, it is necessary to extent the analysis beyond existential constructions and canonical clauses, to include other marked constructions. The marked constructions function in a similar way. It is well-known that speakers of English tend to avoid placing discourse-new entities in initial, subject position. As a result, subjects tend to be 'light' pronouns (Chafe 1980), referring to entities that have been previously mentioned; and speakers use a 'conspiracy of syntactic constructions' (Prince 1981: 228) to position discourse-new entities at the end of the clause. The constructions in 6-8, then, are part of that conspiracy. I also included in this category clauses such as 9. These resemble the structures commonly described as left dislocation (in Carter's terminology both 6 and 10 would be considered as clauses with heads; see Carter 1991:151) and are characteristic of spoken discourse.

9. and their sister her husband owns a shop in MFI

Together, these constructions accounted for just 14.85 per cent of the discourse-new entities in the corpus. There were no consistent sociolinguistic patterns in the use of any of the marked constructions relative to unmarked, canonical constructions. In the Hull middle class group, for example, the girls used more of the marked constructions than the boys, but in the Milton Keynes middle class group male and female speakers used the same number of marked constructions. In the Reading middle class groups, on the other hand, it was the boys who used the higher number of marked constructions, and in all three towns the working class speakers used these forms with approximately equal frequencies (the figures can be seen in the Table in the Appendix). This lack of consistent sociolinguistic variation confirms (if confirmation were needed) the status of the given-before-new principle as a discourse universal (Gundel 1988: 229, Prince 1981, and many others). As expected, in the unmarked constructions NPs referring to discourse-new items were all in non-initial, non-Subject positions in the clause (again, see Chafe 1980; see also Crystal 1980).

Had I been committed to a variationist analysis, the absence of clear social variation could have led to my abandoning the analysis at this point, since the variables traditionally considered most useful to study are not only frequent and integrated into linguistic structure but also socially stratified (Labov 1972: 7-8). Abandoning the analysis, however, would have overlooked some significant social variation, as we will see. In the canonical clauses a wide range of other linguistic forms were used in conjunction with the NPs referring to discourse-new entities; and when these other forms were taken into account clear patterns of gender and social class differentiation in the marking of discourse-new entities emerged.

Since some of these other ways of marking discourse-new items have not previously been discussed in the literature on information structure, I will briefly describe them before turning to the sociolinguistic dimension of the analysis. I will consider the syntactic constructions illustrated in examples 6-9 as forming a first category of discourse-new markers. It is then possible to group the other forms used to introduce discourse-new items into three further categories, making four in total. There is also a fifth ‘miscellaneous’ or ‘other’category.

4. A BROAD BRUSH APPROACH TO THE MARKING OF DISCOURSE-NEW ENTITIES

A second way of marking discourse-new entities was to use a linguistic form that explicitly creates interspeaker involvement. Typically these were set marking tags such as *and stuff* or *and everything*, occurring after the Noun Phrase referring to the discourse-new entity, and pragmatic particles such as *like*, *sort of* or *you know*, in this case occurring before the NP. These forms allow speakers to be imprecise in a way that is interactively felicitous: they function as positive politeness markers, signalling the speaker’s assumption that there is common ground between themselves and the interlocutor that can be drawn on as necessary to identify the referent (Brown and Levinson 1987: 122). Even if the interlocutors do not share enough background knowledge for there to be common ground, the implication that there could be is still interactively felicitous. There is an example in extract 10, where *like* marks the discourse-new items *hygienic baths* and *supermarkets*:

Extract 10

I: do you work with your dad as well?

Karen: no <GIGGLES>

I: what does he do?

Karen: he erm sells like hygienic baths you know to like supermarkets you
know the big ones

In addition to these lexical items, speakers uttered discourse-new items with high rising tones, a further way of creating involvement (Britain 1992). Indefinite *this*, as in 11, could also be considered in this category (Cheshire 1989).

11. my mum and I started having this conversation

This category accounted for a further 181 (18.66 per cent) of the discourse-new items.

Next, speakers sometimes uttered a Noun Phrase and then immediately expanded it, perhaps because they realised that they had not provided enough information for their interlocutor to successfully identify the referent (Clark and Wilkes-Gibbs 1986:4). An example of this strategy is seen in 12, where *the house* is expanded to *somebody's house*, and then modified to *the bedroom*.

12. some nights we sit in the house in somebody's house in the bedroom

Non-restrictive relative clauses served a similar function, allowing speakers to add extra information to the Noun Phrase as the discourse unfolds. Thus in 13 *who's my nan's sister* expands *my Aunt Lucy*:

13. in my family I've got my mum my dad my nan and then my Aunt Lucy who's my nan's sister

Expanded NPs accounted for a further 97 (10.0 per cent) discourse-new NPs.

The fourth category consists of a range of features that could be considered performance phenomena or dysfluencies. They include hesitations, repetitions, filled and unfilled pauses, and false starts. Perhaps, as Arnold et al claim (2000: 47), these are always direct evidence of difficulty in production, but the difficulty can have many causes including, of course, that of accessing from the mental lexicon a noun that has not previously been mentioned in the discourse. In fact, the use of pauses and hesitations in the data fits the predictions of Lindblom's phonetic theory. Lindblom (1990, 1996) is concerned with the adaptive organisation of speech production. He proposes that speakers and listeners co-operate in communicating such that a speaker will include as much co-articulation as the listener can tolerate. If the topic of conversation is known to the listener the speaker will speed up and include phonetic reductions, coarticulations and the like; if the topic is unknown then the speaker will articulate more slowly and carefully. Trudgill (2002), in similar vein, argues that in isolated communities where everyone knows everybody else and speakers have a large fund of shared information there will be a more frequent use of fast speech processes than in other types of community. This, he claims, can result in grammaticalization being more frequent in isolated communities, due to phonetic reduction and the consequent loss of lexical material. Both Trudgill and Lindblom focus on co-articulation phenomena, or fast speech processes, rather than on pauses and hesitations, but pauses and hesitations can also indicate that speakers are taking care with their articulation when introducing new information into the discourse. In any case, even if dysfluencies or performance features reflect a difficulty in production that is unrelated to the introduction of new information, from the perspective of the interlocutor they can still function as a cue indicating that the speaker is about to produce new information (Geluykens 1987).

I included in this category a few instances where speakers marked discourse-new items with slow careful articulation (again, as would be predicted by Lindblom's theory), as in extract 14:

Extract 14

Jake has been talking about his father's job as a university lecturer

I: does he have a lot of lecturing?

Jake: yes he does and he's got to write a lot of uh

I: papers and things?

Jake: yeah and h he's just finished a book so

I: oh really? what about?

Jake: about the ethics of war <SLOW CAREFUL ARTICULATION>

I: the ethics of war

Jake: yeah

Performance cues such as these accounted for a further 81 (8.35 percent) of the discourse-new items.

The 'other' category included occasions where there were explicit efforts at lexical retrieval, as shown in extract 15. These create interpersonal involvement by allowing listeners to witness at first hand the processes the speaker goes through to locate the appropriate lexical item.

Extract 15

I: and what do you want to be when you leave school?

Jeff: either a doctor or a computer .. s.s.scientist well you know make computers programming em computer programmer that's it

I: have you got a computer at home?

In addition, some speakers used multiple strategies, as in 16, where Andrew introduces the discourse-new entity *Australian teenage band* using an existential *there* construction and *like*, with repetition of the construction and a brief pause before the repeat. Although they were not particularly frequent, multiple strategies occurred more frequently in the interviews with adolescent boys than with the girls.

Extract 16

Andrew: well there's this . there is like an Australian teenage band at the moment that play that kind of music

I placed a further 58 NPs (5.98 per cent) in this final miscellaneous category.

We have seen, then, that a full analysis of the forms fulfilling the same discourse function as the existential clauses with which this analysis began shows speakers drawing on a range of linguistic forms. They include syntactic forms made available by the grammar, discourse markers and pragmatic particles associated with discourse management and the construction of interpersonal relations as well as performance features stemming from the demands of on-line production. As mentioned earlier, there were no consistent patterns of sociolinguistic variation in the use of the different categories of forms.

5. SOCIOLINGUISTIC VARIATION IN THE USE OF BARE NOUN PHRASES

Speakers did not always mark their discourse-new entities in the ways described above. As many as 410 discourse-new items (42.27 per cent) were introduced in canonical clauses without any explicit linguistic marking. Examples of such 'bare' NPs are *a golf course* in extract 17, and *a Doberman* in extract 18:

Extract 17

Veronica is talking about her sister's recent visit to Australia

Veronica:	she brought lots of photos like kangaroos and llama. lla.lima whatever they're called <CHUCKLE>
I:	yeah and there are wallabies
Veronica:	yeah and koalas
I:	koalas
Veronica:	and cute animals like that <CHUCKLE>
I:	yeah my son went diving there . he went to a diving school

Veronica: my sister went to a golf course

I: oh a golf course ?

Extract 18

I: er right have you ever been really terrified?

Katie yeah

I: and what happened then?

Katie: erm I got attacked by a Doberman

I: a Doberman oh was that in your street then?

Katie: no it was in my house

It was in the frequency of use of bare NPs such as these that sociolinguistic variation occurred. In all three towns there was a highly significant gender and social class distribution, with working class female adolescents using the highest proportion of bare NPs and middle class male adolescents the lowest proportion. The effect of gender was especially striking for the middle class groups in all three towns, as Figure 1 shows.

Figure 1

The effect of gender was weaker, but still significant for the working class groups, as seen in Figure 2.

Figure 2

The gender distribution can be seen very clearly when the scores for individual speakers are compared. Figure 3, for example, shows the percentage use of bare NPs by the middle class adolescents in Reading. Although there was much individual variation in the use of different forms that marked discourse entities as new, with some speakers using, say, more pragmatic particles and others using more syntactic constructions (and others using all of the forms mentioned above) every female speaker used bare NPs at least once – mostly more than once – and most used bare NPs more frequently than any of the discourse-new markers. By contrast, only three of the middle class boys used bare NPs, and the frequency with which they used them was uniformly low.

Figure 3

More of the working class boys used bare NPs, as seen by the figures for individual speakers in the Reading working-class group, in Figure 4, but although this time 7 of the male speakers use bare NPs, all except one uses them less frequently than the female speakers.

Figure 4

Finally, Figure 5 shows the proportion of bare NPs used by the social class and gender groups in all three towns together, demonstrating clearly that gender has a stronger effect than class on their occurrence, with middle class male speakers using the lowest number of bare NPs and working class female speakers the highest number.

Figure 5

Further evidence for the male preference to mark discourse-new entities comes from the fact that the male adolescents sometimes changed tack in mid course, as in extract 19, where Jake changes both his Noun Phrase, choosing the more specific *friends at work* in place of *everyone up there*, and the syntactic construction, changing to a possessive HAVE GOT that avoids the discourse-new entity being in clause-initial position:

Extract 19

The discussion had been about Jerry's father, an academic whose research area is in the field of politics. Jerry and the other members of the family, however, are not at all interested in politics

I: is your mum interested in politics?

Jake: no

I: no?

Jake: no

I: your poor dad he must feel quite isolated

Jake: well e. everyone up he's got lots of uh friends at work [that do it so

I: [yeah

colleagues

Jake: yeah

Similarly, the use of multiple strategies to mark discourse-new entities, though infrequent overall, occurred more often in the speech of the male adolescents. The fact that the same gender and social class patterns occur in three separate regions of England is compelling evidence of a previously unsuspected tendency for male adolescent speakers, especially middle class males, to mark discourse-new items in their talk[4].

Thus, to return to the starting point of this paper, variation between existential clauses and canonical clauses does not of itself have a role in distinguishing social groups within a community. It does, however, form part of a complex of strategies harnessed by speakers to accomplish a specific discourse function (marking discourse-new entities). When the full complex of forms is taken into account it becomes possible to see sociolinguistic patterning within the community. These social patterns would have been overlooked had I stayed within the confines of the variationist approach. The variable may well be a heuristic construct that does not necessarily map directly onto the units of linguistic structure (see Wolfram 1993, Winford 1996), but it is usual for variants to come from the same component of language: we commonly distinguish, for example, between phonological variables and grammatical variables. It remains to be seen to what extent this methodological point might apply to other types of variation involving syntactic alternants, a point to which I return in section 5. First, however, I will venture some interpretations of the social distribution of the bare NPs in the recordings.

6. SOCIOLINGUISTIC VARIATION

It is of course hazardous to try to explain correlations between linguistic forms and the large-scale social categories of gender and social class. As Milroy and Milroy (1997: 53) have pointed out, in variationist research speaker sex is intended to be a methodological, exploratory variable: in other words, it is a purposely broad, unrefined social variable that can be easily taken into account at the data collection stage of research. The same can be said of the category of social class, especially when, as here, speakers are categorised into just two broadly defined social class groups. The social categories are exploratory, then: if other researchers categorise speakers in the same, albeit simplistic way we can draw useful comparisons across different communities and can replicate studies in different settings.

It is worth pointing out that there was a strong awareness of social class amongst the adolescents who took part in this study (see Kerswill and Williams 1997), as well as of gender. Neverthess, interpreting the results of any analysis based on such large-scale categories is likely to be simplistic. Consider, for example, the many conflicting explanations that researchers have suggested to account for the persistent finding that women tend to use a higher proportion of standard features than men of the same social class (see, for discussion, James 1996). Most of the explanations have some degree of plausibility, but a full understanding of why male and female speakers behave in this way would need to incorporate results from smaller-scale studies based on a more ethnographic approach. AsBergvall (1999: 289) points out, we must also take account of forces larger than local communities, involving the broadly held social and cultural values that are invoked and reified in the national and international media (see further Cheshire 2002: 424).

It is beyond the scope of this paper to undertake such a full exploration of the social embedding of the strategies used by the adolescents to refer to discourse-new entities. Instead, I will follow the example of previous researchers who have analysed sociolinguistic variation in grammatical and discourse features, and consider whether the variation reflects the discourse styles of different social groups. There is evidence, both within the interview data analysed here and in the research literature, that suggests broad differences in the construction of discourse by female and male speakers, as well as by different social classes. I will discuss each of these differences in turn.

Gender

In an earlier analysis of the subset of 16 adolescents from the middle class school in Reading (Cheshire and Williams 2000) we noted the gender distribution in the use of bare NPs and suggested that the adolescent boys were paying greater attention to the referential meaning of the answers they provided to the interviewer’s questions. We proposed that the girls, by contrast, paid greater attention to the interpersonal relationship they were constructing with the interviewer. We pointed out that the choices the boys made to mark their discourse-new items more frequently than the girls corresponded to a general impression we had formed that the boys were doing their best to be helpful, co-operative interviewees. This was shown, for example, by their responses to one of the early questions in the interview, asking where they lived. Many of the boys gave their full address, with the house number and name of the street; the girls, by contrast, rarely did this, instead describing the general area in which they lived and locating their own house very roughly within that area. In addition, the boys often checked the precise focus of a question, as in extract 20.

Extract 20

I:	and I have to ask you what do you like best in school?
Dave:	what…subject do you mean?

The girls never checked the focus of the question in this way. Instead, many of them seemed to construe the interview as an opportunity for a conversation with the interviewer, and they gave as much attention to how she might perceive their answers as to whether they were providing an adequate response to the question. For example, some added metacomments to their replies, such as *I know that sounds a bit strange* in extract 21.

Extract 21

I:	is there anybody you really admire? or you look up to?

Rachel: um . I've never really thought about this before . um . my grandparents . I know that sounds a bit strange but my gran . he's like my stepgrandpa but he's just . don't know why but I just admire him

The boys did not give any comments of this kind. Furthermore they rarely revealed their stance towards the content of their replies, and rarely included any self-disclosure, whereas the girls did this frequently. A typical example is in extract 22.

Extract 22

I: so what about you then what do you want to do . do you want to do politics and international relations or do you want to

Sally oh I I'm just I can't decide what I want to do there's so many things I could do and there's so many things I would like to do but there's things holding me back like I wouldn't want to do that much work I wouldn't want I've got no self discipline and I'm completely indecisive

The boys, then, appeared to see the first part of the interview as a situation where they were being asked for information (which, of course, they were), and they seemed to do their best to provide it in as helpful a way as possible. They interpreted the questions literally, and they tended to give literal, factual replies. Their more frequent flagging of the discourse-new entities could, then, be seen as a reflection of their general orientation to the discourse. Although the girls sometimes gave factual replies, they tended to focus as much – sometimes more – on the interpersonal aspects of the interaction. For them, the boundaries between a conversation and an interview seemed blurred. As a result they perhaps assumed that the interviewer would ask for clarification about discourse-new entities if this was needed, as would happen in a normal conversation. In conversation too great a degree of precision can be cumbersome, sounding excessively impersonal or authoritarian (Channell 1994). Perhaps this is because, as Dressler and Wodak (1982: 716) have argued, in more informal speech styles the needs of the hearer are subordinated to the needs of the speaker. Dressler and Wodak focus on the implications for the use of fast speech processes, but there are also implications for the marking of discourse-new items (as mentioned earlier, Prince's framework of analysis, which I have used here, assumes that speakers form their utterances in line with their assumptions about the needs of the hearer). It is possible that the fact that the main interviewer was female contributed to the construction of the interview as a more conversation-like speech event, with female adolescents assuming shared understandings. However if that were the case we might expect there to be differences between the proportions of bare NPs in the six interviews that Paul Kerswill conducted with Milton Keynes boys, and the interviews with adolescent boys carried out by Ann Williams. There were no differences, in fact.

There is much support in the research literature on language and gender for the idea that female and male speakers may have a different orientation to the expression of referential and affective meaning, realised through the use of speech acts such as compliments and apologies, the use of linguistic hedging and boosting devices, and in turn-taking and other aspects of linguistic behaviour in public and private contexts (Holmes 1995). In the formal writing that is part of the British National Corpus, male authors use a higher proportion of features identified by Biber (1995) as ‘informational’, including more determiners, prepositions and post-head modifiers with *of*-phrases, whereas female writers use more features associated with ‘involvedness’, such as first and second person pronouns (Argamon et al 2003). The literature, then, supports the idea that in the interviews analysed here the boys were more concerned with the referential meaning of the answers they were constructing, and that this is reflected in the care they took when introducing into their discourse entities they assumed to be unfamiliar to their interlocutor. The girls, on the other hand, would then be more concerned with the affective component of the talk they were constructing; as a result they could be expected to pay less attention to the information status of the entities they introduced into the discourse.

Social class

Does this suggest that male working class speakers share the orientation to talk that I have suggested is shown by the girls in this study? Recall that in all three towns the groups of working class boys used a higher proportion of bare NPs referring to discourse-new entities than the groups of middle class boys. There is some support for this interpretation from other studies. For example, Hemphill’s (1999) research into social class differences in narrative form found that middle class speakers used more abstracts than working class speakers: these abstracts framed the upcoming story and summarised it. This, then, suggests a greater concern with the clear expression of referential meaning on the part of the middle class speakers. The working class speakers in Hemphill's study used more codas, and these showed their personal involvement or evaluation of the story. Here, then, the working class preferences parallel the orientation to the discourse said to be more characteristic of female speakers than male speakers.

The wider research literature, however, suggests a somewhat different interpretation for the social class dimension of variation. Some researchers have suggested that middle class speakers have an independent, speaker-oriented speech style that contrasts with a working-class collaborative, addressee-oriented style. For example, Macaulay (2002) reports that middle class speakers in his studies used a higher proportion of adverbials that made the speaker’s stance or personal viewpoint perfectly clear. Working class speakers on the other hand tended to leave their interlocutor to infer their point of view, reflected in a greater amount of detail to ‘paint the scene’ and a more frequent use of syntactic constructions such as fronting to give focus to specific elements in the discourse. A further example comes from Hemphill’s (1989) analysis of group discussions: Hemphill found that working-class speakers used more ellipsis in turn-initial position, such that the ellipted part of their utterance could be found in the previous speaker’s turn, as in extract 23 (though note that I have not analysed the use of ellipsis in the interviews with adolescents).

Extract 23

I:	and did your mum and dad grow up round here or did they grow up in east Hull?
Mick:	grew up round here

Hemphill also found that the working class speakers used pronouns in turn-initial position, to tie their turn to the previous speaker's turn. Middle class speakers preferred a full Noun Phrase, framing their turn as a separate contribution to the discourse, as in extract 24.

Extract 24

I:	and what does your dad do?
Pete:	my dad's an aeronautical engineer

Also relevant is Bernstein's (1971) finding that working class speakers used indefinite *they* more frequently than the middle class speakers in his study. This was castigated by Bernstein as indicating a lack of differentiation, but in fact, as Weiner and Labov (1983: 33) point out, it is a major alternant to the agentless passive, representing simply an alternative way of saying the same thing in an active construction. Taken together, these studies of social class differences suggest a working class approach to the construction of talk that is collaborative, allowing interlocutors the freedom to draw their own conclusions and interpretations, and to take an equal share in the construction of meaning. The use of bare NPs by the working class adolescents in the present study could be seen as a reflection of this general orientation. Insofar as their use constructs meanings that are implicit rather than explicit they can also be seen as illustrating a characteristic of cultures that historically have been more influenced by oral traditions than written ones. Holmes (1998) suggests this as an explanation for the less explicit strategies used in narratives told by New Zealanders of Maori origin. Trudgill's ideas about linguistic and social typology are again relevant: he suggests that some social groups are more characterised by the occurrence of formal situations than others, and that in formal situations speakers are likely to pay more attention to the needs of the hearer (Trudgill 2002:716). These situations, as I argued earlier, would give rise to a more frequent marking of discourse-new entities. Historically, it is middle class groups (and male middle class groups at that) who would have been the dominant group in public life and who would therefore have taken part more frequently in formal situations. A middle class tendency to construct explicit meanings may to some extent be the legacy of the historical dominance of this group in public life.

The linguistic behaviour of the female working class groups and the male middle class groups would then be seen as representing the opposite extremes of these group tendencies. For the female working class adolescents in this study the combination of a focus on affective meanings and a strategy of allowing speakers freedom in interpretation results in their using a higher number of bare NPs to refer to discourse-new entities than any other group of speakers; conversely, for the male middle-class adolescents the combined focus on referential, explicit meanings and their own independent contribution to the construction of the discourse leads them to use the lowest number of bare NPs.

These interpretations are open to question. What is important for the argument of this paper, however, is to note that many researchers have interpreted gender and social class differences in the use of a wide range of linguistic features in terms of different interactive styles, or different orientations to the construction of talk, and that in many of these studies these styles or orientations are constructed by speakers making choices from a diverse range of forms, so diverse that they could not realistically be analysed within the framework of a variationist analysis (although a straightforward quantitative analysis is of course possible, as here). The diverse forms include syntactic constructions: this is the case, for example, for the variation in the use of adverbials and constructions involving focus analysed by Macaulay and, indeed, for the variation between syntactic constructions, lexical items and other forms that mark the discourse new items analysed here. They are not confined to syntactic forms, however; so in order to discover sociolinguistic patterns in the data it is important to look beyond syntactic variation. I suggest that in many cases it may be helpful to analyse syntactic variation, as here, as a first step in a discovery procedure that could then lead to an understanding of the social embedding of syntactic forms. The nature of the social embedding can be quite diverse. In a previous study (Cheshire 2003), for example, an initial focus on unattached *when* clauses (with no main clause) led to an analysis of the story openers used in adolescents' narratives, and this in turn led to an understanding of how male and female adolescents constructed friendship through their storytelling practice.

5. IMPLICATIONS

Language variation

For the analysis of linguistic variation the discovery procedure I have just mentioned may allow progress to be made in our understanding of the social embedding of syntactic variation. Analyses of syntactic variation continue to take linguistic form as the starting point, in line with Weiner and Labov's early claim that 'it is clear that the sharpest analytical conclusions on the conditioning factors that constrain linguistic change and variation can be made when form varies but meaning is constant' (1983: 31).

A focus on form, however, brings with it a methodological problem. Variables are normally set up so that they include one variant that is prescriptively defined as non-standard. This is partly the legacy of early work in the field, which aimed to counter social stereotypes about the nature of stigmatised language varieties. It has also, however, been a way of identifying forms involved in social differentiation so that they can be analysed within the same variationist framework as phonological variation (Winford 1996: 188). Indeed, Dines (1980:17) proposed that the social stigma of a linguistic feature could serve as a heuristic device for initial decisions about what to consider as a variable at the discourse level. A consequence is that there is a tendency in variationist studies to analyse the same features over and over again: typically features reflecting the standard ideology, such as those involving subject-verb agreement (such as, for English, *was/were* variation). I do not wish to argue against analysing features such as these, but simply to point out that the insistence on looking for forms that are both identical in meaning and sociolinguistically salient may prevent us from exploring the social distribution of syntactic forms that are not salient in this way. There are several reasons why forms may not be salient. They may be less frequent than the prescriptively determined non-standard forms; or characteristic of spoken grammar and not, therefore, codified and brought to conscious attention, or e they may be involved in the expression of pragmatic meanings that cannot be handled within a single component of language – such as, of course, information structure.

In addition to the well-known problems of frequency that are associated with the analysis of syntactic variation, then, our conventional frameworks of analysis may prevent us from noticing syntactic variation, let alone exploring it. Winford (1996) has argued that the analysis of syntactic forms that are not socially distributed within a community (in other words, that all speakers have an equal likelihood of using) falls within the scope of pragmatics rather than variation studies. He gives as an example the use of active and passive constructions. The existential construction that was the starting point of the analysis presented in this paper would be a further example. However, the field of pragmatics does not have a tradition of considering sociolinguistic variation. In order not to overlook the social embedding of syntactic constructions it may be necessary, then, to cross disciplinary boundaries, taking as the starting point of an analysis the function of a specific syntactic construction rather than the form, and then exploring the full range of other linguistic forms that speakers use to fulfil the same function.

Pragmatics

As I just mentioned, sociolinguistic variation is not usually taken into account in pragmatic analyses; it has never, as far as I am aware, been included in the analysis of information structure. Approaches to the analysis of information structure can be divided into two broad types, both of which assume that all speakers behave the same way. In the first approach researchers focus on the morphological and syntactic forms that speakers use in their presentation of old and new information, or on the effect of information status on constituent ordering (see, for example, Valduví (1993). Prosody is also taken into account. It is assumed that speakers select referring expressions and syntactic forms in relation to a situated and evolving discourse model of the hearer's current state of knowledge, and that speakers and hearers draw on the phonological, morphological and syntactic structures made available by their grammar in order to present and interpret information. The implicit assumption for many researchers seems to be that this aspect of linguistic behaviour could fall within an expanded concept of linguistic competence (Cameron 1997:30). Lambrecht (1994:3), for example, stresses that information structure is a component of grammar, more specifically of sentence grammar, and that although it is concerned with such psychological phenomena as the speaker's hypotheses about the hearer's mental states: 'such phenomena are relevant to the linguist only inasmuch as they are reflected in grammatical structure (morphosyntax, prosody)'. Thus although the data may come from language in use, potential variation between speakers and within different styles can be ignored. Analyses are often based on a single text, either spoken or written, or on a collection of decontextualised individual examples where spoken and written genres are mixed.

Research in the second tradition sees the identification of a referent as a dynamic, collaborative social act. It focuses, therefore, on language as it is used within an interactional context. For example, Clarke and Wilkes-Gibbs (1986) document how speakers introduce a Noun Phrase that they and their interlocutor may then repair, expand or replace in an iterative process until a mutually acceptable version is arrived at. Smith et al (in press) show how speakers prepare their listeners for the introduction of a new discourse entity in several stages. Sometimes speakers present relevant content to pre-activate the discourse entity; or they may use a variety of interactive devices to ensure that common ground has been activated, such as *you know,* pauses followed by rapid pitch changes, or self repairs. Within this tradition, then, establishing a referent is seen as a joint activity involving a combination of linguistic strategies. Here too, however, potential sociolinguistic variation between speakers has been ignored, this time, perhaps, because analyses are often influenced by Conversation Analysis. Researchers focus on the sequential unfolding of discourse and avoid imposing predetermined categories such as gender or social class unless they are shown in the discourse to be relevant to the speakers. Within this tradition too, then, the default position is that all speakers behave in the same way.

My analysis, however, suggests that it is revealing to draw on both these research traditions, and that speakers of different genders and from different social classes may make different choices about when to select one (or more) of the available strategies to mark the introduction of a discourse-new entity. It would be instructive to perform more detailed qualitative analyses to attempt to determine why speakers mark their discourse-new entities on some occasions and not on others; but for the time being we can simply note that speakers do not, in fact, all act the same way: sociolinguistic variation exists at the pragmatic level as well as in other aspects of language.

Previous studies show that information structure is subject to stylistic or situational factors as well as to social factors. For example, Prince's (1981) analyses of a formal written text and an extended oral narrative found what can be loosely termed stylistic variation between the two texts: the written text contained a lower proportion of discourse-old entities and a higher proportion of inferable entities than the oral narrative, suggesting that in formal academic writing readers are expected to do more inferencing work. In the adolescent interviews analysed here there was a higher proportion of inferable items than in either of Prince's (1981) analyses. I assume that this reflects the question and answer format of the early part of the interviews, where many answers contained an NP referring to an item that was inferable from a trigger contained in the question. By contrast, Arnold et al (2000: 36) report that the Aligned-Hansard Corpus of parliamentary debates contained very few inferable entities (though their analysis focused only on NPs occurring in constructions where dative alternation and heavy NP shift were possible). Perhaps in formal debates where it is crucial to make a point unambiguously speakers choose to rely on inference only rarely. Nevertheless, despite the situational differences in the relative proportions of discourse-old and inferable entities, in all the studies mentioned above discourse-new entities occur infrequently relative to inferable and discourse-old entities, in line with Prince's General Conservation Principle (see section 2).

Some writers have argued that a baseline is needed for analyses of information structure – indeed, of language in general – and that the baseline should be dialogues, with monologues such as reading and writing treated as variation from this norm (Linell 1998, Smith et al in press). The analysis presented here suggests that such a baseline should be more specific still, to take account of the social characteristics of speakers. There is also a case for research on information structure to be based on the language used in more naturalistic settings rather than on the contrived experimental talk on which much previous research has relied. In other words, pragmatics may benefit from paying more attention to both the social and the situational dimensions of language use.

7. CONCLUSION

As we have seen, the adolescents in the study reported here harnessed a wide range of forms to mark discourse-new referents, taking them from all components of their language. The forms included constructions that are normally considered to be generated by the grammar, forms that stem from the demands of on-line performance, and pragmatic particles and discourse markers used in the interactional dimension of discourse. In other words, the speakers happily performed the same discourse function using phenomena that are normally analysed separately from each other.. Perhaps we may learn more about the way speakers make use of syntactic variation in the construction of discourse if as linguists we follow their lead and in our analyses consider abandoning a rigid separation between the different components of language. Such an approach would not fit with a conventional variationist framework, and the focus of the analysis may no longer be syntactic variation. In the illustrative analysis presented in this paper the focus became pragmatic variation, and perhaps this will always be the outcome, given that an empirical analysis of syntactic forms will inevitably involve an investigation of language use. This does not necessarily, in my opinion, put into question the traditional separation between grammar seen as an internal system and grammar seen as a usage-based phenomenon (Newmeyer 2003), but it confirms the view articulated by Pintzuk (2003: 525) that a coherent theory relating grammar and usage can and should be formulated.

It is important, in any event, to recognise that it is not necessarily helpful to consider syntactic variation as if it were a similar phenomenon to phonetic and phonological variation. Speakers use syntactic forms to construct discourse, and the social meaning of syntactic forms relates to their function. This means that the conventional practice of setting up a variable consisting of standard and non-standard syntactic variants may prevent us from conceptualising syntactic variation in terms of discourse function. It is important, then, to think beyond this conventional framework of analysis. Equally, it may be revealing to pluck syntactic forms from their discourse context in order to enter them into a multivariate analysis, but it in this case it is also important to recognise that the only aspects of the discourse context that will be taken into account will be those that we decide ahead of time may be potential constraints on the variation. Other aspects of the discourse contexts will remain unexplored. I have tried to argue here for a different approach to syntactic variation, that uses it as a discovery procedure, focusing on the function of a syntactic form and undertaking a broad-based analysis of other forms with the same function. In some cases we may make no discoveries, but on other occasions we may well be led to new understandings of how speakers use language to create social meaning and social life.

NOTES

1. This paper would not have been possible without Ann Williams and Paul Kerswill, who together recorded all the interviews I analysed in this paper. I would like to thank them for allowing me to use the recordings in this way, as well as for their very helpful comments on an earlier draft of the paper. I am also very grateful for the comments and suggestions made by the referees and the editors of the journal, all of which have much improved the paper. Thanks too to Jennifer Coates for her careful and encouraging reading of the paper. Finally I must thank many colleagues, too numerous to mention, who listened to various oral presentations of different versions of this paper and made very helpful constructive criticisms. I have done my best to take account of all these comments, but of course the ultimate responsibility for the paper has to rest with me.

2. The following transcription conventions are used:

.	short pause (not timed)
..	longer pause (not timed)
?	question marks show the end of a stretch of talk interpreted as an question
<LAUGHTER>	angled brackets give additional information
[	extended square brackets show the beginning of an overlap
	latching
I:	indicates an utterance made by the interviewer

All names have been changed

3. Note that Peter may equally have treated *Amazing X Men* as inferable from the trigger *magazines or anything*; the point remains, however, that his assumptions are shown in the subsequent discourse to have been faulty.

4. I also attempted to analyse social variation in the expressions used to refer to inferable entities: although speakers drew from the same range of strategies as for the discourse-new entities, in most cases the identity of the inferable entities was established interactively, over several turns.

REFERENCES

Argamon, Shlomo, Koppel, Moshe, Fine, Jonathan and Shimoni, Anat Rachel. 2003. Gender, genre, and writing style in formal written Texts. *Text* 23(3): 321-346.

Ariel, Mira 1991. The function of accessibility in a theory of grammar. *Journal of Pragmatics* 16: 443-463.

Arnold, Jennifer E., Wasow, Thomas, Losongo, Anthony and Ginstrom, Ryan. 2000. Heaviness vs. newness: The effects of structural complexity and discourse status on discourse ordering. *Language* 76: 28-55.

Bell, Allan 1984. Language style as audience design. *Language in Society* 13: 145-204.

Bell, Allan 2000. Back in style: reworking audience design. In Penelope Eckert and John R. Rickford (eds.) *Style and Sociolinguistic Variation*. Cambridge: Cambridge University Press. 139-169.

Bergvall, Victoria L. (1999). Toward a comprehensive theory of language and gender. *Language in Society* 28: 273-293.

Bernstein, Basil 1971. *Class, Codes and Control.* St. Albans: Paladin.

Biber, Douglas 1995. *Dimensions of Register Variation: A Cross-linguistic Comparison*. Cambridge: Cambridge University Press.

Birner, Betty J. 1994. Information status and word order: An analysis of English inversion. *Language* 70: 233-259.

Britain, D. 1992 Linguistic Change in Intonation: the use of High Rising Terminals in New Zealand English. *Language Variation and Change* 4: 77-104.

Brown, Penelope and Levinson, Stephen. 1987. *Politeness: Some Universals in Language Usage*. Cambridge: Cambridge University Press.

Cameron, Richard. 1997. Accessibility theory in a variable syntax of Spanish. *Journal of Pragmatics* 28: 29-67.

Carter, Ronald. 1999. Standard grammars, spoken English and education. In Tony Bex and Richard J. Watts (eds.) *Standard English: The widening debate.* London: Routledge. 149-166.

Chafe, Wallace. 1980. The deployment of consciousness in the production of a narrative. In Wallace Chafe (ed.) *The Pear Stories: Cognitive, Cultural and Linguistic Aspects of Narrative Production*. Norwood, New Jersey: Ablex. 9-50.

Channell, Joanna. 1994. *Vague language*. Oxford: Oxford University Press.

Cheshire, Jenny. 1989. Addressee-oriented features in spoken discourse. *York Papers in Linguistics* 13: 49-63.

Cheshire, Jenny. 2003. Sex and gender in variationist research. In J.K.Chambers, Peter Trudgill and Natalie Schilling-Estes (eds.) *Handbook of Language Variation and Change.* Oxford: Blackwell. 423-443.

Cheshire, Jenny. 2003. Social dimensions of syntactic variation: the case of *when* clauses. In David Britain and Jenny Cheshire (eds.) *Social Dialectology*: *In Honour of Peter Trudgill.* Amsterdam: Benjamins, pp. 245-261.

Cheshire, Jenny, Kerswill, Paul and Williams. Ann. 1999. *The role of adolescents in dialect levelling*. Ref. R000236180. Final Report submitted to the Economic and Social Research Council, June 1999.

Cheshire, Jenny and Williams, Ann. 2000. Information structure in male and female adolescent talk. *Journal of English Linguistics* 30: 217-238.

Clark, Herbert H. and Wilkes-Gibbs, Deanna. 1986. Referring as a collaborative process. *Cognition* 22: 1-39.

Collins, Peter. 2002. Some discourse functions of existentials in English. *Proceedings of the 2001 Conference of the Australian Linguistic Society.* 1-6. http://linguistics.anu.edu.au/ALS2001/proceedings.html

Crystal, David. 1980. Neglected grammatical factors in conversational English. In Sidney Greenbaum, Geoffrey Leech and Jan Svartvik (eds.) *Studies in English Linguistics.* Harlow, U.K.: Longman. 153-166.

Coupland, Nikolas. 1983. Patterns of encounter management: Further arguments for discourse variables. *Language in Society* 12: 459-476.

Dines, Elizabeth. 1980. Variation in discourse – “and stuff like that”. *Language in Society* 9: 13-31.

Dressler, Wolfgang and Wodak, Ruth. 1982. Sociophonological methods in the study of sociolinguistic variation in Viennese German. *Language in Society* 11: 339-370.

Ferrara, Kathleen and Bell, Barbara. 1995. Sociolinguistic variation and discourse function of constructed dialogue introducers: the case of *be* + *like. American Speech* 70: 265-289.

Geluykens, Ronald. 1992. *From Discourse Process to Grammatical Construction: On Left Dislocation in English.* Amsterdam: Benjamins.

Gundel, Jeannette. 1988. Universals of topic-comment structure. In M. Hammond et al (eds.) *Studies in Syntactic Typlogy.* Amsterdam: Benjamins. 209-239.

Hemphill, Lowry. 1989. Topic development, syntax and social class. *Discourse Processes* 12: 267-285.

Hemphill, Lowry. 1999. Narrative style, social class and response to poetry. *Research in the Teaching of English* 3: 275-303.

Holmes, Janet. 1995. *Women, Men and Politeness*. Harlow, U.K.: Longman.

Holmes, Janet. 1998. Narrative structure: Some contrasts between Maori and Pakeha story-telling. *Multilingua* 17: 25-57.

Hudson, Richard A. 1996. *Sociolinguistic*s. Cambridge: Cambridge University Press, 2nd edition.

James, Deborah. 1996. Women, men and prestige speech forms: a critical review. In Victoria .L. Bergvall, Janet M. Bing and Alice F. Freed (eds.) *Rethinking Language and Gender Research: Theory and Practice*. Harlow,U.K.: Longman.98-125.

Kerswill, Paul and Williams, Ann 1997. Investigating social and linguistic identity in three British schools. In U-B. Kotsinas, A-B. Stenström and A-M. Malin (eds.) *Ungdomsspråk I Norden. Föredrag från ett forskarsymposium [Youth language in the Nordic countries. Papers from a research symposium].* Series: MINS, no. 43. Stockholm: University of Stockholm, Department of Nordic Languages and Literature. 159-176.

Labov, William. 1990 The intersection of sex and social class in the course of linguistic change. *Language Variation and Change* 2: 205-54.

Lambrecht, Knud 1994. *Information Structure and Sentence Form*. Cambridge: Cambridge University Press.

Lavandera, Beatriz. 1978. The variable component in bilingual performance. In James Alatis (ed.) *International dimensions of Bilingual Education*. Washington, D.C.: Georgetown University Press. 391-411.

Lindblom, Björn. 1990 Explaining phonetic variation: a sketch of the H and H theory. In William Hardcastle and A. Marchal (eds.) *Speech Production and Speech Modelling.* Dordrecht: Kluwer. 403-439.

Lindblom, Björn. 1996. Role of articulation in speech perception: Clues from production. *Journal of the Acoustical Society of America* 99: 1683-1692.

Linell, Per. 1998. *Approaching Dialogue*. Amsterdam: Benjamins.

Macaulay, Ronald. 2002. Extremely interesting, very interesting, or only quite interesting? Adverbs and social class. *Journal of Sociolinguistics* 6: 398-417.

Meyerhoff, Miriam. In press. Syntactic variation. *Encyclopedia of Linguistics.* Berlin: de Gruyter.

Milroy, Lesley and Gordon, M. 2003. *Analyzing and interpreting linguistic variation*. Oxford; Blackwell. 2nd edition.

Milroy, James and Milroy, Lesley (1997). Varieties and variation. In Florian Coulmas (ed.) *The Handbook of Sociolinguistics*. Oxford: Blackwell. 47-64.

Newmeyer, Frederick J. 2003. Grammar is grammar and usage is usage. *Language* 79: 682-703.

Pintzuk, Susan. 2003. Variationist approaches to syntactic change. In Brian D. Joseph and Richard D. Janda (eds.) *Handbook of Historical Linguistics.* Oxford: Blackwell. 509-528.

Prince, Ellen F.1981. Toward a taxonomy of given-new information. In Peter Cole (ed.) *Radical Pragmatics*. New York: Academic Press. 223-54.

Prince, Ellen. 1985. Fancy syntax and shared knowledge. *Journal of Pragmatics* 9: 65-81.

Prince, Ellen. 1992. The ZPG letter: Subjects, definiteness and information status. In William C. Mann and Sandra A. Thompson (eds*.) Discourse description: diverse linguistic analyses of a fundraising text*. Amsterdam: John Benjamins. 295-325.

Rickford, John, Wasow, Thomas A., Mendoza-Denton, Norma and Espinoza, Juli. 1995. Syntactic variation and change in progress: loss of the verbal coda in topic-restricting *as far as* constructions. *Language* 71: 102-131.

Rydén, Mats. 1991. The *be/have* variation with intransitives in its crucial phases. In Dieter Kastovsky (ed.) *Historical English Syntax*. Berlin: Mouton de Gruyter. 343-354.

Sacks, Harvey and Schegloff , Emmanuel 1979. Two preferences in the organization of reference to persons in conversation and their interaction. In George Psathas (ed.) *Everyday Language: Studies in Ethnomethodology.* New York: Irvington. 15-21.

Scherre, Maria Marta Pereira and Naro, Anthony J. 1992. The serial effect on internal and external variables. *Language Variation and Change* 4: 1-14.

Schiffrin, Deborah. 1994. *Approaches to Discourse Analysis.* Oxford: Blackwell.

Smith, Sara W., Noda, Hiromi Pat, Andrews, Steven and Jucker, Andreas H. In press. The role of pre-introductions as reference strategies in monologues and dialogues, *Journal of Pragmatics.*

Trudgill, Peter 2003. Linguistic and social typology. In J.K.Chambers, Peter Trudgill and Natalie Schilling-Estes (eds.) *Handbook of Language Variation and Change.* Oxford: Blackwell, pp.707-728.

Valduví, Enric 1993. Information packaging: a survey. University of Edinburgh Human Communication Research Centre: Research paper HCRC/RP-44.

Ward, Gregory and Birner, Betty J. 2001. Discourse and information structure. In Deborah Schiffrin, Deborah Tannen and Heidi E. Hamilton (eds.) *Handbook of Discourse Analysis.* Oxford: Blackwell. 119-136.

Ward, Gregory, Birner, Betty and Huddleston, Rodney. 2002. Information packaging. In Rodney Huddleston and Geoffrey Pullum (eds.) *Cambridge Grammar of the English Language.* Cambridge: Cambridge University Press. 1363-1447.

Weiner, E. Judith and Labov, William. 1983. Constraints on the agentless passive. *Journal of Linguistics* 19: 29-58.

Williams, Ann and Kerswill, Paul. 1999. Dialect levelling: change and continuity on Milton Keynes, Reading and Hull. In Paul Foulkes and Gerry Docherty (eds.) *Urban Voices: Accent studies in the British Isles*. London: Arnold. 141-162.

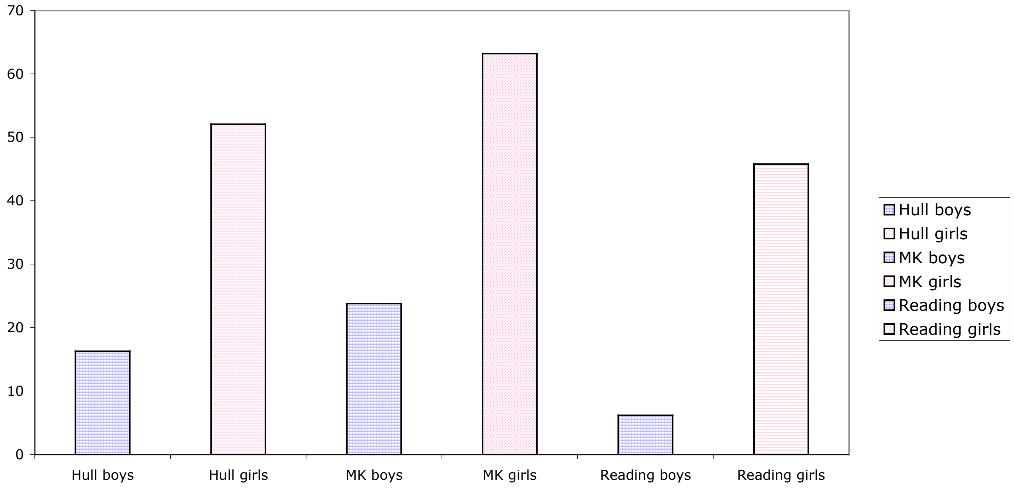
Winford, Donald.1996. The problem of syntactic variation. In Jennifer Arnold, Renée Blake, Brad Davidson, Scott Schwenter and Julie Solomon (eds.) *Sociolinguistic Variation: Data, Theory, and Analysis.* Stanford, USA: CSLI Publications. 177- 192.

Wolfram, Walt. 1993. Identifying and interpreting variables. In Dennis R. Preston (ed.) *American Dialect Research*. Amsterdam: John Benjamins. 193-222.

Appendix. Table1. Percentage of discourse-new markers and bare NPs in the data set (number of tokens)

	syntactic constructions	Addressee-oriented forms	Expanded NPs	Performance cues	Other	Bare NPs	Total
Hull m.cl. boys	10.87 (10)	18.48 (17)	11.96 (11)	19.57 (18)	22.83 (21)	16.3 (15)	100 (92)
Hull m.cl. girls	17.55 (33)	21.28 (40)	3.72 (7)	4.26 (8)	2.13 (4)	52.06 (96)	100 (188)
M.K. m.cl. boys	11.90 (5)	26.19 (11)	11.90 (5)	19.05 (8)	7.14 (3)	23.81 (10)	100 (42)
M.K. m.cl. girls	7.35 (5)	10.29 (7)	6.67 (4)	8.82 (6)	4.41 (3)	63.24 (43)	100 (68)
Reading m.cl.boys	27.7 (18)	26.1 (17)	15.4 (10)	10.8 (7)	13.8 (9)	6.2 (4)	100 (65)
Reading m.cl. girls	16.6 (12)	8.3 (6)	13.9 (10)	5.6 (4)	9.72 (7)	45.8 (33)	100 (72)
Hull w.cl. boys	8.86 (7)	17.72 (14)	24.05 (19)	6.33 (5)	8.86 (7)	34.18 (27)	100 (79)
Hull w.cl girls	12.86 (9)	11.43 (8)	8.57 (6)	4.29 (3)	1.43 (1)	61.43 (43)	100 (70)
M.K. w.cl. boys	12.24 (6)	20.41 (10)	16.33 (8))	10.20 (5)	2.04 (1)	38.78 (19)	100 (49)
M.K. w.cl. girls	12.50 (7)	3.57 (2)	1.79 (1))	14.28 (8)	1.79 (1)	69.64 (37)	100 (56)
Reading w.cl. boys	16.7 (16)	39.6 (38)	12.5 (12)	8.3 (8)	0	22.9 (22)	100 (96)
Reading w.cl. girls	17.2 (16)	11.8 (11)	4.3 (4)	1.1 (1)	0	65.6 (61)	100 (93)
TOTALS	144	181	97	81	58	410	970

Figure 1. Percentage of discourse-new bare NPs (relative to all discourse-new NPS) used by middle class groups

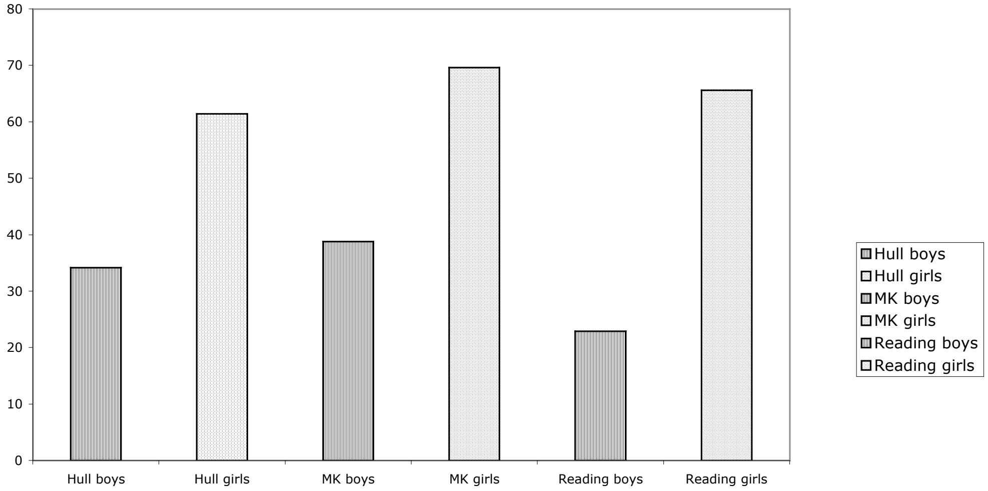


χ^2 value for Hull groups: 31.1918, df1, $p < 0.001$

χ^2 value for Milton Keynes (MK) groups: 16.1644, df 1, $p < 0.001$

χ^2 value for Reading groups: 27.2833, df 1, $p < 0.001$

Figure 2. Percentage of discourse-new bare NPs (relative to all discourse-new NPS) used by working class groups

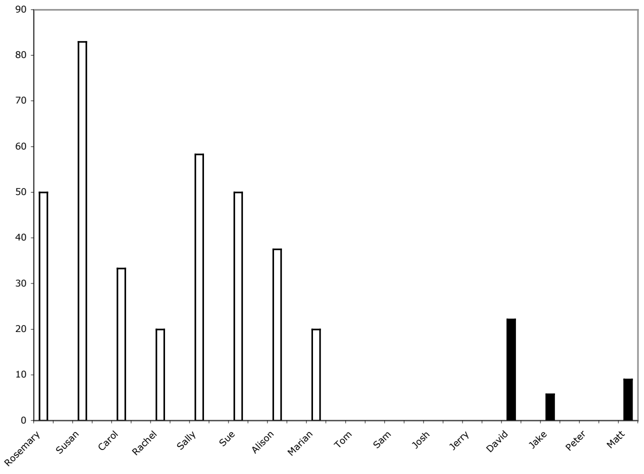


χ^2 value for Hull groups: 11.0653, df 1, $p < 0.001$

χ^2 value for MK groups: 7.8232, df 1, $p < 0.01$

χ^2 value for Reading groups: 34. 9280, df 1, $p < 0.001$

used by individual speakers



clear bars represent percentage of bare NPs used by girls; dark bars represent percentage of bare NPs used by boys

Figure 4. Reading working class groups: proportion of bare NPs used by individual speakers

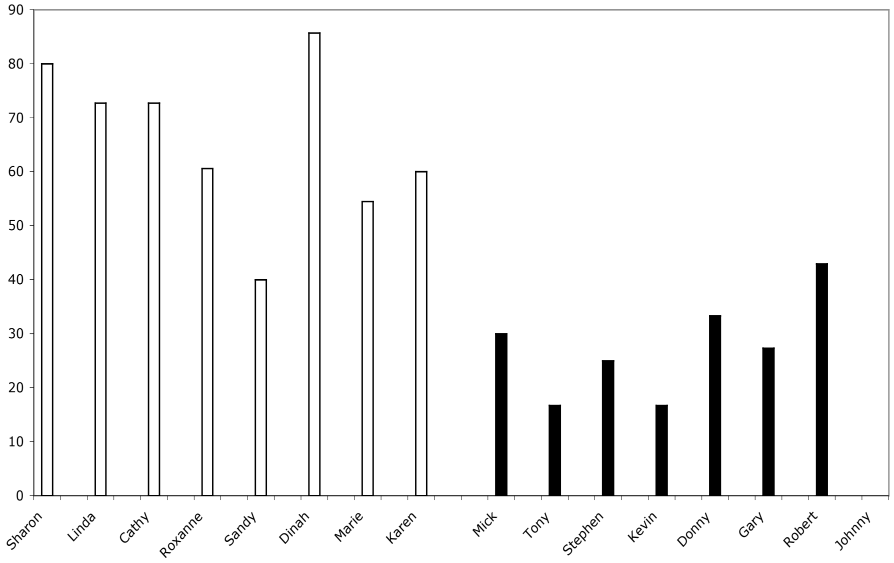


Figure 5. Percentage of discourse-new bare NPs (relative to all discourse-new NPs) by gender and social class (allspeakers)