Article

# Applying corpus linguistics in a health care context

*Svenja Adolphs, Brian Brown, Ronald Carter, Paul Crawford and Opinder Sahota*

## Abstract

*This paper draws on two strands of research and practice in language studies, namely i) studies of communication in health care encounters, and ii) studies of language corpora. It aims to delineate an area of 'applied clinical linguistics' which draws on these existing sub-disciplines so as to enhance our knowledge of communicative events in clinical settings. To illustrate the potential we draw upon a corpus-informed study of communication in staged telephone conversations between callers and advisers in the UK's 'NHS Direct' health advisory service. Here, the combined application of corpus linguistics and conversation analytic techniques has revealed several hitherto undisclosed features concerning strategies used by health advisors to position the caller as the subject of the interaction, give credentials to the advice and terminate the encounter with a 'convergence coda'. The integrated approach which combines corpus linguistic methods with existing frameworks in text analysis of health care contexts seems to offer new possibilities of data and theory building, as well as becoming a resource for practitioners themselves in clinical field settings.*

KEYWORDS: CORPUS LINGUISTICS; CONVERSATIONAL ANALYSIS; HEALTH CARE; CLINICAL COMMUNICATION; APPLIED CLINICAL LINGUISTICS.

## Affiliations and acknowledgement

*Brian Brown:* de Montfort University, UK.

*Other authors*: University of Nottingham, UK.

*Corresponding author*: Ronald Carter, School of English Studies, University of Nottingham, UK.
Email: Ronald.Carter@nottingham.ac.uk

equinox
www.equinoxpub.com

## 1    Introduction

In this paper we aim to map a relatively under-explored area of inquiry in the field of applied linguistics. We will argue that the two fields of health care research and corpus linguistics can be brought together so as to yield fresh insights in both language theory and health care studies, with implications for professional education and working practice in health care. Whereas there is already research in this area, there are a number of interwoven strands which could be brought closer together with benefits to patients, practitioners and the scholarly community as a whole.

This paper will proceed, first, by outlining some issues in the study of communication in health care settings. Secondly, we will describe some results from a study which is based on a research exercise involving staged encounters in a telephone health care environment. Here, we will attempt to show how bodies of text or 'corpora' transcribed from recorded interactions can be subject to computerized analysis and how this process can highlight new avenues of enquiry for the linguist and educator.

## 2    Language, communication and health care

The sheer volume of research on language in health care is phenomenal. The well established focus on doctor-patient interaction has latterly been supplemented by a diversification of enquiries into encounters between patients and nurses (Crawford et al., 1998), physiotherapists (Ballinger et al., 1999), pharmacists (Pilnick, 1998; 1999), occupational therapists (Mattingly, 1994), as well as a variety of alternative practitioners. Yet, despite this frenetic pace of research, most studies so far have been based on relatively small databases, and have not originated in large collections of data. There are some further curious features of this literature which are worth noting. First, it is still dominated by doctor–patient interaction, as many scholars have noted (Candlin, 1997; Candlin & Candlin, 2003), despite the substantial amount of health care being dispensed by professions allied to medicine rather than by medical professionals themselves. There is also a curious hierarchy in the literature, inasmuch as it is studies of doctor–patient interaction which have largely found their way into mainstream social science journals. Studies of interaction between nurses and patients are still mainly found in nursing journals, despite their reliance on many of the same concepts and methodological tools as their counterparts who study doctors. This may reflect the historical priority of research on doctor–patient interaction, as well as broader social hierarchies of gender and professional prestige (Porter, 2001) which have conspired to keep these important developments out of the mainstream.

In addition to this narrowness, as Roberts & Sarangi (2003) note, with a few notable exceptions (Candlin, 2000; Crawford et al., 1998; Elwyn, 2001), much of this research is conducted by health care 'outsiders' rather than 'insiders' and fails to 'foreground a concern for the application of their findings' (2003: 339). In an earlier paper, Roberts and Sarangi (1999) indicate the need for more dialogue between 'research and researched' such that the clinical applications of research are highlighted and the practical concerns of delivering health care are addressed, in the process advancing theory development in linguistics or social science (see also Candlin, 2003).

A good many previous studies of interaction in health care have drawn on conversation analysis which strives to unpack the sequential orderliness of texts as active social phenomena, that is, as part of day-to-day institutional actions. As Drew et al. (2001) characterize it, conversation analysis (CA) focuses largely on the recurrent features of interaction. People are, in this view, attempting to produce meaningful action and to interpret each other's meaning. This approach informs, for example, Heritage and Stivers's (1999) notion of 'online commentary' in the clinical encounter as GPs examine patients. This involves the doctor offering a series of observations as he or she inspects ears or throats and listens to chests. Much of this emphasizes the relatively minor or insignificant nature of the problem, which seems to relate to GPs' unwillingness to prescribe antibiotics unless it is strictly necessary. Thus, conversation analysis and ethnomethodology seek to understand how interactants jointly construct reality in clinical encounters and look at how language facilitates the goals, ambitions and practical procedures of clinical work.

## 3   The development of corpus linguistics

Whilst the developments we have outlined above in the study of health care were taking place, there have been a number of methodological innovations in linguistics, where the compilation of large scale bodies of language or 'corpora' has proceeded apace. This 'corpus revolution' (Leech, 2000) has seen an increasing number of scholars developing large transcribed archives of the terra incognita of the spoken English language (Carter & McCarthy, 1995; McCarthy, 1998). Equipped with such resources and with access to powerful software packages, the present day researcher can explore the spoken word much more readily than in the early days of linguistics. Much of the classic work in language scholarship was performed without the benefit of this quantum leap in language awareness which corpus linguistics affords. The original focus in linguistics on introspective accounts, logical analysis and intuition, incorporating the work of Chomsky (e.g. 1957; 1976; 1993) has been termed the 'internalised' or 'I language' tradition. The other tradition, to which corpus linguistics belongs, and which has latterly been facilitated by the development of computer technology,

has been termed the 'externalised' or 'E language' side of the discipline (Stubbs, 2001). Corpus linguistics leads to a more 'evidence-based' approach to uses of language in different settings.

Whilst it is having an impact on language teaching and learning, the possibilities for corpus research in health care have been relatively under-explored. Some of the possibilities were outlined originally by Thomas and Wilson (1996), in the case of doctor–patient interaction. The following study illustrates how a combined qualitative and quantitative methodology drawing on tools traditionally used for corpus analysis can enhance our understanding of a particular health care setting.

## 4    Background to the NHS Direct study

The NHS Direct service has been in operation in the UK since 1998. Much of existing scholarship on this service relates to factors such as whether the advice was followed (Foster et al., 2003) whether different computer-aided assessment systems yield different outcomes (O'Cathain et al., 2003) and customer satisfaction (Florin and Rosen, 1999), value for money (George, 2002), or the role it plays in the overall system of care (Munro et al., 2000; Nicholl and Munro, 2000). However, we know relatively little about the terra incognita of the consultations themselves which take place within it. Detailed study of this body of health-related communication promises important insights for practitioners, policy makers and health educators as well as patients themselves. In the short term, it is useful to examine issues relating to whether the advisers, nurses and doctors are delivering the service in a thorough and rigorous yet courteous and sympathetic manner. In the longer term, it will be useful as part of broader research projects to examine how we talk about health and illness as a community of sufferers and healers, so that interview protocols, questioning strategies and even services themselves can be designed to maximise the effectiveness of health delivery.

## 5    The NHS Direct corpus

The fieldwork for this study consisted of a series of phone calls made to NHS Direct in Nottingham. The research calls were made between July and September 2002 using a designated phone number. The health advisers and nurses did not know which of the calls they received were made by the researchers. In order to reinforce the concealed identity of the callers, calls were made mainly during extremely busy periods at NHS Direct (9 am to 11 am and 6 pm to 8 pm). The calls were made from a number of different telephone numbers and addresses across the UK Midlands to enhance the anonymity of the callers.

Overall, seventeen calls were made to NHS Direct staff by two male and two female researcher participants (P). The health problems described by the callers covered a wide range of illnesses and predominantly centred on medication advice. This was done to ensure some degree of conceptual coherence in the materials produced for analysis and to allow some degree of comparability between the different sequences of interaction. The callers improvised their performances based on a pre-agreed script with essential features such as age, occupation, place of residence and the nature of the complaint. These were also designed so as to sample a range of ages and social statuses, from a young homeless man, through to a range of manual and white collar workers from a variety of backgrounds. The health advisers (HA), nurses (N) and occasional doctors (D) who took part would be responding using the NHS computer-aided assessment system (O'Cathain et al., 2003) which would prompt them to ask questions in a sequence which guides the professional through a decision making system to enable a thorough assessment to be conducted and a safe course of action to be decided upon.

After recording, the tapes were sent to specialist transcribers who converted them into electronic format suitable for analysis. Overall, the interaction amounted to 61,981 words. For the purpose of corpus analysis the transcripts have been split into utterances made by the health adviser, nurse or doctor (the health professionals' corpus) and those made by the patients (patients' corpus). The health professionals' corpus amounted to 35,014 words in total while the patients' corpus amounted to 26,967 words. Whilst this is small in relation to the usual size of contemporary corpora which attempt to represent the English language as a whole, the corpus developed for this project is relatively specialised and coherent and will suffice as a preliminary vignette into the thousands of hours of NHS Direct interaction which take place and as a means of illustrating what can be discovered in this manner.

## 6   Analysis

The analysis was carried out in three stages. An initial viewing of the transcripts by all members of the research team revealed some patterns in the interactions between health professionals and patients that seemed specific to this particular type of discourse and were analysed in detail using methods from the field of conversation analysis and discourse analysis. In a second stage, the language used in the health professional corpus was compared with a corpus of general spoken English (the five million CANCODE corpus held at the University of Nottingham[1]) to identify linguistic patterns that are unique to the language of health professionals in NHS Direct phone-ins. The analysis employed the Wordsmith Tools software. Having identified a range of linguistic patterns quantitatively we isolated a smaller set of patterns and analysed these in their

discourse environment. All three stages of the analysis revealed an overarching tendency for the nurses and health professionals to use strategies of politeness and the language of convergence in their interactions with the callers. This often involved strategies to minimise the imposition of the advice that was given, as well as strategies of affirmation and acceptance of the patient's situation and concerns. The main results are summarised in the sections below.

## 7    Results and discussion

### 7.1 Keyword analysis

First, a 'keyword' analysis was used to calculate the frequency of each word in the corpus obtained, and to compare this with the occurrence of the same words in the much larger CANCODE corpus of general English. This yielded a set of keywords which in comparison with the general corpus occur with a significantly higher or lower frequency in NHS Direct consultations. This type of analysis is a useful starting point to isolate language patterns that are specific to a particular group of people or type of interaction and can help to identify patterns of communicative style in different contexts. As such this procedure provides a more quantitative and arguably a more systematic point of entry into the data which sets it apart from more traditional approaches such as conversation analysis for example. A keyword analysis also serves as a powerful hypothesis testing device and enables the analyst to cross-reference the results with his/her intuition about the transcripts.

Using the list of keywords – that is, words which appear from the analysis to be used significantly more frequently or less frequently than we would expect by chance – we ran a concordance search of specific items. Leaving aside medical jargon which inevitably featured in the list of keywords, the remaining items fell into the following categories: negatives, imperatives, pronouns, vague language, affirmations/positive backchannels, directives. The initial concordance search presented a point of entry into the text which allowed us to examine how specific items were used by the interactants in an ongoing stretch of discourse. The selected set of items is listed in Table 1. This process then highlights the potential of a corpus approach to enable us to detect hitherto unexplored features of the linguistic landscape in health care. This would be particularly the case when dealing with larger spoken data sets of the kind we are developing in the Nottingham Health Communication Corpus (NHCC).

**Table 1** (*opposite*) Keyword analysis (selected items focusing on negatives, imperatives, pronouns, vague language, affirmations/positive backchannels and directives)

| WORD | FREQ. | NURSE.LST% | FREQ. | CANF.LST% | KEYNESS |
|------|-------|-----------|-------|-----------|---------|
| OK | 120 | 0,34 | 31 | | 1.069,30 |
| YOUR | 407 | 1,16 | 13.868 | 0,25 | 620,1 |
| OKAY | 334 | 0,95 | 9.580 | 0,17 | 599,7 |
| YOU | 1.330 | 3,8 | 128.248 | 2,27 | 306,5 |
| PLEASE | 97 | 0,28 | 1.678 | 0,03 | 256 |
| RIGHT | 433 | 1,24 | 29.401 | 0,52 | 248,3 |
| ADVISE | 27 | 0,08 | 43 | | 182,1 |
| CAN | 259 | 0,74 | 16.570 | 0,29 | 166,2 |
| TAKE | 119 | 0,34 | 4.685 | 0,08 | 154,5 |
| HELP | 62 | 0,18 | 1.329 | 0,02 | 140,9 |
| MAY | 58 | 0,17 | 1.203 | 0,02 | 135 |
| YOU'RE | 166 | 0,47 | 9.889 | 0,17 | 120,9 |
| IF | 284 | 0,81 | 22.694 | 0,4 | 112,5 |
| AVOID | 19 | 0,05 | 66 | | 104 |
| ANYTHING | 94 | 0,27 | 4.334 | 0,08 | 100,5 |
| OBLIGED | 13 | 0,04 | 15 | | 93,9 |
| PATIENCE | 12 | 0,03 | 21 | | 79,2 |
| HOWEVER | 18 | 0,05 | 146 | | 71,6 |
| SUGGEST | 18 | 0,05 | 168 | | 67,1 |
| OBVIOUSLY | 50 | 0,14 | 2.122 | 0,04 | 59,3 |
| WILL | 82 | 0,23 | 5.356 | 0,09 | 50,4 |
| ADVICE | 14 | 0,04 | 142 | | 50,1 |
| LET | 41 | 0,12 | 1.738 | 0,03 | 48,7 |
| MANAGE | 14 | 0,04 | 204 | | 41,1 |
| THEY'LL | 29 | 0,08 | 1.076 | 0,02 | 40,2 |
| JUST | 271 | 0,77 | 29.020 | 0,51 | 40,1 |
| THANK | 42 | 0,12 | 2.134 | 0,04 | 39,2 |
| NORMALLY | 21 | 0,06 | 625 | 0,01 | 36,3 |
| CERTAINLY | 22 | 0,06 | 718 | 0,01 | 34,8 |
| YOU'VE | 90 | 0,26 | 7.326 | 0,13 | 33,9 |
| PERSONAL | 13 | 0,04 | 269 | | 30,3 |
| USUALLY | 19 | 0,05 | 685 | 0,01 | 27,2 |
| ALRIGHT | 17 | 0,05 | 551 | | 27,1 |
| TRY | 34 | 0,1 | 1.932 | 0,03 | 26,7 |
| FINE | 28 | 0,08 | 1.425 | 0,03 | 26,1 |
| NICE | 4 | 0,01 | 4.406 | 0,08 | 31,1 |
| GOOD | 20 | 0,06 | 9.124 | 0,16 | 31,3 |
| WE | 126 | 0,36 | 32.499 | 0,57 | 32,4 |
| THOUGHT | 3 | | 5.684 | 0,1 | 49,4 |
| REALLY | 38 | 0,11 | 16.223 | 0,29 | 50,9 |
| THINK | 57 | 0,16 | 22.527 | 0,4 | 62,8 |
| LAUGHS | 54 | 0,15 | 33.171 | 0,59 | 158,4 |
| I | 481 | 1,37 | 142.090 | 2,51 | 220 |

## 7.2 Using the keyword analysis to detect qualitative features

As a result of the keyword analysis presented above, we were prompted to search for concrete examples of the features which the numerical patterns suggest were frequent in the consultations. Even more intriguing, examining the presence of these features in terms of where they occurred in the consultation and focusing on when they occurred most often, suggested that many of them related to particular phases in the consultation. Some are found in the process of assessment, others in the phase of advice giving and others are predominantly used to wrap up the conversation. Thus, the following exposition of features we have identified as a result of the keyword analysis will follow the sequence of an imaginary consultation between a health professional and a client showing the predominant features in the order in which they are likely to take place.

## 7.3 Securing the consultation: fixing the caller and credentialing the advice

On the face of it, the consultations between callers to NHS Direct and the health advisers and nurses appear to involve the listing of symptoms and screening for potentially serious problems such as meningitis. Moreover, the very fact that the caller might usually be expected to have called to obtain advice would imply that most of the consultation would be about the caller. However, examining some of the features revealed in the corpus analysis suggests that these are not simply a set of screening questions, inventories or assessments. They also have a crucial role in establishing a relationship in situ between caller and adviser, and in establishing the severity of the symptoms, both formally through description and in terms of the way they are talked about. Another feature of the consultations, especially in their early stages, is to further secure the consultation to the caller. This may seem obvious, but involving and enlisting the participation or involvement of the recipients of health advice is by no means automatic or straightforward. In the consultations in the present study there were two major classes of technique which were noticeable in the early stages of a consultation: to involve the caller and to elicit a set of symptoms logically ordered in terms of the classes of problem which NHS Direct are able to deal with.

### i) Hearer involvement

The first major class of devices used to secure the hearer's involvement was detectable through the use of personal pronouns 'you' and 'your' which were amongst the fifteen most significantly frequent items disclosed by the keyword analysis. This implied the interactions were strongly centred on the caller. Moving from the word frequency counts to an examination of context of occurrence disclosed that the term 'you' was used extensively when giving

instructions, identifying courses of action and using colloquial forms. The following extract illustrates these features:

HA:    Yeah, you see you have to do the whole course, you see. Right. What I'm gonna do is just take some details of you for our confidential files.

P:     Eh ha

HA:    If I may, and then get a nurse to call you back it will be

P:     OK

HA:    Approximately around about 40, 45 minutes at the moment. Or, a little later

HA:    […] Thank you very much. Right, have you called us before about yourself?

The use of 'you' then, helps to maintain the focus relentlessly on the client and the client's actions. Intriguingly, there are some asymmetries. The use of 'you' is less intense in the callers' speech; they tend, for example, to use it a good deal as a tag – 'you know':

P:     I was just wondering if it could be an allergy should, I mean what should I do, first to get tested obviously I hope, you know phew.

The 'you knows' in this context seem to have a function as tag questions, mitigators, or codas to the caller's turns of the kind described by Lakoff (1973) as a characteristic of disempowered speech styles, perhaps reflecting the asymmetries in power and knowledge between professional and layperson. In any event, the use of 'you know' in the caller's discourse is dwarfed by comparison with the overwhelming use of the term in the professionals' corpus. The frequent use of 'you', even when the callers are not specifically being told to do something, is believed to be a feature of persuasive discourse (Storey, 1997) and serves to secure the presence of the caller as an object of scrutiny and the subject of future advice. The use of 'you' then is a kind of anchoring device.

## ii) Modalizers and logical operators

Politeness markers

In any type of discourse politeness can be achieved through various communicative strategies and the use of a range of lexical items.

The number of backchannel responses which signal accord and active listenership is significantly higher in the health professional corpus compared with the corpus of general English. Items such as 'OK', 'okay' and 'right' are typical examples which arguably could indicate that health professionals are allowing the patient to explain their symptoms and invite further elaboration.

The increased use of modal items is a further way of marking politeness in discourse. Harris (2003) notes how modalizers and mitigators can be used in asymmetric situations by the professional, perhaps as a means of minimising the threats to 'face' inherent in categorical statements. Modal terms such as 'can' and 'may' introduce optionality into the conversation and thus give the appearance of allowing the patient to make their own decision on whether or not to follow the advice that is given. Below are some concordance lines from the health professional corpus which include the modal term 'may'.

```
   And they also say cool baths may help itching and just gently pat your skin and em a
             It may be that there may be some other course for it.
it said taking with this medication may cause flushing nausea vomiting abdominal pain
          and diarrhoea and rashes may also occur.
                     Tetracycline may discolour developing teeth if it is taken by children
                         They may dry the skin out and they make the itching worse.
         stopping it tonight may not  reduce your symptoms tonight
                 I think you may find useful and there is sort of sort of one and a half
   finger's actually improving you may still need to have a course the course of
```

These examples show that 'may' is used mainly to soften the more or less categorical listing of side effects of certain treatments or conditions or to suggest further action on the part of the patient. As such it serves a dual role as an epistemic softener and perhaps less obviously as a politeness device.

The terms 'if' and 'or' as modalizers and logical operators

The high frequency of the word 'if' signals a similar tendency to 'may'. By introducing hypotheticality into the discourse it creates options for the patient and it also softens or mitigates any advice that is given. Thus it has some allegiances with politeness phenomena. 'If' is also a term which is used in the diagnostic and screening procedure. In this respect it resembles a logical term. For example, the rash in meningitis tends not to fade under pressure. A client phoning in with a rash yielded the following question:

N:    If you push on them [the spots] do they fade and come back again, the rash?

P:    Em, yeah.

The 'if' here is a kind of invitation to perhaps investigate and further refine the reports of symptoms. As part of meningitis screening this is pivotal in the sequencing of further interaction.

There are other uses for the term 'if'. One of these kinds of use is the chaining together of possible events into a logical sequence. This is rather like syllogistic reasoning in formal logic. Here, the caller is describing an earache and an encounter with her GP:

P:      […] he made it sound quite scary. He's made it sound like my my my ear was going to explode or something.

D:      That's always a possibility, that the eardrum does burst if it were if that were to happen

P:      Em

D:      It's just the infection they usually heal anyway

P:      Yeah

Here the causal chain effectively downgrades the potential difficulty of a burst eardrum or a badly infected ear. The possibly catastrophic event is headed off with an 'if' statement leading to the assertion that it is 'just' the infection and that healing will usually take place. Thus, the source of concern is downgraded. More speculatively, we could see this use of 'if' as being part of a system of emotional management that has been noted in other studies of health professionals (Dube et al., 2003) and their commentary on symptoms, and accords with Ferguson's (2000) account of 'if conditionals' as politeness strategies.

A further use of 'if' as a conditional term occurs, for example, in the phrases:

D:      And if necessary get off to the emergency surgery

Or alternatively

D:      [...] If you are in pain in the morning see Dr Carl or whoever.

'If' in such cases is part of a process of adding coherence to the illness experience, suggesting that in the worst case scenario – that the pain continues or increases – then this is manageable and there is a course of action to be taken to remedy it.

The use of the word 'or' has some similar characteristics to 'if'. 'Or' appeared frequently in the corpus of material from NHS Direct personnel. The frequent use of the word 'or', in particular as part of a binominal, is striking and adds to the overall impression that the patient is being offered a range of possible scenarios that may apply to them. The examples below illustrate this:

N:      Right okay. What about any deep burning or aching pain in a band … around …

N:      Are you going hot and cold or sweating or feeling clammy?

N:      Do you feel confused or disorientated?

A common phrase that recurred in this context is the phrase 'or anything', a vague expression mainly used as a tag question which again leaves room for the patient to add their own description of the situation:

N:      And so there's no swelling anywhere to your face or anything?

This apparent vagueness encoded by means of language items features prominently in the whole health professionals' corpus. It may serve as a deference strategy by softening the imposition on the caller and leaving room for elaboration or retraction from any particular question or suggestion. It may also casualize the symptom reports so as to downgrade their seriousness. For example, one would not say, 'Are you having a heart attack or anything?' Yet one might say, 'Are you coughing or anything?' The vagueness represents a marked invitation to the caller to disambiguate the previous utterance and clarify the nature of the symptoms. The disambiguation is invited from the caller after the nurse or adviser has already listed one or more symptoms. This can be seen as a way of minimising the intrusion represented by the questions – the basic shape or form which might be taken by the possible symptoms is preformulated as if common knowledge between caller and nurse and the caller is merely being called upon to clarify an already known situation.

Thus, the impression of politeness and professionalism gained during the fieldwork is one which is sustained by the subsequent analysis of a variety of politeness markers used by the staff in the elicitation of symptoms.

The elicitation of symptoms in itself serves to focus the sequence of interaction even more strongly on the caller as an object of scrutiny. This asymmetry is a mundane but nevertheless notable feature of health care encounters, in that the clients' problems are described but typically no such disclosure is made by the professional.

Once the symptoms have been asymmetrically elicited in this way, it might be appropriate to offer some advice. Again, there were particular conventions used to dispense knowledge. In this case they seemed to serve the function of making it more credible and perhaps more suitable for following.

### iii) Credentialing: the deployment of sources of authority

Once involvement and focus have been secured, it is then the task of the adviser or nurse to make the advice appear authoritative. One of the strategies frequently used by the health professionals was the depersonalisation of advice or information by referring to prestigious secondary sources or third parties. In this context it is interesting to note that the lexical item 'able' is used most frequently in the construction 'They'll be able to advise/confirm/look up …' Other phrases include 'It says here …' or 'The question here is …' where the nurse reads out instructions from another source of information. These strategies may successfully secure the advice to an external source of authority. That is, they are a way of saying that it is more than just one person's opinion – it is the considered view of prestigious bodies or individuals.

Let us look at how external sources of information are deployed. Consider the following example, where the caller is concerned with whether it is possible to drink alcohol whilst taking antibiotics:

HA:    Here you're there now you're just interested in how much alcohol would be safe to drink with metronidazole

P:     Yeah, yeah

HA:    Okay now I've had a look at two sources of information for you. One of them is the British Medical Association their new guide to medicine and drugs.

P:     Eh ha

HA:    Now under the alcohol chapter it does suggest that you should avoid it really it said taking with this medication may cause flushing, nausea, vomiting, abdominal pain or headache and I also checked it on the British National Formulary which is a drug interaction checker.

P:     Yeah

HA:    And they also said that you'd get a reaction there as well eh so you need to have to be aware if you were to drink then it's probable

P:     Right

HA:    They'll react badly together and sort of give you those symptoms

P:     Right

HA:    And it doesn't really say if there is a safe limit, it's just to avoid altogether really.

In this sequence the sources of authority are combined to provide a synergistic prohibition. The individual contributions are modalized by the terms used to describe their claims. The British Medical Association guide 'suggests' whereas the British National Formulary says it is 'probable' – both terms used usually to mitigate the strength of a claim – yet the overall cumulative weight of the recommendations is to 'avoid it altogether'. Indeed, a third source of authority is added later in the interaction:

HA:    You know you could always check with another pharmacist …

But the degree of closure imposed by 'altogether' implies that the result of further inquiries would be redundant – you could ask but you would get the same answer.

Thus, the credibility of the sources of advice is anchored to concrete items such as books, which are described in some detail – even complementary therapies were anchored to a book called 'medicinal herbs' – which helps to foreground the presence of this authority in the conversation, as if they were actors who speak. The use of sources of authority in book form in this way

is rather quaint in some respects, especially in an age of telemedicine when clinicians and researchers themselves are just as likely to use online databases. Nevertheless, it is potent in that it reflects the cultural authority of the written word. Thus, the authority of the advice is established. Furthermore, it helps to sustain the continuity of the interaction, ensuring that there are no gaps in the conversation whilst the material in question is consulted, as these are typically less well tolerated in telephone conversations than when the parties are co-present.

Moreover, this further consolidates the asymmetries we have mentioned earlier, where the focus upon the caller is maintained by the nurse and the complaint is elicited from the caller symptom by symptom. The nurse in this scenario is not necessarily the source of advice, nor even of some of the assessment questions. Instead, they are subtly effaced as agents in the scenario.

### iv) Convergence codas

Having given advice, the nurses and advisers show interest in whether the caller was at all likely to follow it. There was in many of the interactions a sequence at the end which involved a kind of summary of what had been achieved so far so as to encourage the adoption of a course of action. This is often seen in the form of a 'convergence coda' at the end of a stretch of interaction shortly before the phone is put down. For example:

HA:    […] I certainly learned something by speaking to you tonight. But certainly yeah like I say it may be you know that you might find something helpful

P:    Yeah

HA:    In the things I am about to send to you.

HA:    But if not it's always worthwhile popping them just perhaps to see another GP.

P:    Yeah

HA:    To see if there is anything else they can do for him, is it OK

P:    I'll try that then, great.

HA:    No problem, I'll pop in the post to your work then

P:    Lovely

HA:    OK

P:    Thanks a lot.

HA:    No problem, bye

P:    Bye.

In this particular conversation the adviser had addressed the issue of why the caller's husband was getting recurrent earaches and the advisability of seeking a referral for further investigation from his regular GP. In this extract there is a first position invitation to convergence at the suggestion that an opinion from another GP in the practice be sought, yet this yields a 'yeah' rather than an active commitment to do something. Nevertheless, the neat embedding of the 'yeah' in the audio-taped record suggests it is in this case more than merely a backchannel agreement. It is followed by a second position invitation 'To see if there is anything they can do for him, is it OK' yields the active agreement to try this approach. Then, the termination sequence can proceed. This kind of termination sequence, where the health adviser, nurse or doctor actively encourages a vocal assent on the part of the client to perform some course of action, may have implications for the study of compliance or concordance. It offers a reprise of the necessary actions on the part of the professional and the client and frequently seemed to involve a progressive alignment of client and professional – hence the term 'convergence codas'. Like many of the 'closings' characterised by Schegloff and Sacks (1973) they contain in miniature a summary of the events of the preceding conversation.

## 8   Suggestions for further research

The study of these conversations between NHS Direct staff and role-playing clients described above was necessarily limited in terms of the number of transactions and the type of scenarios that the researchers presented in their phone-ins. Yet despite these limitations and the staged nature of the discussions, a corpus linguistic analysis has disclosed a number of features which seem to be characteristic of these kinds of interactions and which lead to new ideas as to how concepts such as hearer involvement, modalizing, credentialing and the convergence coda are deployed in the interaction. These ideas have implications for the further study of health care encounters and for education and training in the emerging field of applied clinical linguistics. They also provide us with some clues as to the recurrent features of health care communication over and above what we can find in models of good consultation practice which have been derived a priori. Thus, a 'data driven' learning approach could alert practitioners, educators and researchers to features of accomplished professional practice which were not hitherto obvious. Once they have been rendered visible in this way, some debate can follow as to whether these were desirable features, perhaps by relating them to subsequent outcomes, and if so how they can be refined and improved.

As we have seen, these features include recurrent politeness strategies, as well as methods of involving and enlisting the callers into the consultation. Further research in applied linguistics could address the following issues:

- a more detailed analysis of linguistic patterning in the language of health care professionals;

- an analysis of the language of emergency calls where it may be inappropriate for the health professional to be vague and where politeness strategies may have to compete with concerns over efficiency of information transfer;

- an analysis of more complex encounters such as phone-ins related to mental health issues;

- an analysis of the effect of strategies of patient empowerment on issues of compliance with a prescribed course of action suggested to the patient;

- an analysis and categorisation of the types of questions asked by the health care professionals, as well as the types of answers they yield.

There is scope, then, for further investigation into the language of nurses, doctors and health advisers working at NHS Direct. The types of analysis described above would add not only to our understanding of health communication in general but they have potential for generating guidelines for best communicative practice.

A future stage of our work would be to see if any of the linguistic patterns we have identified would correspond to actual behaviour in the world outside the health care encounter. In this way it might be possible to address the impact that certain styles of communication have on the level of adherence to the advice or the recommended treatment regime. There are considerable concerns about low rates of compliance across a whole range of clinical specialisms (Claxton et al., 2001): blood pressure (Bremner, 2002), diabetes (Campbell et al., 2003) post-transplant surgery (Chisholm, 2002) and mental health (Coriss et al., 1999). Sometimes fewer than 50 per cent of patients are believed to be following the optimal course of action. While studies of compliance and concordance focusing on attitudinal and cognitive issues have not yet yielded decisive results, it is our hope that examining language and relating it to behaviour will provide some clues as to the 'compliance signatures' which distinguish those who adhere to health care advice from those who do not.

These developments would require advances to be made in corpus linguistics too. It would entail going beyond word frequency and word distribution studies and would involve the development of techniques to identify more complex, outcome-relevant linguistic patterns. Given the present state of technology, this might involve proceeding as we have done, using the software to identify patterns and manual coding to characterise their nature and conversational function, combining in other words both quantitative and more qualitative methodologies.

## 9    Users of health communication corpora

As we have noted, a good deal of existing published literature on the study of language in social settings has been dominated by discourse analysis and conversation analysis. These data are rich from a qualitative point of view and help at an exploratory level and to understand the context of usage. However, the analysis is rendered much stronger if it emerges from a larger corpus of material which allows us to see the peaks and troughs of occurrence of the devices, strategies, lexical choices, patterns, fixed expressions and phrases in comparison to general English usage. In analysing a large corpus it is possible to 'take the pulse' of the interaction under investigation so as to guide subsequent qualitative work. This is methodologically powerful because it can ground qualitative insights in a firm grasp of their regularity, frequency and significance.

The corpus linguistics approach in the study of health language has the potential also to deliver the kinds of data which would be of interest to auditors and evaluators seeking to gain a picture of practice in a particular setting or evaluate the effectiveness of educational interventions and good practice initiatives. There are many in the field of language learning who stress the need to 'present real examples only', as Sinclair (1997: 30) exhorts. However, further understanding of the teaching and learning process may be necessary, as this does not form a complete pedagogy in its own right (Widdowson, 2000). Yet, as we have argued, a data driven learning approach based in the kind of applied clinical linguistics we are advocating might have a great deal to offer an increasingly beleaguered NHS (see also Chant et al., 2002a; 2002b; Crawford et al., 1995, 1999; Brown et al., 1999; Maguire & Pitceathly, 2002). It is becoming increasingly urgent to address education and training of health service personnel. If we take the view that language is both interpersonal and transactional, it is important to consider how the language of a health care encounter is recipient-tailored (Brown & Fraser, 1979).

It is particularly important to examine the issue of health language closely at present because there are some important changes afoot in the health communication field. For example, the emphasis on working with clients and taking their views into account has gained favour with policy makers. It is through careful attention to the language of health care encounters that we will be able to document the shift from information giving to working with the patient and suggest how it might best be expedited.

Although linguists have in the past been the main users of corpora, they certainly need not be the sole users in the future. Health care providers and researchers will increasingly require access to naturalistic data which cannot be reproduced in laboratory conditions, while at the same time they are under pressure to quantify and test their theories rather than rely wholly on qualitative data.

In this paper then, through the illustrative study of NHS Direct consultations, we have attempted to highlight the possible contribution of corpus linguistics and data driven learning to the field. The kind of 'applied clinical linguistics' we have advocated highlights a convergence of some already well established traditions in the health and social sciences, but represents a novel convergence of ideas that should add to the debate about what exactly is going on in health care encounters. Moreover, it is through attention to the language used that we will be able to grasp the jointly formulated irrationality of health care and begin the process of mapping the terra incognita of spoken health care work. This will enable policy makers to establish whether indeed the politically desirable ideals in health care have been met, and it will enable practitioners to guide their interactions down the most advantageous channels so as to ensure that clients are empowered to make the most of the treatments and advice they are given.

## Notes

1    CANCODE stands for Cambridge and Nottingham Corpus of Discourse in English. CANCODE is a five million word computerised corpus of spoken English, made up of recordings from a variety of settings in the countries of the United Kingdom and Ireland. The corpus is designed with a substantial organised database giving information on participants, settings and conversational goals. CANCODE was built by Cambridge University Press and the University of Nottingham and it forms part of the Cambridge International Corpus (CIC). Sole copyright of the corpus resides with Cambridge University Press, from whom all permission to reproduce material must be obtained.

## References

Ballinger, C., Ashburn, A., Low, J. and Roderick, P. (1999) Unpacking the black box of therapy – a pilot study to describe occupational therapy and physiotherapy interventions for people with stroke. *Clinical Rehabilitation* 13(1): 301–9.

Bremner, A. D. (2002) Antihypertensive medication and quality of life – silent treatment of a silent killer. *Cardiovascular Drugs and Therapy* 16: 353–64.

Brown, B., Crawford, P., Richards, K. and Nolan, P. (1999) Holding a mirror up to caring: Language and reflective practice. *Mental Health Care* 4(3): 27–32.

Brown, P. and Fraser, C. (1979) Speech as a marker of situation. In K. R. Scherer and H. Giles (eds) *Social Markers in Speech* 33–62. Cambridge: Cambridge University Press.

Campbell, R., Pound, P., Pope, C., Britten, N., Pill, R., Morgan, M. and Donovan, J. (2003) Evaluating meta ethnography: a synthesis of qualitative research on lay experiences of diabetes and diabetes care. *Social Science and Medicine* 56(4): 671–84.

Candlin, S. (1997) Towards excellence in nursing: an analysis of the discourse of nurses and patients in assessment situations. Unpublished PhD thesis, Lancaster University.

Candlin, S. (2000) New dynamics in the nurse-patient relationship? In S. Sarangi and M. Coulthard (eds) *Discourse and Social life* 230–45. London: Longman.

Candlin, S. (2003) Issues arising when the professional workplace is the site of applied linguistic research. *Applied Linguistics* 24(3): 386–94.

Candlin, C. N. and Candlin, S. (2003) Health care communication: a problematic site for applied linguistics research. *Annual Review of Applied Linguistics* 23: 134–54.

Carter, R. and McCarthy, M. (1995) Grammar and the spoken language. *Applied Linguistics* 16(2): 141–58.

Chant, S., Jenkinson, T., Randle, J., Russell, G. and Webb, C. (2002a) Communication skills training in health care; a review of the literature. *Nurse Education Today* 22: 189–202.

Chant, S. Jenkinson, T., Randle, J. and Russell, G. (2002b) Communication skills: some problems in nursing education and practice. *Journal of Clinical Nursing* 11: 1–12.

Chisholm, M. (2002) Enhancing transplant patients' adherence to medication therapy. *Clinical Transplantation* 16: 30–8.

Chomsky, N. (1957) *Syntactic Structures*. The Hague: Mouton.

Chomsky, N. (1976) *Reflections on Language*. London: Fontana.

Chomsky, N. (1993) *Language and Thought*. London: Moyer Bell.

Claxton, A. J., Cramer, J. and Pierce, C. (2001) A systematic review of the relationship between dose regimens and medication compliance. *Clinical Therapeutics* 23(8): 1296–310.

Corriss, D. J., Smith, T. E., Hull, J. W., Lim, R. W., Pratt, S. I. and Romanelli, S. (1999) Interactive risk factors for treatment adherence in a chronic psychotic disorders population. *Psychiatry Research* 89: 269–74.

Crawford, P., Brown, B. and Nolan, P. (1998) *Communicating Care: the language of nursing*. Cheltenham: Stanley Thornes.

Crawford, P., Johnson, A.J., Brown, B. and Nolan, P. (1999) The language of mental health nursing reports: firing paper bullets? *Journal of Advanced Nursing* 29(2): 331–40.

Crawford, P., Nolan, P. and Brown, B. (1995) Linguistic entrapment: medico-nursing biographies as fictions. *Journal of Advanced Nursing* 22: 1141–8.

Drew, P., Chatwin, J. and Collins, S. (2001) Conversation analysis: a method for research into interactions between patients and health care professionals. *Health Expectations* 4: 58–70.

Dube, L., Ferland, G. and Moskowitz, D. S. (2003) *Emotional and Interpersonal Dimensions of Health Services*. New York: McGill-Queens University Press.

Elwyn, G. (2001) *Shared Decision Making: patient involvement in clinical practice*. Nijmegen: WOK.

Ferguson. G. (2000) If you pop over there: a corpus-based study of conditionals in medical discourse. *English for Specific Purposes* 20: 61–82.

Florin, D. and Rosen, R. (1999) Evaluating NHS Direct: early findings raise questions about expanding the service. *British Medical Journal* 319: 5–6.

Foster, J., Jessopp, L. and Chakraborti, S. (2003) Do callers to NHS Direct follow the advice to attend an accident and emergency department? *Emergency Medical Journal* 20: 285–8.

George, S. (2002) NHS Direct audited: customer satisfaction, but at what price? *British Medical Journal* 324: 558–9.

Harris, S. (2003) Politeness and power: making and responding to 'requests' in institutional settings. *Text* 23(1): 27–52.

Heritage, J. and Stivers, T. (1999) Online commentary in acute medical visits: a method of shaping patient expectations. *Social Science and Medicine* 49: 1501–17.

Lakoff, R. (1973) *Language and Woman's Place*. New York: Harper & Row.

Leech, G. (1994) Students' grammar, teachers' grammar, learners' grammar. In M. Bygate, A. Tonkyn and E. Williams (eds) *Grammar and the Language Teacher* 17–30. London: Prentice Hall.

Leech, G. (2000) Grammars of spoken English: new outcomes of corpus oriented research. *Language Learning* 50(4): 675–724.

Maguire, P. and Pitceathly, C. (2002) Key communication skills and how to acquire them. *British Medical Journal* 325: 697–700.

Mangione-Smith, R. Stivers, T., Elliott, M., McDonald, L. and Heritage, J. (2003) Online commentary during the physical examination: a communication tool for avoiding inappropriate antibiotic prescribing? *Social Science and Medicine* 56: 313–20.

Mattingly, C. (1994) The concept of therapeutic emplotment. *Social Science and Medicine* 38(6): 811–22.

McCarthy, M. (1998) *Spoken Language and Applied Linguistics*. Cambridge: Cambridge University Press.

Munro, J. Nicholl, J., O'Cathain, A. and Knowles, E. (2000) Impact of NHS Direct on demand for immediate care: observational study. *British Medical Journal* 321: 150–3.

Nicholl, J. and Munro, J. (2000) Systems for emergency care. *British Medical Journal* 320: 955–6.

O'Cathain, A., Webber, E., Nicholl, J., Munro, J. and Knowles, E. (2003) NHS Direct: consistency of triage outcomes. *Emergency Medical Journal* 20: 289–92.

Pilnick, A. (1998) 'Why didn't you say just that?' Dealing with issues of asymmetry, knowledge and competence in the pharmacist/ client encounter. *Sociology of Health and Illness* 20(1): 29–51.

Pilnick, A. (1999) 'Patient counselling' by pharmacists: advice, information, or instruction? *The Sociological Quarterly* 40(4): 613–22.

Porter, S. (2001) Women in a women's job: The gendered experience of nurses. In W.C. Cockerham  and M.L. Glasser (eds) *Readings in Medical Sociology* 407–19. New Jersey: Prentice-Hall.

Roberts, C. and Sarangi, S. (1999) Hybridity in gatekeeping discourse: issues of practical relevance for the researcher. In S. Sarangi and C. Roberts (eds) *Talk, Work and Institutional Order: discourse in medical, mediation and management Settings* 473–503. Berlin: Mouton de Gruyter.

Roberts, C. and Sarangi, S. (2002) Mapping and assessing medical students' interactional involvement styles with patients. In K. Spelman Miller and P. Thompson (eds) *Unity and Diversity in Language Use* 99–124. London: Continuum.

Roberts, C. and Sarangi, S. (2003) Uptake of discourse research in interprofessional settings: reporting from medical consultancy. *Applied Linguistics* 24(3): 338–59.

Schegloff, E. A. and Sacks, H. (1973) Opening up closings. *Semiotica* 7: 289–327.

Silverman, D. (1987) *Communication and Medical Practice: social relations in the clinic*. London: Sage.

Sinclair, J. M. (1997) Corpus evidence in language description. In A. Wachtman (ed.) *Teaching and Language Corpora*. London: Longman.

Storey, R. (1997) *The Art of Persuasive Communication*. Aldershot: Gower.

Stubbs, M. (2001) Texts, corpora and problems of interpretation. *Applied Linguistics* 22(2): 149–72.

Thomas, J. and Wilson, A. (1996) Methodologies for studying a corpus of doctor–patient interaction. In J. Thomas and M. Short (eds) *Using Corpora for Language Research* 92–109. London: Longman.

Widdowson, H. G. (2000) On the limitations of linguistics applied. *Applied Linguistics* 21(1): 3–25.