

## **Clinical Linguistics? Corpus Linguistics in Health Care Settings**

**Svenja Adolphs, Brian Brown, Ron Carter, Paul Crawford**  
**Centre for Health Language Research (CHLR)**

### **Abstract**

This paper draws on three strands of research and practice in language studies namely i) studies of communication in health care encounters; ii) studies of language corpora, and iii) education and training for health care professionals to propose a new discipline of 'clinical linguistics'. This new field of enquiry will draw on these existing sub fields so as to enhance our knowledge of communicative events in clinical settings. To illustrate the potential we draw upon a study of communication in telephone conversations between callers and advisors in the UK's 'NHS Direct' health advisory service. Here, the application of corpus linguistics and conversation analytic techniques has revealed several hitherto undisclosed features concerning strategies used by health advisors to fix and secure the caller as the subject of the interaction, give credentials to the advice and terminate the encounter with a 'convergence coda'. This has implications for teaching health care professionals in the future and we discuss the possibilities that a data driven approach to learning might have advantages. The new field of clinical linguistics then has the potential to offer a new source of data and theory building as well as a resource for practitioners in clinical field settings themselves.

## **Clinical Linguistics? Corpus Linguistics in Health Care Settings**

**Svenja Adolphs, Brian Brown, Ron Carter, Paul Crawford  
Centre for Health Language Research (CHLR)**

### **Introduction**

In this paper we aim to map out a new area of inquiry in the field of applied linguistics. We will make an argument that the two fields of health care research and corpus linguistics should be brought together so as to yield fresh insights in both language theory and health care studies. Whereas there is so far little data in this terrain of investigation, there are 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 and language study and suggest how they might be brought together with advantages for language study and the training of health care practitioners. Secondly, we will describe some results from a study of our own, based on role-played exercises in a telephone health care environment. Here, we will attempt to show how bodies of material or 'corpora' transcribed from recorded interactions can be subject to computerised analysis and how this can highlight new avenues of enquiry for the linguist and educator. Finally we shall map out how this can be expanded to create a new discipline of 'clinical linguistics' where material and insights from a variety of academic and professional sources can be brought together to provide a resource for both clinical practice and theory development.

### **Language, communication and health care**

The sheer volume of research on health care from a socio-linguistic perspective is phenomenal. The age old focus on doctor-patient interaction has latterly been supplemented by a diversification of enquiries into encounters between clients and nurses (Crawford et al, 1998), physiotherapists (Ballinger et al, 1999), dentists (Nettleton, 1992), pharmacists, (Pilnick 1998; 1999), occupational therapists (Mattingly, 1994), as well as a

variety of alternative practitioners. Yet despite this frenetic pace of research, much of what has been achieved has tended to focus on relatively small databases of clinical encounters or interviews and has not so far been addressed to large scale collections of data. This is despite the fact that overall, due to the popularity of this topic area; there must be many millions of words of clinical interaction languishing on cassette tape or on forgotten sectors of researchers' hard disks. The literature on health care interaction and is too vast for us to attempt a summary here. The exponential rise in publications on this subject in the last twenty years or so has involved researchers concentrating on issues such as the interactive achievement of diagnosis in clinical encounters (Korsch et al, 1968, Maynard, 1992; Wallston, 1978, Rate, 1994; Pitts, 1998); compliance with recommendations (Hussey & Gilland, 1989); controlling frame structures (Goffman 1974, Fisher 1991, Coupland et al 1994); and interactional management of encounters (Coupland et al 1994, Gill & Maynard, 1995).

There are some curious features of this literature which are worth noting. First, this literature is still dominated by studies of doctor patient interaction, as many scholars have noted (Candlin, 1997; Candlin & Candlin, 2003), despite so much health care being dispensed by professions allied to medicine rather than medics 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 ghettoised into 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, but may also reflect hierarchies of gender and professional prestige (Porter, 2001).

In addition to this narrowness, 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' (Roberts & Sarangi 2003, p. 339). There are some notable exceptions, where research has been conducted by practitioners themselves, for example Candlin (2000), Crawford et al (1998) and Elwyn (2001) but such efforts are exceptions rather than the rule.

This kind of focus has limited the field of applied linguistics. There are however some exciting possibilities for positive change. As Roberts and Sarangi (1999) indicate, there is a need for a more dialogical arrangement between 'research and researched' such that the clinical applications of research are highlighted to a much greater extent. This emphasis on dialogue between research traditions is currently modish, even Bakhtinian, and it is our intention in this paper to suggest ways in which new subfield of 'clinical linguistics' may be mapped out and colonised by fruitful new research approaches. An interdisciplinary stance which encompasses practitioners' concerns, patients' agenda's and scholarly rigour is vital if applied linguistics is to yield this new discipline of what we have chosen to call 'clinical linguistics'. That is, language research aimed at improving clinical practices rather than merely accruing academic knowledge. This kind of approach has been flagged up by other, similarly minded scholars (e.g. Candlin, 2003) who seek to contribute to the practical concerns of delivering health care as well as to theory development in linguistics or social science. Let us first examine some of the analytic tools which are currently deployed in the analysis of health care language. A good deal of what has been done previously in studies of interaction in health care has proceeded along the lines of methodologies derived from conversation analysis. It is therefore appropriate to describe this in some detail. As Drew et al (2001) characterise it, CA (conversation analysis) is a method which focuses largely on the verbal communications which people recurrently use in interacting with one another. People are, in this view attempting to produce meaningful action and to interpret the other's meaning. In Drew et al's view, there are three key features of CA:

- 1) Any utterances are considered to be performing social actions, such as maintaining agreement between the participants, finding out the reasons for the present situation and securing the interactant's identity as a creditable person.
- 2) Utterances and actions are considered to be part of sequences of action, so that what one participant says and does is occasioned by what the others have just said and done. CA thus focuses on dynamic processes of interaction from which sequences are built up.
- 3) These sequences appear to have stable patterns. How one participant acts and speaks can be shown to have regular, predictable consequences for how the other responds. Social interactions are meaningful for the participants who produce them and they have a natural, repeatable organisation that can be discovered and the analyst is interested in understanding the machinery, the rules and the structures that produce or constitute this orderliness. Moreover, as far as conversation analysts are concerned that's the only order

there is. From the point of view of ethnomethodology and especially conversation analysis, '...the primordial site of social order is found in members' use of methodical practices to produce, make sense of and thereby render accountable, features of their local circumstances ...The socially structured character of ...any enterprise undertaken by members is thus not exterior or extrinsic to their everyday workings, but interior and intrinsic, residing in the local and particular detail of practical actions undertaken by members uniquely competent to do so. (Boden and Zimmerman, 1991, ps. 6 - 7)

This then is the way that a good deal of research h into health care encounters proceeds, on the basis of a fine grained inspection of issues such as turn taking and reality construction in the clinical encounter. This kind of methodology informs, for example Heritage and Stivers' (1999) work on what they call '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. Many of these online commentaries emphasise that the problem is relatively minor or that all is well. This seems to relate to GPs unwillingness to prescribe antibiotics unless it is strictly necessary. In this way, relatively fine grained conversation analysis gives some insight into the way that realities are constructed in clinical encounters.

### **The development of corpus linguistics**

Whilst the developments we have outlined above in the study of health care were talking place, there were a number of methodological innovations in linguistics, where over the last few decades the development of large scale bodies of language has proceeded apace. A number of these large corpora have been assembled in the last few years, but the idea itself is not new. It can be traced back to the German linguist Kading, who in 1897 used a large corpus of German - 11 million words - to collate frequency distributions of letters and sequences of letters. This corpus, by size alone, is impressive for its time, and compares favourably with more modern corpora (McEnery and Wilson, 1996). In the late 20<sup>th</sup> century there have been attempts to develop a number of large corpora of language which can be electronically searched.

Early signs of the modern era of corpus linguistics are dated by McEnery and Wilson (1996) from 1960 when Quirk (1960) planned and implemented his ambitious Survey of

English Usage (SEU). At the same time, Francis and Kucera began compiling the Brown corpus, which was developed over the following two decades. These researchers were in a minority, but they were not universally regarded as peculiar and others followed their lead. In 1975 Jan Svartvik began extending the work of the SEU and the Brown corpus to construct the London-Lund corpus.

The present-day interest in corpora has been described as a 'corpus revolution' (Leech, 2000) such that an increasing number of scholars are concerned to develop large transcribed archives of the spoken English language. This offers the opportunity to probe into the 'terra incognita' of spoken language (Carter & McCarthy, 1995). Whereas conversation analysis has sometimes had the ambition to examine regular, repeatable features of interaction, it is the corpus revolution that makes this ambition possible through the availability of larger scale bodies of the spoken language.

Most famously in the UK the British National Corpus is possibly the largest collection of English language available for scholarship. Presently, according to its website (<http://info.ox.ac.uk/bnc/>) it contains about 100 million words of English. The publishers claim that it provides a unique and authoritative view of the state of contemporary English language, which has been compiled so as to represent as many different varieties of English as possible. It is largely based on written texts. About 10% of the BNC is based on spoken language, so like much language study in general, it is still dominated by the written forms of English.

At Nottingham University where the present authors are based, Nottingham has developed differently. In 1994 the Cambridge and Nottingham Corpus of Discourse in English (Cancode) began, funded by Cambridge University Press and presently contains 5 million words of transcribed data. The recordings were collected in Britain between 1995 and 2000, keyboarded by trained transcribers, coded, and stored in a computerised database which can be searched with the publishers' specially designed software. So far the material has been used for projects concerning the teaching of spoken English. Nearly three years were spent developing the corpus which is hosted at Cambridge and Nottingham Universities. It is organised so as to facilitate searching on the basis of a number of extra-linguistic features. For example, the social context of the language samples has been recorded so the corpus can be used to analyse how language varies

across different social contexts. Moreover the relationship between the speakers is also recorded, so questions of familiarity or power differences can be explored. It is possible to retrieve individual text files and identify who is speaking where and to whom. This socio-linguistic profiling which can be derived from these data is a new feature unique to the Cancode corpus.

Equipped with such tools, 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 the quantum leap in language awareness which corpus linguistics affords.

The history of linguistics can be seen as the interaction of two different traditions. The first involves scholarship based on introspective accounts, logical analysis and intuition. This tradition incorporates the germinal work of Chomsky (e.g. 1957; 1976) and has been termed the Internalised or I language tradition. The other tradition, to which corpus linguistics belongs, comes from what has been termed the externalised or E language side of the discipline (Stubbs, 2001). Corpus linguistics perhaps offers the possibility for a more fully evidence-based approach to linguistics.

As Hadley (2002) argues, this has already been observed to have had a profound effect on language teaching. Indeed according to some, there has been a major paradigm shift (Woodward 1996). Starting as far back as the mid 1980's (Swan 1985), some began to question many aspects of the way the English language was being taught to non-native speakers. This has led to an increasingly desperate search for new ways of addressing the problems encountered in language teaching. One solution, described by a number of authors, including Hadley (2002), is to make increasing use of corpora of the language being taught.

Johns and King (1991) formulate the situation even more explicitly. In their work the " . . . language-learner is also, essentially, a research worker whose learning needs to be driven by access to linguistic data - hence the term 'data-driven learning' (DDL) to describe the approach" (p. 2). Data-driven learning makes explicit use of the kinds of corpora we have just been describing where learners investigate language with concordancing software. This enables learners to isolate common patterns in authentic language samples. It has

been termed a new form of grammatical 'consciousness-raising' (Rutherford and Smith 1988) that attempts to move learners along the pedagogic continuum from product to process. While still very much a new methodology, DDL has been argued to be a powerful tool in teaching grammar successfully.

Apart from being a source of empirical teaching data, corpora can be used to look critically at existing language teaching materials. Kennedy (1987a, 1987b) has looked at ways of expressing quantification and frequency in ESL (English as a second language) textbooks. Holmes (1988) has examined ways of expressing doubt and certainty in ESL textbooks, while Mindt (1992) has looked at future time expressions in German textbooks of English. These studies analyse the relevant constructions or vocabularies, both in the sample text books and in standard English corpora and then they compare their findings between the two sets. Most studies found that there were considerable differences between what textbooks are teaching and how native speakers actually use language as evidenced in the corpora. Some textbooks gloss over important aspects of usage, or foreground less frequent stylistic choices at the expense of more common ones. Thus McEnery and Wilson (1996) argue that non-empirically based teaching materials can be misleading and that corpus studies should be used to inform the production of material so that the more common choices of usage are given more attention than those which are less common.

Whilst it is having an impact on the teaching of language, the possibilities for corpus research in health care has been relatively under-explored. Some of the possibilities were outlined originally by Thomas and Wilson (1996), in the case of doctor patient interaction, but these possibilities have not so far been exploited fully. However, the recent formation of the Centre for Health Language Research at Nottingham University has precipitated a new drive toward the emerging discipline of 'clinical linguistics' by creating a mixed research group of 'outsiders' (linguists with an interest in health care) and 'insiders' (health care professionals with expertise in discourse-based research). Perhaps, more interestingly, and innovatively, the group has begun to advance the utility of Corpus Linguistics within a multimodal research approach in the field of health care. This marks a new and potentially high impact initiative. This approach builds on early work by Crawford et al (1999) who studied a sub-corpus of mental health nursing reports. Crawford et al's paper remains relatively unknown amongst applied linguists, perhaps because as Candlin and Candlin (2003) have noted scholars and practitioners rarely 'look over the fence' to

examine what those in other disciplines are doing. We will take some steps towards remedying this deficiency in the study of health care language and argue for a rapprochement between the disciplines of health care and applied linguistics in the pages that follow.

### **Health care education: The professionals of tomorrow?**

There are a number of reasons why the study of communication in health care is timely and apposite from the point of view of practitioners, policymakers and in these litigious times, jurists and other legal professionals too.

For example, there is an increasing tendency for communication skills to be foregrounded in medical education (Makoul, 2003; Silwa et al, 2003). In the present era of devolved funding and local management of budgets, a number of for-profit organisations have been set up with the purpose of providing training and quality audits.

Allied to this, there is a growing interest in the use of teaching media other than the conventional chalk and talk of lectures, for example distance learning initiatives, role plays both with actors and real patients, interactive computer generated exercises and much more. Some communication skills training techniques have involved initiatives in training patients so that they can participate in education for practitioners (Wykurz & Kelly, 2002). This trend on the part of the professions and those who teach their members is particularly evident in the UK where the desire to refigure tomorrow's professional as a communications expert is especially strong. Thus there are a variety of initiatives in communication skills for professions allied to medicine with the intention of encouraging somehow better communication between professionals and clients and between different groups of professionals.

As teamwork and interdisciplinary liaison becomes more widespread in health care, teams become decision making bodies who may discuss diagnoses, prognoses and care plans for clients (Chant et al, 2002a; 2002b). In addition the role of the health care professional as a comforter, educator and counsellor has been emphasised in many contemporary accounts of the purpose of health care professionals. Furthermore, the debate around health care policy has further emphasised the role of the practitioner in being the front line

person who takes account of service users' views in the form of documents such as National Service Frameworks. Over the last few years the UK government has in these documents foregrounded the practice of taking users views into account across a range of fields of practice such as Mental Health Care

A growing number of practitioners are also interested in exploring the phenomenological world of the client, in some cases through the use of videos demonstrating the patient's narratives in their own words (Makoul, 2003), as well as conventional research based in ethnographic or sociolinguistic paradigms. This interest in capturing novel formats and frameworks for teaching purposes comes from health care educators seeking new ways of engaging their students and departs somewhat from the researcher's usual technique of presenting transcribed extracts or frequency tables.

However, from the inception of the present day concern with health care communication a decade or so ago, a number of concerns were raised. For example, the use of role play was not thought to be appropriate for all cultural and ethnic groups, and the styles of communicative practice preferred in most western approaches to interpersonal encounters in medicine may equally not be appropriate in other parts of the world (Curr, 1994).

These shortcomings have perhaps arisen as a result of the curriculum for communication skills being derived from a set of theoretical concerns based in the study of Western models of counselling and communicative action. Certainly, for the last decade it has often been a complaint that those who write textbooks on communication are lacking in experience of talking to patients about difficult and sometimes life threatening issues (Toghill, 1994).

These ongoing debates about education for practitioners are still current. Models of desirable professional practice tend to be displayed prominently in accounts of clinical communication skills (e.g. Gask and Usherwood, 2002) and the reader is left wondering how these idealisations might be made to fit the more chaotic experience of everyday health care encounters.

These dilemmas and difficulties might be more easily solved, we would suggest, if there were a greater reliance on what is actually done in clinical encounters. The data we have

about interactions between health professionals and clients is often the result of carefully staged investigations in specific research programmes. Consequently, a wider use of more lifelike clinical encounters for teaching purposes might be advantageous. Learning the skills of the clinical encounter is a little like learning a new language. In the same way that the use of corpus linguistics has revolutionised the study of language learning and has highlighted the way that some of what is taught in conventional curricula may well be actively misleading. This is why some contemporary scholars of language learning have been so keen to advocate a data driven learning approach. Perhaps we could make a similar plea for health care language and healthcare education.

In the teaching of language data driven learning (DDL) approaches see grammar as a flexible system of recurring and interrelated prototypes rather than a static set of rules. Leech (1994) calls this a "fuzzy" view of grammar that draws upon prototype theories of mental representations in cognitive psychology (Rosch, 1975) and is closely related to schema theory (Cook, 1997). The Prototype Theory states humans recognize reality in abstract types. For example, the prototypical dog for an American might be a brown four-legged canine with short hair, rope-like tail, long nose and standing about 50 centimetres in height. Of course, dogs vary widely from this norm, yet we still recognize them as dogs because the prototype allows for variety in the identification process. Langacker (1991) suggests that the prototype theory can be applied to grammar, and we would add that health care encounters might be structured along similar lines. Doctor Finlay's classic 'What seems to be the trouble?' does not occur in all health care encounters, yet usually we notice an opportunity for patients' concerns to be elicited, even if it takes a different form. By this token, the kinds of linguistic events one sees in health care encounters are prototypical in nature, and should not be seen as lists of inflexible rules.

This view of the grammar system of healthcare encounters is, according to Leech (1994. ps. 19-20) "organic" in nature, not mechanistic. A DDL approach suggests that learning the 'grammar' of a health care encounter might proceed like learning the grammar of a language. In this framework, researchers and educators see the learning as best achieved through consciousness raising activities rather than teaching rules for good practice. Consciousness raising is defined by Rutherford and Smith (1988) as ". . . the deliberate attempt to draw the learner's attention specifically to the formal properties of the target language" (p. 107). This is why in DDL learners are not seen simply as recipients of

knowledge, but as researchers studying the regularity of the language. Teachers help the learners' research without knowing in advance what patterns they will discover. A DDL approach expects learners to get a "feel" for the language, or by extension the healthcare encounter, by personally experiencing a focused study of the target language's organic consistencies (Chalker 1994, Johns & King, 1991).

Thus there are implications here for how health care communication might be taught in future, if corpus linguistics, the study of health care language and medical teaching are brought progressively closer together.

Learning a new discipline such as medicine or nursing may well turnout to have much in common with learning a new language, and becoming proficient in one's chosen discipline involves new ways of thinking talking and writing, as revealed in any comparison of novice with experienced practitioners (Crawford et al, 1999). Medical educators have for a long while advocated practice at communication skills involving role playing exercises, videotaping critical review and feedback as a way of refining these skills (e.g. Maguire and Pitceathly, 2002). However, the concerns medical educators have with the health care encounter are often rather different from those of linguists. Maguire and Pitceathly (2002) for example note that doctors may elicit only half of their patients' concerns and that they may explore little of the physical, social or emotional impact of the problems. Less than half of the patients' psychological morbidity may be recognised. Whilst it may seem obtuse to quantify these kinds of variables, Maguire and Pitceathly's account echoes half a century of concern about the deficiencies in health care communication.

Whereas there is a growing body of work which originates in recordings of the health care encounters, equally much of the work which emphasises communication in nursing and allied disciplines is based on retrospective reports of people's experiences rather than the actual communications themselves. Whilst post hoc reports of feelings experiences and satisfaction are important, these issues are one step removed from the encounter from which they originate. Corpus Linguistics can provide a detailed account of encounters between health care professionals and clients in terms of the language used. Healthcare is an enterprise deeply affected by language choices and strategies, spanning a wide range of problems and interactional styles. The study of any aspect of health care language would be enhanced considerably if this language use were exhaustively

characterised in qualitative, quantitative and stylistic terms in order to advance a deeper understanding of the central role that language plays in achieving health care outcomes. Greater scrutiny and awareness of language use may illuminate, interrogate and potentially transform how professionals communicate with each other and patients/ clients, how agreement on treatments is reached with patients and ensure that such treatments or interventions are sustained or maintained by willing, informed patients. Thus, we may arrive at clearer frameworks for which language strategies result in or contribute towards desired clinical outcomes.

Communication between health professionals and patients is a special type of institutional discourse where participants differ both in the stocks of knowledge and the linguistic resources available to verbalize this knowledge. In order for medical work to proceed successfully it is presumably necessary that the patients' knowledge about their body and complaints is adapted to the purposes of the medical institution through language in interaction. On the other hand it is also necessary that the professional knowledge of the health professionals is adapted to the lay knowledge of the patients. In this mutual exchange there may be a whole variety of speech actions, which each have their own functionality. A simple question such as 'how are you?' for example may be a conversation opener or a request for a display of symptoms, depending on the context (Coupland et al, 1994). Lindwall et al (2003) describe the use of communication innovations such attempts to engage with the patient's feelings thoughts and at the time they are scheduled for operations.

By building a large corpus or computational collection of key interactions in treatment provision or advice, it becomes possible to perform an in-depth analysis of vocabulary, interactional structure and reality construction. We can thereby advance an understanding of the conversational practices of the interactants as they achieve their mutual understanding or even mutual miscommunications. Once these formulations are accomplished, it is clear from previous work that their implications can have far reaching effects if they are put into practice (Crawford et al 1995; Brown et al, 1999).

The internal organisation of conversation structure and content then is a vital sphere of study in its own right. Insights from this can guide our search for the elements of an

encounter, such that an optimal course of action is established which maximises preferred clinical outcomes.

This line of research has a variety of potential applications. In order to facilitate subsequent discussion of the issues involved let us summarise some of the possible areas of enquiry as a kind of corpus linguistics 'twelve step programme'. Corpus linguistics can for example help us to:

- i) Provide a detailed description and analysis of the language of prescribing or treatment consultations, and to characterise the unique features of professional and patient language in this context.
- ii) Identify possible strategies for more effective language use in any treatment consultation with patients.
- iii) Identify linguistic difficulties between participants at the interface of a professional or technical lexicon and a lay or non-technical one.
- iv) Enhance our understanding of underlying linguistic dynamics that could influence how patients react to their treatment or medication regime.
- v) Analyse how emotions are conveyed during health professional-patient interactions and how this could affect patient concordance.
- vi) Examine features such as turn-taking, turn-length, topic control, congruence between topic and language style, interruptions, intonational information and meaning;
- vii) Determine preference structures and paralinguistic features;
- viii) Analyse the interactional processes which are undertaken by professionals and identify the linguistic strategies that professionals adopt in treatment and prescribing activity;
- ix) Investigate the strategies by which professionals attempt to secure compliance with advice and the strategies by which patients signal their acceptance of, or resistance to, that advice in the conversational encounters themselves;
- x) Determine the genre and register of consultation language which are likely to differ in systematic ways from the use of English in general;
- xi) Examine the kinds of vocabularies, 'fixed expressions' and common collocations associated with different kinds of consultation scenarios;
- xii) Identify features of 'successful' communication and offer recommendations for future training and best practice.

It is important to stress here that the overall methods of linguistic research can be applied flexibly and be used to address a whole variety of questions, topics and ideas. These aims will perhaps best be achieved by means of a willingness to work flexibly with partners in the education system, in health policy and in the commercial sector so as to explore topics of mutual interest and reach conclusions which lead to tangible benefits in terms which make sense to policymakers, patient groups, practitioners and commercial partners. Whilst within the discipline of linguistics, researchers have long recognised the value of compiling a large corpus of spoken language and subjecting this to computerised analysis to discern patterns of language use across a broad range of human social practice, such an application is a rarity at present.

Using corpus linguistics, it is possible to provide a nuanced explication of communication dynamics or 'linguistic signatures' directly associated with a variety of health care interventions or inputs. These can then be used to educate professionals and patients, potentially leading to better clinical outcomes. There is the potential of correlating these measures of language with a variety of clinical outcomes. It is hoped that the systematic study of a large body of language will explicate the forms of health care language and will yield greater insight into the meaning of health care interaction. The promise of corpus linguistics is that it will allow a detailed analysis of a variety of health care language styles and interactions, which can then be utilised in communication training programmes. This creative synthesis between health care and linguistics could provide a wide variety of health practitioners with the information they need to make substantial improvements in care delivery.

Whilst it is commonplace to advocate training in communication for health professionals, it is less easy to be specific about what this should involve. Adopting a method such as data driven learning which has an established pedigree in the teaching of language could well offer some new insights for those involved in communication skills training for health care personnel.

This then raises the question of what a learner adopting a data driven approach might discover of the lexis, grammar and choreography of a health care encounter. Therefore, in the following section of this paper we will give an account of an analysis of the language of NHS Direct consultations. This will illustrate the application of Corpus Linguistics as a key

part of a multi modal methodology which might also include discourse analytic and conversation analytic techniques. We will examine how this approach may be utilised to influence and change health care practices. Corpus linguistics can be used, then as an educational device as well as a tool for monitoring and quality control in health care.

### **Background to the NHS Direct study**

The NHS direct service has been in operation for several years now, yet we know relatively little about the fine detail of the kinds of consultations which take place within the service. Like so much of our spoken language it is a 'terra incognita'. Yet the promise of detailed study of this vast body of health-related communication is very great. There may be important insights for practitioners, policymakers 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, rigorous yet courteous and sympathetic manner. In the longer term it will be useful to examine how we as a culture 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.

### **Methods and techniques**

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 advisors 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 -11 am and 6 pm - 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 still further.

Overall seventeen calls were made by two male and two female researchers. 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 NHS Direct corpus**

Overall the callers described seventeen scenarios to NHS-Direct staff all of which have been transcribed and subjected to linguistic analysis. After recording, the tapes were sent to specialist transcribers who converted them into electronic format suitable for analysis. The overall word count of all the interactions amounts to 61,981 words. For the purpose of corpus analysis the transcripts have been split into utterances made by the health advisor/nurse (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. This reflects the success of the scripts used by the fieldworkers in eliciting assessment questions and advice from the staff at NHS Direct

### **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 was 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 to identify linguistic patterns that are unique to the language of health professionals in NHS-Direct phone-ins. This type of analysis was carried out with the use of corpus linguistic tools and statistics. 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.

## **Results and Discussion**

### **1) Keyword analysis**

A key word analysis calculates the frequency of every word in a specific corpus and compares this with the occurrence of the same words in a much larger corpus of general English. It then identifies a set of 'key words' which means that in comparison with the general corpus these words occur with a significantly higher or lower frequency. 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 interactions. We carried out a key word analysis using the health professionals' corpus and compared this with a 5 million word corpus of general spoken English (See appendix 2). The computer software provides a list of these words and a measure of how outstanding they are. A complete list of key words is provided in appendix 3.

Using the list of key words – that is, words which appear from the analysis to be used significantly more than we would expect by chance – we examined the instances where they were used in an effort to determine what exactly the interactants were doing with them. This then highlights the potential of a corpus approach to enable us to detect hitherto unexplored features of the linguistic landscape in health care.

### **2) 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 Advisors and Nurses appear to involve the listing of symptoms and screening for potentially serious problems such as meningitis. 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 between caller and advisor, and in establishing the severity of the symptoms, both formally through description and in terms of the way they are talked about. A further feature of the consultations is to 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 to involve the caller and credential the advice.

### ***j) Hearer involvement***

The first major class of devices used to secure the hearer's involvement was detectable through the use of personal pronouns. The personal pronouns 'you' and 'your' are amongst the 15 most significantly frequent items disclosed by the key word analysis, appearing over and above the frequency one would expect by chance in corpora of spoken English. This means that the interactions are very much centred on the caller. 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.

FP: Eh ha

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

FP: 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 in the sense of 'you know':

FP: 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. Some theorists such as Lakoff (1973) have described the use of tag questions 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 a feature of persuasive discourse in general and serves to secure the presence of the caller as an object of scrutiny and the subject of future advice. The use of you them is a kind of anchoring device.

***ii) Credentialing: The deployment of sources of authority***

Once the involvement and focus in the caller have been secured, it is then the task of the adviser or nurse to make the advice appear authoritative. One of the strategies that was frequently used by the health professionals in this data is the depersonalisation of advice or information by either third party advice or by referring to secondary sources. That is, it is made to seem as if knowledge, advice or information come from some prestigious source. 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 do two things. First they 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. Secondly, it is a way of mitigating the possible intrusion of personal questions. If it is clear that the NHS Direct employee is merely following a script or interview schedule, then the element of personal intrusion may be minimised.

To pursue the first of these possibilities, let us look at how external sources of information are deployed in the discourse of the nurse advisers. For example, consider the following sequence of discourse from the transcribed conversations. Here 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

FP: 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.

FP: 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.

FP: 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

FP: Right

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

FP: Right

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

In this sequence we can see the sources of authority combined to provide a synergistic prohibition. The individual contributions from the various sources of authority are themselves modalised 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 yield the same answer and that they would be redundant - you could ask but you'd 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 complimentary 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.

### ***iii) Modalisers and logical operators***

The terms 'if' and 'or' as modalisers 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 in assessing a rash, the interaction proceeded as follows:

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

FP: Em, yeah.

The 'if' here is a kind of invitation to perhaps investigate and further refine the reports of symptoms. In this case a rash that does or does not fade under pressure is part of a screening sequence of meningitis, so the piece of interaction here 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. Let us examine an example of this from a sequence of interaction. Here, the caller is describing an earache and an encounter with her GP.

FP: [. . .] 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

FP: Em

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

FP: 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 and their commentary on symptoms.

A further use of 'if' is as a conditional term, for example as 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:

Right okay. What about any deep burning or aching pain in a band <\$E> 1 sec <\\$E> around+

<\$4> Are you going hot and cold or sweating or feeling clammy?

<N>: Do you feel confused or disorientated?

A common phrase that is recurring 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, e.g.:

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 serves as a deference strategy on the part of the health professional who softens the imposition on the caller and leaves room for elaboration or retraction from any particular question or suggestion. It may also casualise 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 pre-formulated 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.

### **3) 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

FP: 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.

FP: Yeah

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

FP: I'll try that then, great.

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

FP Lovely

HA: OK

FP: Thanks a lot.

HA: No problem, bye

FP: 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. The 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.

## **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 role played 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 concepts such as hearer involvement, credentialing, modalising and the convergence coda. These ideas have implications for both the further study of health care encounters and for education and training. These patterns in communication used by the health professionals at NHS-Direct 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. It is possible to decide whether they are desirable features and if so how they can be refined and improved.

These include recurrent politeness strategies, as well as methods of patient involvement and empowerment. We would expect that further investigations of larger samples of data will yield additional results, as will the analysis of authentic phone-ins as opposed to those that are carried out by researchers. A second stage of this project is thus desirable and 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 conditions 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 advisors 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 sets of guidelines for best communicative practice in the NHS-Direct context and informing training courses and continuing professional development.

As corpus linguistics is only recently becoming used in health communication it is difficult to know how to evaluate it. Moreover, once we detect sociolinguistic features in these encounters it is an even more problematic task to decide whether they are desirable and whether they are features we wish to encourage in health care professionals in the future.

One example of a feature which language study has detected with training implications is Heritage's example of an 'online commentary' mentioned earlier (Heritage and Stivers, 1999) where the doctor comments on the observations he or she makes whilst examining the patient usually offering reassurance on the relative normality of the vital signs. In a subsequent paper Heritage and his colleagues go on to note that where an online commentary emphasises the normality of the symptoms then he or she is far less likely to prescribe antibiotics (Mangione-Smith et al, 2003). This has important policy consequences. There is much concern about the extensive use of antibiotics and its potential for encouraging microbes to evolve resistance. Thus, the use of a reassuring online commentary may help to avoid the pressure to prescribe.

With a clearer idea of what takes place in health care encounters based on the intensive study of corpora of language there may be many more such examples awaiting discovery. The next 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 issue of compliance or concordance in health care. The level of adherence to the advice or the recommended treatment regime is often quite low. There are considerable concerns about low rates of compliance across a whole range of clinical specialisms: Blood pressure (Bremner, 2002), diabetes (Campbell et al 2003) post-transplant surgery (Chisolm, 2002) and mental health (Coriss et al, 1999). Compliance rates are lower where more medication doses have to be taken (Claxton et al, 2001). Sometimes fewer than 50% of patients are believed to be

following the optimal course of action with their medication or other therapeutic recommendations. Once lifestyle issues such as diet, smoking and exercise are taken into consideration, rates of compliance with medical advice may be even more disappointing for clinicians. Whilst studies of compliance and concordance focusing on attitudinal and cognitive issues have not yet yielded decisive results it is out 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.

The interest in corpora of language originated in linguistics, but the developments at Nottingham and elsewhere are signalling that it is beginning to expand into other fields too, where its value has yet to be fully appreciated. This is why we are keen to establish the importance of clinical linguistics and a data driven approach to the study of health care communication at Nottingham is not yet appreciated. The potential for using corpora of health care language in education will be exploited by the Nottingham team so as to aid some new developments in education. At Nottingham a distance-learning MSc on Health Communication is under development, which is intended to offer this data-driven approach to the study of health care language and to act as a showcase for innovation. The aim is to develop a more reciprocal relationship between linguists and practitioners via this MSc so as to provide opportunities for linguistics findings in healthcare to be returned to the clinical setting.

It is hoped that these clinical links will facilitate the development of the Nottingham health language corpus which is presently in its infancy. The ethos of the group is non-competitive, with an emphasis on collaborative working and interpersonal support. The aim to develop a substantial health care language corpus is facilitated by Nottingham's experience with developing the CANCODE corpus. Once it is established, the Nottingham corpus, allied to a panoply of software tools to enable electronic concordancing, can offer to the health care research community is a dynamic, infinitely searchable resource for determining what health care language looks like in its natural setting. Rather than having only clues from disparate small scale studies as at present, the approach can take the study of language further quantitatively and qualitatively. Whilst a variety of corpora are emerging across the world we are not aware any so far that focus exclusively on health care.

A good deal of existing published literature on the study of language in social settings has been dominated by discourse analysis and conversation analysis. This is rich from a qualitative point of view and helps at an exploratory level and to understand context of use. A corpus can provide a numerical account of health care language which allows us to see the peaks and troughs of occurrence of whatever devices, strategies, lexical choice, patterns, fixed expressions and phrases in which the researcher is interested. Moreover the analyst can compare the pattern and frequency of the features of interest between a specific context such as a GP consultation and general English usage. In analysing a large corpus it is possible to take the pulse of the setting under investigation and indicate broad discursal patterns. This can then guide qualitative work. This is methodologically powerful because it can ground qualitative insights in a firm grasp of their regularity frequency and significance. Thus it is possible for researchers to counter the charge that communication research is idiosyncratic to the analyst. A stronger claim to validity can be made if we can show that an analysis is representative of a much more general experience.

It is the intention of the Nottingham group to focus on health care language in action rather than just at written language. It is harder to catch language in vivo, but it may well turn out that the fleeting conversations in hospital cubicles, consulting rooms and client's homes have more important clinical implications than any amount of policy, guidelines or education leaflets. It is also possible to compare language of health professionals between or across disciplines and sites.

The previous tendency within language and communication studies for a good deal of the work to be performed in a qualitative fashion. With the development of quantitative corpus linguistics there is an opportunity to show that the divisions between numbers and words are not as large as has previously been thought. The approach envisaged by the Nottingham group does not involve a dichotomy or hierarchy but instead promotes the value of complementarity. Once patterns have been demonstrated in very large corpora of millions of words, this can be a signal for qualitative research to further investigate the patterns. The examples we have presented from the study of NHS Direct project is one such and example, where an initial analysis of the corpus of material suggested points of attention for discourse and conversation analytic techniques. It is important to note the

direction of this approach however. Here it is envisaged that quantification can be a preliminary stage to qualification, rather than being added on as a kind of methodological scaffolding around qualitative research afterwards.

If it is pursued on a sufficiently large scale, 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. In this way it might be possible to audit the effectiveness of educational interventions and good practice initiative and detect the kinds of communicative strategies which might lead to either more complaints from clients or greater satisfaction.

There are a number of researchers, theorists and pedagogues in the field of language learning who stress the need to use real life examples in the teaching of new languages as Sinclair (1997, p.30) exhorts 'present real examples only'. Widdowson (2000) highlights that this cannot form a complete pedagogy in its own right and that further understanding of the teaching and learning process may be necessary. Yet as we have argued, a data driven learning approach might have a great deal to offer an increasingly beleaguered NHS. It is becoming increasingly urgent to address education and training of health service personnel. There seems to be little support for non-native speakers of English in the NHS, despite the large number of such employees in the organization. For example it is estimated that 25% of doctors do not have English as a first language. At present the situation is addressed by means of an examination, yet this is only a single point measure, and is only applied to speakers of non-EU foreign languages. However, there is considerably more to health care practice than simply being able to translate from one language to another and having the relevant professional qualifications. If we take the view that language is transactional, it is important to consider how the language of a health care encounter is 'recipient tailored' (Brown and Fraser, 1979). Indeed, it may well be that clients give different accounts of themselves in response to different health care professionals, even if the latter are following more or less the same assessment script.

Corpus Linguistics developed as a branch of applied linguistics and involves analysis of naturally occurring texts. As we have tried to show, it runs counter to the paradigm of Chomskian linguistics which has usually had a much more intuitive approach to analysing language with its search for generative grammar and deep structure. In this tradition some

theorists, such as Chomsky (1976; 1993) and Pinker (1994) have argued that there are somehow 'hard wired' cognitive and neural structures that enable the grammar and lexicon to be learned so readily; in Chomsky's phrase, a 'Language Acquisition Device' or Pinker's 'language instinct'. This is debatable however. Some critics charge Chomsky and Pinker with paying insufficient attention to the sheer diversity of the world's languages or with not considering how difficult it is to explain language in evolutionary terms (Allott, 2001). Corpus linguistics offers a way out of such impasses. By taking an avowedly bottom up empiricist approach and seeking to derive theory which is grounded in actual instances of language use corpus linguistics offers a new approach to deriving grammar based on actual language in use. Historically, many linguists have tended to support Chomsky's claim that corpora had little to offer the study of linguistics. Due to the infinite variety of language any corpus would only be a partial and therefore unrepresentative sample of language. Whereas there are still many linguists who adhere to this ethos, there are many who are turning away from the Chomskian idea that language reflects some inner universal 'mentalese' or 'language of thought'. For example at Birmingham University's Centre for Corpus Linguistics the researchers believe that meaning is socially constructed in the discourse, in texts that continually shape and reshape the interactions among people and the complex institutions they establish. It is through these texts that the members of a discourse community negotiate the meaning of what they say.

John Sinclair at the University of Birmingham started the contemporary era of the collection and analysis of large quantities of corpus data. The initial impetus for these ventures in the 1970s was not theoretical but commercial. The publishers of dictionaries, like their illustrious predecessor Dr Johnson wished to use actual examples of language as it was contemporarily spoken on which to base their dictionaries. Thus funding initially came from publishers. Whilst Sinclair has retired, Birmingham still hosts a Centre for Corpus Linguistics which has helped originate the COBUILD corpus and the Bank of English collaboration between Harper Collins and Birmingham University. The commercial publisher's corpora continue to expand and the Cambridge International Corpus supported by Cambridge University Press now boasts 600 million words, including 30 million words of Business English, 20 million words of academic English and 15 million words of 'learner English'. At Wolverhampton University the computational linguistics group has compiled the Wolverhampton Business English corpus. Thus it is clear that whilst corpora are expanding, the emphasis has hitherto been on specialist collections with implications for

commerce or language teaching. No one has so far specialised in health care language. However, the popularity of health care as an arena for small scale study there must already be a great deal of transcript in researchers' desk drawers and in neglected sectors of hard drives. This affords some intriguing opportunities for data sharing if the health care research community could be convinced of the value of compiling the existing data into a larger corpus.

The enormity of the task facing anyone seeking to build a corpus should not be underestimated. It is labour intensive, and results may not be immediately apparent as the task is complicated by the slow incremental build up of the corpus. However, a turning point comes once the corpus incorporates a million words, whereupon it is believed to be sufficient to perform 'significant and symbolically worthwhile' analysis. However, in some circumstances, smaller corpora may be equally useful for the investigation of local speech phenomena. As we have demonstrated above, a relatively small corpus of 60,000 words can be useful if the socio-cultural features are carefully mapped out.

To any researcher seeking to build a corpus of health care language there may be a variety of sources of funding available to cover the costs of equipment and transcription. There is a great deal of funding currently available for research in health care, yet a tradition of devoting this to language research has proved harder to establish. Equally, there is a good deal of interest in funding training and professional development programmes for students and staff which highlight communication, so it could be these monies which eventually fund the development of corpora for teaching purposes. Perhaps also, in the same way that publishers of dictionaries have supported the compilation of a number of corpora, publishers of textbooks might be able to sponsor the creation of health language corpora.

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 policymakers. 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 scientists 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 on qualitative data. One topic of enquiry that is germane to this kind of study of health care relates to the kinds of explanations people give for events. The question of how and why people explain phenomena is of interest from the point of view of health psychology and health promotion is that of how and why people attempt to explain things. Explanations (or *attributions*) have been important to psychologists because they reveal the ways in which people regard their environment. To obtain data for studying explanations many researchers have relied on naturally-occurring texts such as newspapers, diaries, company reports and so on. These are written texts, yet most everyday human interaction, including the explanation of health phenomena takes place through the medium of speech. To solve this problem Antaki and Naji (1987) used the London-Lund corpus (of spoken language) as a source of data for explanations in everyday conversation. They took 200,000 words of conversation and retrieved all instances of the commonest causal conjunction *because* (and its variant 'cos). An analysis of a pilot sample derived a classification scheme for the data, which was then used to classify all the explanations according to what was being explained. For example "actions of speaker or speaker's group", "general states of affairs" and so on. A frequency analysis of the explanation types in the corpus showed that explanations of general states of affairs were the most common type of explanation (33.8%) followed by actions of speaker and speaker's group (28.8%) and actions of others (17.7%). This refuted previous theories that the prototypical type of explanation is the explanation of a person's actions. As McEnery and Wilson (1996) remark, work such as this shows the potential of corpora to test and modify theories, especially where researchers are interested in naturalistic quantifiable language data. This implies that corpus techniques have great potential for investigating the kinds of explanations people tender for their illness. Allied to this there is also a potential for using this knowledge to design health care practice in a manner which suits clients' ways of thinking. Some recent work on everyday explanations for illness has disclosed a rich variety of explanatory frameworks in use by laypeople (Popay et al 2003). Explanations for ill health ranged from 'beer fags egg and chips' (p. 7) to 'worry and stress' (p. 8) and included 'worse housing, high unemployment and a lack of hope in the area. (p. 9). As with the features identified in clinical encounters themselves, it will be possible to use a

large corpus of health care language to identify the kinds of explanations people use and the role they play in the health care process.

In the light of this diverse range of possible applications for a corpus of health care language, the Nottingham Centre for Health Language Research has plans to build a shared resource for qualitative and quantitative research. This corpus will be useful in that it will combine the development of knowledge in linguistics, policy and practice in health care and facilitate cross-overs between the humanities, social sciences and professional practice. The application of corpus linguistics to the study of health care language constitutes a methodological innovation. The development of the corpus will also, as we have discussed, have great potential in facilitating a data driven learning approach to education in health communication.

In this paper then we have attempted to outline some of the limitations of existing forms of inquiry concerning health care language and highlight the possible contribution of corpus linguistics and data driven learning to the field, concluding with some remarks about how such a corpus could be developed. This kind of language study should enable the debate to proceed with some clarity 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 oral health care work. This will enable policymakers 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.

### **Acknowledgements**

To all members of the HLRG who assisted with this project.

### **References**

- Allott, R. (2001) *The natural origin of language*, Hertfordshire: Able Publishers.
- Antaki, C. and Naji, S. (1987) "Events explained in conversational "because" statements", *British Journal of Social Psychology*, 26, 119-126.

- Ballinger, C., Ashburn, A., Low, J. & 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), 301-309.
- Boden, D. & Zimmerman, D. H. (1991) *Talk and social structure: Studies in ethnomethodology and conversation analysis*, Cambridge: Polity Press.
- Bremner, A.D. (2002) Antihypertensive medication and quality of life – silent treatment of a silent killer, *Cardiovascular Drugs and Therapy*, 16, 353-364.
- Brown, B., Crawford, P., Richards, K., & Nolan, P. (1999) Holding a mirror up to caring: Language and reflective practice, *Mental Health Care*, 4, (3) 27 -32.
- Campbell, R., Pound, P., Pope, C., Britten, N., Pill, R., Morgan, M. & 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), 671-684.
- 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*. London: Longman, pp. 230-45.
- Candlin, S. (2003) Issues arising when the professional workplace is the site of applied linguistic research, *Applied Linguistics* 24 (3): 386-394.
- Candlin, C.N. & Candlin, S. (2003) Health care communication: a problematic site for applied linguistics research. *Annual Review of Applied Linguistics* 23: 134-154.
- Carter, R. & McCarthy, M. (1995) Grammar and the Spoken Language, *Applied Linguistics*, 16, (2), 141-158.
- Chalker, S. (1994) Pedagogical grammar: Principles and problems, in Bygate, M., Tonkyn, A. & Williams, E. (Eds.), *Grammar and the Language Teacher*, London: Prentice Hall, ps 31-40.
- Chant S, Jenkinson T, Randle J, Russell G & 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. & 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, 30-38.
- Chomsky, N. (1957) *Syntactic structures*, The Hague: Mouton De Gruyter.
- Chomsky, N. (1976) *Reflections on language*, London: Fontana.
- Chomsky, N. (1993) *Language and thought*, London: Moyer Bell.

- Claxton, A.J., Cramer, J., & Pierce, C. (2001) A systematic review of the relationship between dose regimens and medication compliance, *Clinical Therapeutics*, 23, (8), 1296-1310.
- Cook, G. (1997) Key concepts in ELT: Schemata, *ELT Journal*, 51, (1), 86.
- Corriss, D.J., Smith, T.E., Hull, J.W., Lim, R.W., Pratt, S.I. & Romanelli, S. (1999) Interactive risk factors for treatment adherence in a chronic psychotic disorders population, *Psychiatry Research*, 89, 269-274.
- Coupland, J., Robinson, J.D. & Coupland, N. (1994) Frame negotiation in doctor-elderly patient interactions, *Discourse and Society*, 5, 89-124.
- Crawford, P., Brown, B. & Nolan, P. (1998) *Communicating Care: The Language of Nursing*, Cheltenham: Stanley Thornes.
- Crawford, P., Johnson, A.J., Brown, B. & Nolan, P. (1999) The language of mental health nursing reports: firing paper bullets? *Journal of Advanced Nursing* 29 (2):331-340.
- Crawford, P., Nolan, P. & Brown, B. (1995) linguistic entrapment: Medico-nursing biographies as fictions, *Journal of Advanced Nursing*, 22, 1141-1148.
- Curr, D. (1994) Role play, *British Medical Journal* 308, 725.
- Drew, P., Chatwin, J. & Collins, S. (2001) Conversation analysis: A method for research into interactions between patients and health care professionals, *Health Expectations*, 4, 58-70.
- Elwyn, G. (2001) *Shared Decision Making: Patient Involvement in Clinical Practice*. Nijmegen: WOK.
- Fisher, S. (1991) A discourse of the social: Medical talk/power talk/oppositional talk, *Discourse and Society*, 2, 157 -82.
- Gask, L. & Usherwood, T. (2002) The ABC of psychological medicine: The consultation, *British Medical Journal*, 324, 1567-1569.
- Gill, V.T. & Maynard, D.W. (1995) On labelling in actual interaction: Delivering and receiving diagnoses of developmental disabilities, *Social Problems*, 42, (1) 11-37.
- Goffman, E. (1974) *Frame Analysis*, New York: Harper and Row.
- Hadley, G. (2002) Sensing the Winds of Change: An introduction to Data-driven learning, *RELC Journal*, 33, (2), 99-124.
- Heritage, J. & Stivers, T. (1999) Online commentary in acute medical visits: A method of shaping patient expectations, *Social Science and Medicine*, 49, 1501-1517.
- Holmes, J. (1988) Doubt and certainty in ESL textbooks, *Applied Linguistics*, 9, 21-44.

- Hussey, L.C. & Gilland, K. (1989) Compliance, low literacy and locus of control, *Nursing Clinics of North America*, 24, (3), 605-611.
- Johns, T.F. and King, P. (1991) *Classroom Concordancing*, Birmingham: ELR.
- Johnson, J. D. (1997) *Cancer-related information seeking*, Cresskill, N.J.: Hampton Press.
- Kennedy, G. (1987a) Expressing temporal frequency in academic English, *TESOL Quarterly*, 21, 69-86.
- Kennedy, G. (1987b) Quantification and the use of English: a case study of one aspect of the learner's task, *Applied Linguistics*, 8, 264-86.
- Korsch, B.M., Gozzi, E.K., & Francis, V. (1968) Gaps in doctor-patient communication, *Paediatrics*, 42, 855-871.
- Lakoff, R. (1973) *Language and Woman's Place*, New York: Harper & Row.
- Langacker, R.W. (1991) *Concept, Image and Symbol: The Cognitive Basis of Grammar*, Berlin: Mouton de Gruyter.
- Leech, G. (1994) Students' grammar, teachers' grammar, learners' grammar, in Bygate, M., Tonkyn, A. & Williams, E. (Eds.) *Grammar and the Language Teacher*, London: Prentice Hall, ps. 17-30.
- Leech, G. (2000) Grammars of spoken English: New Outcomes of corpus oriented research, *Language Learning*, 50, (4), 675-724.
- Lindwall, L., von Post, I. & Bergbom, I. (2003) Patients' and nurses' experience of perioperative dialogues, *Journal of Advanced Nursing*, 43, (3), 246-253.
- Maguire, P. & Pitceathly, C. (2002) Key communication skills and how to acquire them, *British Medical Journal*, 325, 697-700.
- Makoul, G. (2003) Communication skills education in medical school and beyond, *JAMA*, 289, 93.
- Mangione-Smith, R. Stivers, T., Elliott, M. McDonald, L. & Heritage, J. (2003) Online commentary during the physical examination: a communication tool for avoiding inappropriate antibiotic prescribing? *Social Science and Medicine*, 56, 313-320.
- Mattingly, C. (1994) The concept of therapeutic emplotment, *Social Science and Medicine*, 38, (6), 811-822.
- McEnery, T. & Wilson, A. (1996) *Corpus linguistics*, Edinburgh: Edinburgh University Press.
- Maynard, D.W. (1992) On clinicians co-implicating recipients' perspective in the delivery of diagnostic news. In P. Drew & J. Heritage (eds.), *Talk at Work: Interaction in Institutional Settings*. Cambridge: Cambridge University Press, pp. 331-358.

- Mindt, D. (1991) Syntactic evidence for semantic distinctions in English, in Aijmer, K. and Altenberg, B. (Eds.) (1991) *English Corpus Linguistics: Studies in Honour of Jan Svartvik*, London: Longman, pp 182-96.
- Nettleton, S. (1992) *Power, Pain and Dentistry*, Open University Press.
- Northouse, L.L. (1998) *Health Communication: Strategies for Health Professionals*, Stamford, Connecticut, London: Appleton & Lange.
- 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), 613-622.
- Pinker, S (1994) *the language instinct*, New York.: W. Morrow & Co.
- Pitts, M. (1998) The medical consultation, In Pitts, M. & Phillips, K. (Eds.) *The Psychology of Health*, Routledge, London.
- Popay, J., Bennett, S., Thomas, C., Williams, G. Gatrell, A. & Bostock, L. (2003) Beyond 'beer, fags, egg and chips'? Exploring lay understandings of social inequalities in health, *Sociology of Health & Illness*, 25, (1), 1-23.
- Porter, S. (2001) Women in a Women's Job: The Gendered Experience of Nurses, in Cockerham, W.C. & Glasser, M.L. (Eds.), *Readings in Medical Sociology*, New Jersey: Prentice-Hall, ps. 407-419.
- Quirk, R. (1960) Towards a description of English usage, *Transactions of the Philological Society*, 56, 40-61.
- Rate, P. (1994) *The doctor's communication handbook*, Oxford: Radcliffe Medical Press
- Roberts, C. & 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*. Berlin: Mouton de Gruyter. pp. 473-503.
- Roberts, C. & Sarangi, S. (2003) Uptake of discourse research in interprofessional settings: reporting from medical consultancy, *Applied Linguistics* 24 (3): 338-359.
- Rosch, E. (1975) Cognitive representations of semantic categories, *Journal of Experimental Psychology: General*. 104, 192-233.
- Rutherford, W. and Smith, M. (1988) Consciousness raising and universal grammar, in Rutherford, W. and Smith, M. (Eds.) *Grammar and Second Language Teacher: A Book of Readings*. Boston: Heinle and Heinle, ps. 107-116.

- Schegloff, E. A., & Sacks, H. (1973) Opening up closings. *Semiotica*, 7, 289–327.
- Silverman, D. (1987) *Communication and Medical Practice: Social Relations in the Clinic*, London: Sage.
- Sliwa, J., Makoul, G. & Betts, H. (2002) Rehabilitation-specific communication skills training: Improving the physician-patient relationship. *American Journal of Physical Medicine & Rehabilitation*. 81, 126-132.
- Sinclair, J.M. (1997) Corpus evidence in language description, In Wachtman, A. (Ed.) *Teaching and language corpora*, London: Longman.
- Stubbs, M. (2001) Texts, corpora and problems of interpretation, *Applied Linguistics*, 22, (2), 149-172.
- Swan, M. (1985) A critical look at the communicative approach, in *ELT Journal*, 39, (1), 2-12.
- Thomas, J. and Wilson, A. (1996) Methodologies for studying a corpus of doctor-patient interaction. In Thomas, J. and Short, M. (Eds.) *Using Corpora for Language Research*, London: Longman, pp. 92-109.
- Toghill, P.J. (1994) How to break bad news, *British Medical Journal*, 308, 144.
- Wallston, T.S. (1978) *Three biases in the processing of diagnostic information*, Psychometric Laboratory University of North Carolina, Chapel Hill, North Carolina.
- Widdowson, H.G. (2000) on the limitations of linguistics applied, *Applied Linguistics*, 21, (1), 3-25.
- Woodward, T. (1996) Paradigm shift and the language teaching profession, in Willis, J. & Willis, D.(Eds.), *Challenge and Change in Language Teaching*, Oxford: Heinemann, ps. 4-9.
- Wykurz, G. & Kelly, D. (2002) Developing the role of patients as teachers, *British Medical Journal*, 325, 818-821.