Unit 5 Case studies

5.3 Corpus Linguistics and Communication in Health Care

Sarah Atkins and Kevin Harvey, University of Nottingham

Introduction
The following study looks at the ways in which corpus linguistic methods can be employed to facilitate research on language and communication in health care. It describes the data collection and research conducted at the University of Nottingham surrounding the construction and analysis of the *Nottingham Health Communication Corpus* (NHCC). In particular, it discusses a sub-collection of this corpus, a series of transcribed NHS Direct phone call interactions, looking in more detail at some of the quantitative and qualitative linguistic analyses that have been made of these data.

Corpus studies in healthcare communication
Linguists have, for some time, recognised the value of compiling large corpora of spoken language and subjecting these to computerised analysis to discern patterns of language use across a broad range of human social practice. However, corpus linguistic research methods are a relatively new application in relation to health care texts. The benefits of corpus work is quickly being recognised by medical researchers and professionals and recently there has been a variety of insightful corpus studies into the domain of health care discourse (Thomas and Wilson 1996; Skelton and Hobbs 1999; Skelton, Wearn and Hobbs 2002; Adolphs et al 2004).

One of the principal reasons for this expansion is that corpus linguistic research can be applied flexibly to health care data and can be used to address a whole variety of questions, topics and ideas. For instance, one of the first major corpus studies of health care communication (Thomas and Wilson 1996) made use of a series of practitioner-patient exchanges totalling 1.25 million words in order to evaluate the efficacy of communication between a wide range of health professionals (including doctors, nurses, therapists) and cancer patients. Combining both quantitative and qualitative analysis, the study identified features of the professionals' language which contributed to a more informal, interpersonally-oriented style. Similarly, Skelton and colleagues' work (2002) has explored how GPs interact with patients: specifically how they refer to both patients and the medical profession and implicate patients and themselves in decision making processes during the consultation. Corpus studies such as these have enabled the identification of recurring patterns of medical interaction and, consequently, provide medical educators with linguistic information with which to inform communications training and development.

The Nottingham Health Communication Corpus
The NHCC currently consists of more than 500,000 words of transcribed interactions from a variety of health care contexts and is an ongoing, interdisciplinary corpus project involving staff from the School of English, School of Sociology and Social Policy, School of Nursing, and the Medical School at the University of Nottingham. Whilst in previous health care communication research, there has been a well-established focus on doctor-patient interactions, the NHCC seeks to diversify this research focus and consider a range of interaction types. Consequently, it is made up of a number of sub-collections containing interactions between a wide variety of participants, such as nurses, pharmacists, NHS Direct health advisers, a hospital chaplain and patients. The corpus is therefore unique in that it contains both written and spoken modes of communication in health care and represents the communicative routines of non-physician personnel.

One particular sub-section of the NHCC is a collection of transcribed NHS Direct interactions. NHS Direct is a relatively new communicative context to consider in relation to health care in the UK. It is a phone line, launched in 1998, which enables members of the public to pose their medical queries to an NHS health adviser or nurse on the telephone. In 2002, researchers at The University of Nottingham obtained permission from NHS Direct Nottingham to record phone calls to health advisers on this line (see Adolphs et al 2004). Researchers were given a pre-arranged outline of the medical query to discuss in each of the

phone calls, which largely focused on asking for medication advice to establish some consistency and comparability between the interactions. The NHS Direct health advisers and nurses were unaware which of their callers were researchers and calls were made from a variety of phone numbers and addresses in the Midlands to further conceal their identity. In total, a 61,981-word corpus of dialogue was collected and transcribed into an electronic format suitable for analysis. To assist further analysis, the corpus was split into the dialogue of the health care professionals (35,014 words) and the dialogue of patients (26,967 words). Although these numbers are relatively small compared to many corpora, the specialised nature of this health care dialogue made this collection sufficient for an initial, corpus linguistic investigation into the language data (Adolphs et al 2004: 13).

Analysis of the NHS Direct Interactions

Adolphs et al (2004) made a preliminary investigation of the collected NHS Direct data employing corpus linguistic methods and WordSmith Tools software. A word frequency count was obtained for the health professionals section of the NHS Direct data. A 'keyword' analysis was then made by comparing these word frequencies with the same words in the CANCODE corpus. CANCODE is a five-million word corpus held at the University of Nottingham which aims to be a representative sample of spoken English and is thus a more general corpus than the specialised interactions of the NHS Direct phone calls. The comparative analysis enabled researchers to identify words in the NHS Direct corpus which had a significantly higher or lower frequency than in the more general CANCODE corpus, that is, words which occur more or less frequently than would be expected by chance. This enabled the identification of language patterns that were specific to this more specialised health care interaction type, although highly specialised medical jargon was excluded from the keyword list (see table 5.3.1 below).

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

Table: Keyword Analysis

The keyword list features a range of linguistic elements which were resolved into the following types: imperatives ('try'), pronouns ('you', 'your'), vague language ('anything'), affirmations/positive backchannels ('right', 'okay') and directives ('avoid'). These features are more prominent in the language of NHS Direct operatives than they are in general spoken English. An initial examination of these linguistic items arguably suggests a style of discourse which is involved and interpersonal (note, for instance, the high frequency of the second person pronoun 'you' and 'your' and the high occurrence of backchannelling items). The discourse is also directive (the use of imperatives) but it would appear that, owing to the high frequency of modal verbs (such as 'may'), instances of vague language and softening words such as 'just', such direction is mitigated and courses of action are likely to be negotiated between participants rather than imposed by the health professionals. However, at this stage, this is speculation since this list of key words appears out of context (removed from the actual discourse environments in which they appear and function). In order to describe the characteristics of the practitioner's interactional style, it is necessary to examine the key words in context, to see how they operate in their discourse environment, in the consultations themselves.

Concordance lines

Concordance searches were made to look at the use of these keywords in the context of the surrounding discourse. Through this, researchers were able to see in greater depth how the NHS Direct health professionals employed these keywords in language strategies that secured the involvement of the hearer, as well as the tendency 'to use politeness and the language of convergence' (Adolphs et al 2004: 14). For example, the following is a concordance search for the word 'may', which appears frequently in the language of NHS Direct professionals;

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

One of the significant functions of this modal auxiliary 'may' is to introduce optionality into the conversation, giving the appearance of allowing the patient to make their own decision on whether or not to follow the advice given. Adolphs et al (2004) argue that 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. It also worth noting that 'may' is used in conjunction with other modalisers which encode further politeness and help prevent the operative from sounding too authoritative, a consequence which would probably result in interpersonal distance between the two parties. For example, in concordance line number 8 above, the utterance is further modalised by clause 'I think'. In this instance, the health advisor personally registers a more

tenuous level of commitment to the proposition than its absence would afford. 'I think' enables the speaker to be less categorical and therefore less imposing, while at the same personalising the utterance, with the operative referring to herself individually and more intimately, rather than as a general and impersonal voice of medical knowledge and authority.

Vague language in the health professionals' language

Building on the findings of Adolphs et al (2004), in particular the observation that there was a high frequency of vague language in the dialogue of the NHS Direct health professionals, a further, more qualitative study was made of the data. Adolphs et al (forthcoming) make a 1000-word section from the NHS Direct corpus and perform a detailed study of the vague language patterns employed. The study also employs the very broad framework for identifying vague language outlined by Channel (1994).

A particularly common phrase in the NHS Direct data, observed by Adolphs et al, is the vague expression *or anything*, 'mainly used as a tag question which …leaves room for the patient to add their own description of the situation' (2004: 19). For example, the following occurrences of *or anything* can be noted within a few lines of each other in 1000-word extract, at a stage where the health professional is attempting to elicit the patient's description of his symptoms:

NHS Nurse: Er any intense headache or mental confusion or anything?
NHS Nurse: No shortness of breath or gasping for breath or anything?
NHS Nurse: And so there's no swelling anywhere to your face or anything?

These utterances are all of a similar formation and occur frequently throughout the corpus. They typically follow the structure of listing examples of symptoms (*intense headache or mental confusion*), followed by a clause final tag question, most commonly *or anything?*. Channell (1994: 122) offers an analysis of this type of construction, where there is an EXEMPLAR followed by a TAG, arguing that they direct 'the hearer to access a set, of which the given item is a member whose characteristics will enable the hearer to identify the set'. These constructions therefore designate a category encompassing a range of medical symptoms of which the symptoms listed are understood to be prototypical examples. For example, with *any intense headache or mental confusion or anything?*', the caller is asked to consider a whole class of symptoms encompassing feelings of illness related to the head. What particularly distinguishes these EXEMPLAR + TAG formations in the NHS Direct data is that often the health professional provides two exemplars in a binomial construction, such as *any intense headache or mental confusion*. These numerous binomials account for the frequency of the word *or*, the 17th most frequent word in the NHS Direct corpus. The occurrence of such binominals signals categorical incompleteness; the two symptoms listed are 'relevantly incomplete' (Jefferson, 1990: 68) in that they do not exhaust the possible array of symptoms potentially experienced by the patient. These occurrences of the vague category identifiers, therefore, may encourage the patient to disclose symptoms which he or she might not otherwise have considered or thought relevant to the practitioner had he or she simply listed discrete sets of symptoms. Such a questioning strategy fulfils the need to elicit as much information about the patient's symptoms as possible and may preclude the danger of the patient failing to provide what might be important information.

Categorisation has long been thought to be an important cognitive process (Neisser 1987: 1) and it is 'not a surprise to discover that human language has several ways of referring to categories' as in the way outlined above (Channell 1994: 122). However, this categorization does reveal an interesting feature of the way in which symptoms of illness are conceptualised, categorising them rather generally, either in terms of parts of the body, such as the head, or particular ailments, such as swelling, rather than as discrete entities.

It has also been suggested that this vague language item, 'or anything', might serve as a deference strategy to reduce the imposition on the patient and also 'casualize the symptom reports so as to downgrade their seriousness' (Adolphs et al 2004: 20). Indeed, Overstreet (1999) has emphasised the interpersonal function of these categorizing tags, or what she calls 'general extenders', describing, for example, their use by speakers to mark an attitude towards hearers, specifically their function as a strategy of politeness and conversational

cooperation. In the NHS phone-ins, this politeness strategy would seem to accord with some of the other politeness strategies identified.

Conclusions and applications

The example analyses made above illustrate how corpus linguistic techniques can be used to direct and assist more in-depth, qualitative analyses of language corpora. With a scope and reliability of analysis not otherwise possible (Biber and Conrad, 2004), corpus linguistics is able to provide a nuanced explication of communication dynamics or 'linguistic signatures' directly associated with a variety of health care interventions or inputs. Not only can these be used to provide insights into the texture of the interaction in question, they can further be used to educate professionals and patients, potentially leading to better information exchanges and clinical outcomes. There is the potential for correlating these measures of language with a variety of clinical outcomes. It is hoped that the systematic study of a large body of language (such as the NHCC) will explicate the many and diverse 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 corpus linguistics has the potential to provide a wide variety of health practitioners with the information they need to make substantial improvements in care delivery in a range of settings.

References:

Adolphs, S, Brown, B., Carter, R., Crawford, C. and Sahota, O. (2004) 'Applying Corpus Linguistics in a health care context', *Journal of Applied Linguistics,* 1, 1: 9-28

Adolphs, S., Atkins, S., Harvey, K. (forthcoming). 'Caught between professional requirements and interpersonal needs: vague language in healthcare contexts'. In J. Cutting (ed.) *Vague Language Explored* Basingstoke: Palgrave

Biber, D. and Conrad, S. (2004) 'Corpus-Based Comparisons of Registers', in C. Coffin, A. Hewings, and K. O'Halloran (eds) *Applying English Grammar: Functional and Corpus Approaches* London: Arnold

Channell, J. (1994). *Vague Language* Oxford: Oxford University Press

Jefferson, G. (1990) 'List-construction as a task and resource' In G. Psathas (ed.) *Studies in Ethnomethodology and Conversation Analysis* Washington: University Press of America

Neisser, U, (ed.) (1987) *Concepts and conceptual development: Ecolological and intellectual bases of categorization* New York: Cambridge University Press

Overstreet, M. (1999) *Whales, Candlelight, and Stuff Like That* Oxford, New York: Oxford University Press

Skelton, J. and Hobbs, F.D.R (1999) 'Concordancing: the use of language-based research in medical communication'. *The Lancet*, 353: 108-111

Skelton, J.R., Wearn, A. M., and F.D.R. Hobbs (2002) ''I' and ' we': a concordancing analysis of doctors and patients use first person pronouns in primary care consultations ', *Family* Practice, 19, 5: 484-488

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