

## Health communication: Corpus linguistics, data driven learning and education for health professionals

Paul Crawford  
Nottingham University

Brian Brown  
De Montfort University

---

### Abstract

In this paper we describe how the advances in corpus usage in second language learning have implications for the education of health professionals. The growing understanding that practitioners in the healthcare disciplines must also be effective communicators means that education for these practitioners is ripe for innovation in educational practice. To illustrate what may be learned in this way, we describe some of our own investigations into the idiomatic use of a particular term which occurred frequently in one of our corpora, concerned with health enquiries directed at an online agony aunt. Those inquiring were frequently concerned with whether the issue they described was 'normal' yet idiomatically this was loaded with considerably more meaning than would be disclosed from a mere dictionary definition of the term. Inspection of the contexts in which it occurred suggest that 'normal' is a term drawn upon to mark major life transitions and to aid the identification of aspects of physical or mental health where medical intervention is merited. Normalising practices may be enabling, and part of what has been described as a process of 'civilizing' the body. It is through corpora such as ours that practitioners can understand the concepts employed by actual and potential clients, the vocabularies used and what they likely mean, in a fruitful convergence of data driven learning and education for healthcare practitioners.

*Keywords: health communication, corpus linguistics, data driven learning, health professionals*

---

### 1. Introduction

In this paper we explore some of the possibilities for providing an evidence-based educational experience for students and professionals in advancing their knowledge and

practice in healthcare communication. We contend that approaches which originated in language learning have the potential to inform the process of learning about healthcare language too. In particular, the enthusiasm for empirical, corpus-based language learning which has infused the field of second language instruction over the last two decades has a good deal of potential in education for healthcare practitioners who seek to enhance the language and skills of communication in healthcare contexts.

The paper argues for a much closer integration of teaching, learning, research and practice to ensure that health professionals of the future are situated in a continuous process of enquiry about health communication. This is preferable to practitioners merely recycling introductory, basic concepts and skills that they learn about in college, and an emphasis on continuing inquiry allows them to respond effectively to calls for improved quality of communication in patient care.

## 2. Literature review: Integrating experience and learning

Healthcare training often aims to make students technically proficient and enable them to master the mechanisms of health and illness. Yet they are also increasingly expected to be skilled, compassionate communicators with patients and their families, to keep themselves updated on new research for evidence based practice, and to undertake their own investigations when puzzling questions in healthcare emerge. In achieving this, educators have to contend with shifts in pedagogic practice too, which have emphasized the value of a transition to a student-centred approach. Rather than merely transmitting knowledge, the purpose of teaching has been formulated as being to help the learner develop their own active ways of engaging with the subject matter (D'Eon, 2004; Pratt, 1998). This kind of active learning is the goal and the reason for our instructional efforts (D'Eon, 2004, p. 604).

The learning of language and healthcare practice represents a kind of process which has been called 'enculturation' such that the learner comes to fulfill the norms and values within the cultural context in which they live and work (Kottak, 2004). This

process also involves individuals becoming accepted group members and fulfilling the functions and roles of the group. As part of this course of action, individuals come to know and establish a context of boundaries and accepted conduct which informs and consolidates what is acceptable and unacceptable within the framework of that social group (Grusec & Hastings, 2007) through a mechanism which has been termed social/relational learning (Mackway-Jones & Walker, 1999). As D'Eon (2004) puts it, what we are aiming for in education for healthcare professionals in interdisciplinary settings is for 'learners to adopt effective and successful strategies and where applicable shed and/or modify less effective ones in all domains of learning' (D'Eon, 2004, p. 604).

In the healthcare professions, but especially nursing, experiential learning has often been intimately related to styles of instruction that emphasize reflection and the development of the learner as a self aware being whose growing expertise is paralleled by an increasing capacity for reflection (Kolb, 1984; Ghaye & Lillyman, 1997). This is often encouraged as a means to integrate knowledge from the literature with knowledge gained from practice.

Communicative action and communication learning always take place in some kind of social context. This context can be especially important if it facilitates or inhibits learning or if it presents problems which limit the scope and effectiveness of the communication which can take place. In this section, therefore, we will deal with some educational efforts to understand that context, both from the point of view of those learning, and from that of people whose opportunities to communicate are restricted by circumstances.

The social context of learning has been seen as important by many educators, and identified as a resource that can facilitate the learning process. The milieu of learning is important in this view. Observing colleagues at work, participating in the local work culture of practice and taking part in the learning communities of the workplace all enhance the 'zone of proximal development' (Vygotsky, 1978). Thus, advice to learn by watching more experienced colleagues (Patvardhan, 2005), the literature on the value of learning communities in workplaces (Eraut, 1994; 2000) and the value attached to informal learning all address this issue. Milieu-based learning and related notions such

as situated learning (Lave & Wenger, 1991) and distributed cognition (Salomon, 1993) have valuable contributions to make to health communication learning. Since people think in and through relationships with others and use a variety of socially and culturally available tools, different cognitions will emerge in different situations, and language use will need to be carefully tailored and context-nuanced, not least to issues of power and material inequalities (Wallerstein, 1992). Yet this is only part of the picture. Those who conceive of learning as an inherently social process have been accused of presenting an overly passive model of the learner, and of minimizing the nature of the *content* of what is learned (Yakhelf, 2010). Moreover, in healthcare, the communities of learning and practice can legitimately be argued to include the clients too, as they are important contributors to the communicative environment and making sense of their concerns – however colloquially and idiomatically expressed – is an important part of their work.

We can certainly learn a lot about what is going on in healthcare through the study of the language which is used there. By learning more about the language of healthcare and the kinds of problems which people bring to healthcare practitioners, we can gain a good idea about the essential features which are important to clients and practitioners. A careful study of healthcare language has clear benefits for practitioners in training. By analogy with the process of acquiring a second language, there are possibilities for using so-called ‘data-driven learning’ where students can explore the kinds of communication that take place in healthcare settings, using data archives and large bodies of previously collected data.

The study of large bodies or ‘corpora’ of language takes its place among a number of empirical innovations in linguistics; over the last few decades, the development of such large scale databases of language has proceeded apace. The present-day interest in corpora has been described as a ‘corpus revolution’ (Leech, 2000). 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 (Psathas, 1995, p.3), it is the corpus revolution that makes this ambition possible, through the availability of larger scale bodies of

spoken language. Corpus linguistics offers the possibility for a more fully evidence-based approach to studying and learning about the uses of language in different settings. So far this has been of great interest to lexicographers, linguists and sociologists of language (Rietveld *et al.*, 2004), but is of growing importance to educators too.

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 a 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. This, allied with the deployment of concordancing software with which to examine the occurrence and usage of words have already had a profound effect on language teaching in Europe and the United States (Hadley, 2002; Varley, 2009). This is supplemented by initiatives to encourage teachers to use corpora routinely in their teaching (Breyer, 2009). According to some, there has been a major paradigm shift (Woodward, 1996).

In the healthcare context, a data-driven learning approach involves learners getting a ‘feel’ for the language of the healthcare encounter, by personally experiencing a focused study of the target language’s organic consistencies rather than a mechanistic approach to learning rules (Chalker 1994, Johns & King, 1991). The concept of data driven learning as a way of teaching was originally developed in the learning of second languages, where, rather than learn, say, how verbs behave in different tenses, it is believed to be better to enable learners to explore the language and test out their own theories about how it works, actively engaging – albeit at a distance – with the community of linguistic practice concerned. One of the early exponents of this approach, Johns (1991, p.2) wrote that 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’.

We believe that there are important lessons in data-driven learning which could productively inform both the study and the learning of health language, and which could enhance the education of healthcare practitioners. To the novice, the healthcare work

environment may well be unfamiliar. Given the pre-eminence of health as a domain of human activity, it is perhaps surprising that the field of English for specific purposes has not hitherto developed this as a major focus, compared to say, business English or – taking its cue from John Swales (1990) – genre analysis. At a more practical level, where instruction is concerned, there are much higher levels of activity. There are courses of study and textbooks for medical English (Glendenning & Holmstrom, 1987; Ribes & Ros, 2005), English for nurses (Allum & Wrathall, 2008; Grice, 2003; Parkinson & Brooker, 2004), radiology (Ribes & Ros, 2006), cardiovascular medicine (Ribes & Mejia, 2008) and pharmacy (Diaz-Gilbert, 2008). Yet many of these are concerned with the technical aspects of vocabulary and syntax one may encounter in these fields rather than the everyday language in which patients' concerns may be expressed. The challenge then is to collect more everyday, colloquial expressions of patients' ailments which may be encountered when providing services, and it is here that the study of corpora of naturally occurring language can make an important contribution.

Data-driven learning makes explicit use of the kinds of corpora we have described; learners investigate language with concordancing software. This enables them to isolate common patterns in authentic language samples. DDL has been termed a new form of grammatical 'consciousness-raising' (Rutherford & Smith, 1988). While still a new methodology, it has been argued to be a powerful tool in teaching grammar successfully. It is believed that a good deal can be learned from looking at the context in which a word occurs, and therefore there is much to be gained from studying the terms with which a target word is co-located or collocated. The elicitation of common collocations from bodies of language is seen by Wu *et al.* (2010) as a key mechanism for enhancing the knowledge of novice language learners.

Whilst it is having an impact on language teaching and learning, the possibilities for corpus research and teaching in healthcare itself have been less well explored. The academic language of medicine has been explored using a variety of corpus-related techniques. For example, academic word lists for the study of healthcare are under development (Chen & Ge, 2007; Wang *et al.*, 2008). These have been applied, for

example, to examine the collocations typically used in medical research papers (Marco, 2000), and to establish the phraseological units typically used in medical prose (Mendez-Cendon, 2009). Ferguson (2001) bridges the divide between academic writing and the more colloquial language in which healthcare is undertaken, suggesting that conditionals (expressions using 'if') are used to operationalise concepts in writing, but in face to face encounters they serve to enhance politeness and sensitivity.

This trend towards the study of the practice of healthcare communication was outlined originally by Thomas and Wilson (1996) in the case of doctor patient interaction, followed by Crawford *et al.* (1999) in the case of mental health nursing, and Adolphs *et al.* (2004) in terms of NHS Direct consultations. Wiggins (2009) adopted the approach to the study of conversation about weight management and Watermeyer and Penn (2009) apply it in exploring interactions between pharmacists and clients, and Da Silva and Dennick (2010) have examined problem-based learning discussions. Skelton and his colleagues have examined face to face activities in medicine using a number of techniques familiar to the corpus linguist, such as the role of concordancing in analysing general practice consultations (Skelton & Hobbs, 1999), metaphorical expressions used in general practice consultations (Skelton *et al.*, 2002a) and the use of pronouns in consultations (Skelton *et al.*, 2002b). What is clear from much of this exploration is that in talk with clients there are many nonstandard colloquial forms and idioms present, and these represent an additional challenge to the language learner seeking to be able to administer healthcare through the medium of English, over and above mastering the technical vocabulary applied in this field.

The parallels are even stronger if we consider that learning a new discipline such as nursing may well 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., Glendenning & Holmstrom, 1987; Maguire & Pitceathly, 2002; Roberts & Sarangi, 2002; Simpson, 1985; Wass *et al.*,

2003). However, the concerns medical educators have with the healthcare 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 might not seem meaningful to quantify these kinds of variables, Maguire and Pitceathly's account echoes half a century of concern about the deficiencies in healthcare communication.

A great many of these initiatives in education for healthcare professionals in training rely on role playing exercises. These range from students taking turns in playing practitioners and patients (as described in e.g., Glendenning & Holmstrom, 2005; Skelton, 2008), through to better-funded examples employing actors who are briefed to display particular symptoms for the exercise in question. The use of standardised patients for use in objective structured clinical examinations (OSCEs) has a long history (Harden & Gleeson, 1979; Stillman *et al.*, 1986). This means that the focus is on a limited range of educator-defined problems and forms of expression, once again often leaving out the more idiosyncratic or idiomatic ways in which distress can be expressed. Corpus explorations on the other hand can be a rich source of unexpected or idiomatic forms of language (Grant, 2005; Moon, 1998). Moon (1991, p.3) notes that the term 'idiom' can refer to a variety of features of language including a particular manner of expressing something which characterises a person or group, perhaps via a fixed expression or recurring lexeme. Idioms convey more than a strict semantic interpretation of their wording would suggest. Indeed, sometimes the meaning may be opaque to those unfamiliar with the culture within which it is used. Thus, there is clear potential to discover aspects of the language of healthcare that may otherwise remain hidden. Whilst the concept of authenticity is contested (Shomoossi & Ketabi, 2007) there is a widespread consensus supporting the use of materials which reflect the richness and complexity of the target language. A number of researchers, theorists and pedagogues in the field of language learning follow Sinclair's (1997, p.30) exhortation to 'present real examples only'.

### 3. Methodology

To provide an idea of how a corpus of language could be used in healthcare, particularly with regard to the exploration of idiomatic structures, let us take an example from a study we performed with members of the Health Language Research Group at Nottingham University, which was originally reported in Harvey (2007; 2008). Although not necessarily universally applicable, it nonetheless shows how particular healthcare facilities and projects can use research on an ongoing basis to discern the language forms that are specific to their realm of practice and their constituent groups. Our example, then, concerns a rather specific corpus that has been assembled in collaboration with the proprietors of a UK based website (<http://www.teenagehealthfreak.com>) containing health advice for adolescents. The website features an electronic equivalent of an 'agony aunt' page to which readers are invited to send messages with problems, selected ones being answered on the website. The service has been very popular, and whilst only a tiny fraction of the messages are answered the number of messages written in to the website has resulted in over a million words of emails becoming available.

The examples we shall use in this paper derive from a subsample of 400,000 words of messages inquiring about health matters (Harvey *et al.*, 2007) compared with a five million word collection of general spoken English from the Cambridge and Nottingham Corpus of Discourse in English (CANCODE), to identify 'keywords' which appeared significantly more frequently in the health message database than in the CANCODE Corpus. This pattern of word frequencies can be used to establish the 'genre' or 'register' of communication in this context, and to discern the common vocabularies, concepts and idiomatic forms of words which the website's correspondents used to characterise their problems.

#### 4. Results

Table 1 lists the absolute frequencies of the keywords salient to the teenage health messages and their corresponding frequencies in the CANCODE corpus. The range of keywords provides an indication of the central health themes and verbal choices appearing in the adolescent corpus and is thus able to identify patterns of communicative style that are unique to this particular health language context.

It is no surprise that in the teenage email corpus first person pronouns and connected terms like 'am' are very frequently observed. What is interesting though is the sheer frequency of terms related to sexual health. Although the website solicits input across a whole range of health issues, it is sexuality and reproductive health that predominates in the emails, and terms like 'penis', 'pregnant', 'period' and 'gay' feature far more frequently than they do in English as a whole. A few other oddities relate to the website itself. 'Ann' corresponds to the character 'Dr Ann' who acts as the agony aunt, and 'quiz' corresponds to the quizzes the website presents as a means of providing educational material - 'So you think you know about AIDS?' and so on.

The analysis of corpus material allows a process of 'drilling down' to address progressively more detailed and contextualised levels of analysis once the main features of the terrain have been apprehended. To illustrate our concern with the idiomatic nature of language, let us take as an example the word 'normal' which appears more frequently in our corpus than in English as a whole. Some examples of how the authors of the messages used the terms are given in Table 2. The original text has not been corrected for grammar, word choice, spelling or punctuation.

Table 1 *Frequencies of words in the 'teenage health freak' corpus compared to the CANCODE corpus*

Rank	Word	Teenage corpus		CANCODE corpus	
		Frequency	%	Frequency	%
1	My	9,775	2.55	24,435	0.47
2	I	25,287	6.59	167,377	3.24
3	Am	3,594	0.94	4,691	0.09
4	Sex	3,208	0.84	3,421	0.07
5	Im	2,234	0.58	2,265	0.04
6	Me	4,659	1.21	19,068	0.37
7	Penis	1,480	0.39	1,508	0.03
8	Help	1,834	0.48	3,163	0.06
9	Quiz	1,273	0.33	1,293	0.03
10	Ann	1,184	0.31	1,335	0.03
11	Don't	1,110	0.29	1,110	0.02
12	Asked	1,143	0.38	2,353	0.02
13	Pregnant	1,092	0.28	1,214	0.02
14	Question	1,374	0.36	2,325	0.02
15	Have	6,237	1.63	39,324	0.76
16	Is	6,924	1.81	47,740	0.92
17	Do	5,655	1.47	36,146	0.7
18	Boyfriend	848	0.22	965	0.02
19	Bullying	785	0.20	793	0.02
20	Period	884	0.23	1,155	0.02
21	Dr	821	0.21	958	0.02
22	Please	1,265	0.33	2,943	0.06
23	Drugs	757	0.20	980	0.02
24	How	2,377	0.62	11,492	0.22
25	Worried	744	0.19	1,193	0.02
26	U	743	0.19	1,193	0.02
27	Gay	616	0.16	727	0.01
28	Normal	746	0.19	1,291	0.02

Table 2 *Randomly selected concordance lines of 'normal'*


---

I'm 12, I'm 5, 3 ft and 42kg is this a <b>normal</b> weight or is it too light? or being flat chested. worried boobs arent <b>normal</b> size. episodes of Bulimia and oft seen that is normal. But i dont want to b <b>normal</b> i want to be thin. I find it insulting 14 and i havent started my period am i <b>normal</b> 8687. i masterbate 8688. i use sick for no apparant reason 13583. is it <b>normal</b> to miss a period for 3 months i thinking about becoming a transexual. Is it <b>normal</b> to do this? i still have alot of discharge is this <b>normal</b> ?? been a little depressed recently. Is this <b>normal</b> ? ward and i havent got any pubic hair am i <b>normal</b> ? 19030. i am 13 and my name I this white stuff in my under wear, am i <b>normal</b> or do i have a disease
---

---

From our point of view, the term normal is interesting, because, in line with Moon's (1998) account of idiomatic usage, we can see from the contexts in which it occurs that it means somewhat more than merely being usual, average or 'the norm'. It does of course refer to a variety of mood, sex or age related norms, such as when it is usual to begin one's periods or how regular they should be, presumably when one is in the early stages of maturity. Equally, one might say that the term normal is loaded with other meanings and positive valuations. Asking about the 'normality' of relatively rare experiences, such as being a transsexual or other things which they perceive to be unusual such as not having pubic hair, is a question about more than actuarial likelihood. In his work on discipline, regulation and the production of social and scientific order, Foucault (e.g. 1979, 1980) described what he called 'normalising practices' where particular sets of behaviours are learned and regulated through a process of comparison. There are a variety of disciplines that may be involved in this process of normalisation, for example education, professional training, the hospital system and various other systems that are capable of setting normative standards for health and wellbeing. As Hardin (2001) notes 'rewards...are measured out for behaviours considered socially permissible. Paradoxically, one of those rewards is being considered normal. To avoid the consequences of being labelled abnormal, and to secure rewards in our culture, people learn to monitor their interactions within a prescribed set of normative standards.' (Hardin, 2001, p. 16). In this context a question about whether something is normal could even denote a process of seeking alternatives to identities which have

previously been characterised as deviant and unhealthy (Cohler & Hammack, 2007). Equally, the 'normal' question can disclose something important about the asker's expectations concerning their body. That is, to someone concerned about their body's secretions, 'normal' might connote an imagined normative cleanliness and dryness whereas to someone concerned about the onset of periods the normativity might refer to the desirability of not being left out when one's peers begin menstruating. If we follow Flowerdew's (2009) advice and consider longer meaning units, showing more of the contexts in which 'normal' occurred, its function is clearer in correspondents' questions, a selection of which are presented in Table 3.

Table 3 *Questions from the teenage health corpus using the with the word 'normal'*


---

I am addicted to cerial. Is this normal?
hi only one of my balls have dropped, is this normal? will the other one drop in time?
Dear Dr. Ann, My cum is not white, its clear with some white in it. Do i have a problem? Is this normal? Whata can i do to change this?
Hi dr A. i fink iv started my periods but im not sure coz it woz brown. i told my mum and she sed i had started and it was normal 4 my fist period. but im stil worried coz sumtimes ders brown stuf and sumtimes ders nowt der. plz help
Hey I am a 16 female, When me and my boyfriend have sex I sometimes fart at of my vigena, is this normal? right by the hole in my vagina, there is. . . a thing like a ball. Is this my g-spot or is this normal??
Dr. Ann, I have a question about my inner lips in my vagina one inner lip is small and the other inner lip is bigger and it kinda sticks out what do I do and is that normal?

---

Looking at the way normality and its opposites are formulated in these questions it appears that it is often part of a contrastive statement, such that the event or phenomenon is described and then an invited contrast or comparison is posed 'is this normal?' It may not be speculating too much to see the term 'normal' as one which indicates stance (Precht, 2003) and appraisal (Macken-Horarik, 2003). That is, when people say 'normal' in this corpus it is often in the context of something they think is

wrong. The preoccupation with being normal is not merely self absorption but part of an attentiveness towards the body which is a necessary condition of effective citizenship. Subjecting oneself to normalising practices is not merely connected with making people and problems tractable and manageable – as Foucault might propose – but is also about enabling and rendering the individual capable (Heyes, 2006). Normalisation is a process often associated with transitions in the life course, when people are subject to ‘increases in the intensity of the inspecting gaze’ (Powell & Biggs, 2000, p.9).

Being ‘normal’ in the context of teenagers seeking advice is inflected with other factors such as feeling connected and autonomous (Wisdom *et al.*, 2006) and abnormality is susceptible to medical solutions. Normalising matters of the body can be seen as a process of civilising or bettering the person who is subject to these interventions (Hansen & Philo, 2007). To ask whether something is normal is to ask whether something has gone wrong with one’s health (Horden & Street, 2007), or one’s accommodation to the process of bodily maturation. It perhaps presages a process of medically mediated reconstruction, either of the physical body or one’s biography (Horden & Street, 2007). Normality, inasmuch as it represents a kind of regime of truth or power, may be drawn stringently, such that a whole variety of relatively commonplace developmental phenomena may be judged abnormal through a process that ‘compares, differentiates, hierarchises, homogenises, excludes. In short, it normalises’ (Foucault, 1979, p.183). In doing so, it establishes the ‘frontier of the abnormal’ (Foucault, 1979, p.183).

Thus, exploring the contexts in which the term occurs and exploring what it is that people are doing with the term can tell us about the idiomatic qualities of a word, in this case ‘normal’. It is clear that it represents a great deal more than something which is merely ‘average’ and is therefore loaded with meanings over and above a lexical definition, especially perhaps in health contexts. As this example shows, it is not only more frequent in these requests for advice than in the English language as a whole, but is also deployed in a way which clears the ground for medically inflected advice and possibly intervention. Now this is not necessarily obvious to those commencing

healthcare careers, or even experienced practitioners. Using techniques of this kind, it should be possible to learn the kinds of problems which are expressed to such healthcare facilities, the concerns which users have, the concepts they deploy, the vocabulary they use, and the kinds of help they are expecting. In contrast to their reticence in face to face settings, at least judging from the messages received by Teenage Health Freak, there appears to be considerably more inclination to disclose problems online, often in graphic terms.

Work on idioms and idiomatic usage in healthcare can therefore usefully be based on the evidence of corpus research. Indeed, to anyone working in health with young people in the medium of English there could be valuable lessons to be learned. Of course, the material here, derived from a UK resource, does not necessarily tell us how for example American or Australian teenagers would talk about the issues, nor does the data collection process allow the questions and problems to be traced to particular class, ethnic or faith subgroups. This would have to wait upon the creation of corpora where researchers have proactively collected demographic and contextual information to accompany the language and annotate it.

## 5. Corpus linguistics in health research and education

What we have hoped to do here is highlight the role of corpus-based language study as a tool for research and learning activities which can be used to get researchers, practitioners and students to understand the issues in healthcare encounters. Data driven learning (DDL) encourages learners to engage with the corpus via research tasks, and is significant in that it is a pedagogical application of a research method, originating with Tim Johns’ famous contention that ‘research is too serious to be left to the researchers’ (1991, p.2). As Mishan (2004) describes it, the key feature of the methodology is this ‘aura’ of research, and this research agenda is what gives the sense of authenticity. In the initial tastes of this data we have given to health psychologists in training, nurses and counsellors there is generally a fascination with the data and a tendency to ask



questions like ‘How did they. . .?’, ‘Did anybody say . . .?’ and ‘Were there very many. . .?’ which are, of course empirical questions which they can then go on to investigate with the data itself. The kinds of questions asked by the correspondents to the website have the air of the ever popular problem features page of magazines, which gives a sense of authenticity, while the research element confers a sense of the genuine discoveries which still remain to be made concerning the use of language in healthcare. Teaching and training sessions with this data can take a variety of forms. The authors have used it to illustrate conventional lectures and presentations, and have implemented interactive sessions where students have explored all or part of the data themselves. In the latter case, students have looked for how the website’s users formulated particular problems, for example those relating to sexual health, emotional distress or suicidality. These generate considerable interest and excitement among students who are often keen to share their findings with the rest of the class in plenary sessions.

Despite this enthusiasm for the use of authentic examples and material from corpora in teaching about healthcare communication, it is important to note that there are some important differences between the material included in a corpus of language and the original. ‘The texts which are collected in a corpus have only a reflected reality’ (Widdowson, 2000, p.7), for ‘Reality [. . .] does not travel with the text’ (Widdowson, 1998, p.711-712). The reality, the authenticity, of text is tarnished by transposition (Mishan, 2004). There is a need for pedagogic mediation and a willingness to extend the material under consideration over longer meaning units than individual concordance lines if the full value of corpus approaches is to be realized (Flowerdew, 2009). In addition, as we have noted there are some research questions which it would be very difficult to address in this way without considerable further research to ascertain the demographic characteristics of the correspondents.

Yet as we have argued, a data driven learning approach might have a great deal to offer health services. Understanding the idiomatic expressions which people in the UK (and elsewhere) use to describe their health problems, grasping the lay theories and folkways they use to interpret what has gone wrong and getting a grip on the rhythms of complaint might well be facilitated by the use of corpora of healthcare language. In

the UK it is becoming increasingly urgent to come up with new ways of educating and training health service personnel, especially in the light of changes in the NHS workforce and rapid demographic, epidemiological and linguistic changes in the population. 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 Cangiano *et al.* (2009) reports that 25% of doctors do not have English as a first language and overseas care workers account for 19% of the workforce nationally, rising to 60% in London. At present the situation is addressed largely by means of third-party examinations in English such as IELTS, yet this is only a single point measure, and is only applied to speakers of non-EU languages. Changes in European regulatory processes may relax this language requirement. Consequently, exploratory exercises like the ones we have described can be integrated into curricula, and may have a role to play in educating practitioners and trainees in the idioms of particular kinds of healthcare work, over and above that offered by generic second language teaching.

As corpus research reveals, there is considerably more to health concerns and healthcare 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 healthcare encounter can be ‘recipient tailored’ (Brown & Fraser, 1979). Indeed, it may well be that clients give different accounts of themselves in response to different healthcare professionals, even if the latter are following more or less the same assessment script.

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, each of which has its 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 in attempts to engage with the patient's feelings and thoughts at the time they are scheduled for operations.

It is important to stress here that the overall methods of linguistic research can be applied and 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. This will enable exploration of 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. We are not suggesting that the corpus presented here would be applicable to everyone. The point is rather to stress the value of educators collecting corpora from the kinds of settings and reflecting the kinds of problems that practitioners will be working with.

While it is clear that the compilation of corpora is expanding, the emphasis has hitherto often been on language collections with implications for commerce or language teaching. Accordingly, with the data here, and other research materials, we have begun to compile a corpus specialised in healthcare language, the Nottingham Health Communication Corpus. However, the popularity of healthcare as an arena for small scale study means that a good deal of material must have already been collected on a 'cottage industry' basis. This affords some intriguing opportunities for data sharing, if the healthcare research community could be convinced of the value of ethically sound compiling of the existing data into a larger corpus.

It is particularly important to examine the issue of health language closely via corpus-based research at present because there are some important changes afoot in the health communication field. For example the emphasis on taking clients' views into account has gained favour with policymakers. It is through careful attention to the language of healthcare encounters that we will be able to document and enhance the shift from older models of information giving to more modern approaches to 'working with the client'. 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.

By building a large corpus of 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 arrive at their mutual understanding or occasional miscommunication. Once these formulations are put into practice it is clear from previous work that they can have far reaching effects (Crawford *et al.*, 1995; Brown *et al.*, 1999).

We argue that progress in the healthcare disciplines may be well served by taking a leaf out of the corpus linguists' book, and using a similar approach to deal with teaching and learning healthcare language. Moreover it might well be possible to link communicative styles, strategies and motifs to data concerning the effectiveness of healthcare interventions. In this way a more effective and evidence-based approach to healthcare language can be developed which will promote the best use of class time for trainees, and of scarce and expensive resources such as drugs and treatment facilities.

## 6. Concluding remarks: Learning lessons

In this paper we have attempted to highlight some new trends and possibilities in the learning of healthcare communication skills which have some promise to shift the envelope of debate about healthcare communication and the way in which it can be taught. Our account of corpus-based and data-driven learning in healthcare, and the case of adolescent health advice emails, has shown how new ways of making sense of language have a place in the education of healthcare practitioners. These techniques can extend the work of educators and raise the consciousness of practitioners.

Since corpus linguistics has only recently been applied in the study of health communication it is difficult yet to know how to evaluate it. Moreover, once we detect sociolinguistic features in healthcare encounters it is an even more problematic task to

decide whether they are desirable and whether they are features we wish to encourage in healthcare professionals in the future.

The data that many authors have so far collected concerning 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 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, 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. We wish to make a similar plea for healthcare language and healthcare education.

Healthcare education is a long-term investment with multiple beneficiaries who are not always identified as stakeholders at the outset. As Shipengrover and James (1999) note, these vague and contradictory visions can lead to difficulty in characterizing and measuring the outcomes of education in healthcare.

Finally, for ESP related to healthcare, it is vital that health language learning and practice advance beyond functionality and basic competence to an increasingly evidence-based and context specific efficacy – in other words that it contributes to advancement of and not simply attunes to current health communication practices. ESP programmes rooted in data-driven learning may not only provide a useful model for health communication learners where English is a first language but also provide a clear opportunity to influence and change complacent health communication practices in the English-speaking world. Techniques such as corpus linguistics which have been popularised to assist second language learners in English may turn out to offer some important new ideas for health care education.

Data-driven learning might yield important and unexpected opportunities for advancing healthcare communication by converging second language learning in English for the specific purpose of healthcare work, where this approach is gaining momentum, and its wider applicability in health communication learning generally. It

would be a valuable innovation if the learning of English as a second language via data-driven learning were to contribute to a more critical and evidence-based review of health language practices among practitioners with English as a first language.

### Acknowledgements

We would like to acknowledge a former PhD student, Dr Kevin Harvey who played a central role in the original study on adolescent health advice emails.

### References

- Adolphs, S., Brown, Carter, R., Crawford, P. & Sahota, O. (2004). Applying corpus linguistics in a health care context. *Journal of Applied Linguistics*, 1(1), 9-28.
- Allum, V. and Wrathall, P. (2008). *Cambridge English for nursing*, Cambridge: Cambridge University Press.
- Breyer, Y. (2009). Learning and teaching with corpora: reflections by student teachers, *Computer Assisted Language Learning*, 22(2), 153-172.
- 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.
- Brown, P. & Fraser, C. (1979). Speech as a marker of situation. In Scherer, K.R. and Giles, H. (Eds.) *Social markers in speech*, (pp. 33-62) Cambridge: Cambridge University Press
- Carter, R. & McCarthy, M. (1995). Grammar and the Spoken Language. *Applied Linguistics*, 16(2), 141-158.
- Cangiano, A., Shutes, I., Spencer, S. & Leeson, G. (2009). *Migrant Care workers in ageing societies*. Oxford: Centre on Migration, Policy and Society.
- Chalker, S. (1994). Pedagogical grammar: Principles and problems. In M. Bygate, A. Tonkyn, and E. Williams, (Eds.), *Grammar and the Language Teacher*. (pp. 31-44). London: Prentice Hall.
- Chen, Q., & Ge, G. (2007). A corpus-based lexical study on frequency and distribution of Coxhead's AWL word families in medical research articles (RAs). *English for Specific Purposes*, 26(4), 502-514.
- Cohler, B.J. & Hammack, P.L. (2007). The Psychological World of the Gay Teenager: Social Change, Narrative, and "Normality". *Journal of Youth and Adolescence*, 36, 47-59.
- Coupland, J., Robinson, J.D. & Coupland, N. (1994). Frame negotiation in doctor-elderly patient interactions. *Discourse and Society*, 5, 89-124.
- 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.
- Da Silva, A.L. & Dennick, R. (2010). Corpus analysis of problem-based learning transcripts: an exploratory study. *Medical Education*, 44, 280-288.
- D'Eon, M. (2004). A blueprint for interprofessional learning. *Medical Teacher*, 26(7), 604-609.
- Diaz-Gilbert, M. (2008) *English for Pharmacy: Writing and Oral Communication*. New York: Lippincott Williams and Wilkins.
- Eraut, M. (1994). *Developing Professional Knowledge and Competence*. London: Falmer Press.
- Eraut, M. (2000). Non-formal learning, implicit learning and tacit knowledge in professional work. In F. Coffield (Ed.) *The Necessity of Informal Learning* (pp. 12-31) Bristol: The Policy Press.
- Ferguson, G. (2001). If you pop over there: a corpus-based study of conditionals in medical discourse. *English for Specific Purposes*, 20, 61-82
- Flowerdew, L. (2009). Applying corpus linguistics to pedagogy: A critical evaluation. *International Journal of Corpus Linguistics*, 14 (3), 393-417.
- Foucault, M. (1979). *Discipline and punish*. New York: Vintage.
- Foucault, M. (1980). Body/power. In C. Gordon.(Ed.), *Power/Knowledge: Selected Interviews and Other Writings, 1972-1977*.(pp.55-62). New York: Pantheon.
- Ghaye, T. & Lillyman, S. (1997). *Learning Journals & Critical Incidents: Reflective practice for health care professionals: Key management Skills in Nursing*. Dinton, Wiltshire Quay Books, Mark Allen Publishing Group.
- Glendinning, F. & Holmstrom, R. (1987) *English in Medicine: a course in communication skills*. Cambridge: Cambridge University Press.
- Glendinning, F. & Holmstrom, R. (2005) *English in Medicine: a course in communication skills*. (3<sup>rd</sup> Edition) Cambridge: Cambridge University Press.
- Grant, L.E. (2005). Frequency of 'core idioms' in the British National Corpus (BNC). *International Journal of Corpus Linguistics*, 10(4), 429-451.
- Grice, T. (2003). *Everyday English for Nursing*. London: Balliere Tindall.
- Grusec, J.E. & Hastings, P.D. (2007). *Handbook of Socialization: Theory and Research*. New York: Guilford Press.
- Hadley, G. (2002). Sensing the Winds of Change: An introduction to Data-driven learning, *RELC Journal*, 33(2), 99-124.
- Hansen, N. & Philo, C. (2007). The normality of doing things differently: Bodies, space and disability geography. *Tijdschrift voor Economische en Sociale Geografie*, 98(4), 493-506.
- Harden, R.M. & Gleeson, F.A. (1979). Assessment of clinical competence using an observed structured clinical examination. *Medical Education*, 13, 41- 47.
- Hardin, P.K. (2001). Theory and language: Locating agency between free will and discursive marionettes. *Nursing Inquiry*, 8, 11-18.
- Harvey, K., Brown, B. & Crawford, P. (2007). 'Am I Normal?' Teenagers, Sexual Health and the Internet. *Social Science & Medicine*, 65(4), 771 - 781.
- Heyes, C. (2006). Foucault Goes to Weight Watchers. *Hypatia*, 21(2), 126-149.
- Horden, A.J. & Street, A.F. (2007). Communicating about patient sexuality and intimacy after cancer: mismatched expectations and unmet needs. *Medical Journal of Australia*, 186, 224-227.

- Johns, T.F. (1991). Should you be persuaded: Two examples of data-driven learning. in Johns, T.F. and King, P. (Eds.) *Classroom Concordancing* (pp. 1-13). Birmingham: Empirical Language Research.
- Johns, T.F. and King, P. (1991). *Classroom Concordancing*, Birmingham: Empirical Language Research.
- Kolb, D.A. (1984). *Experiential Learning: Experience as the Source of Learning and Development*. Englewood Cliffs; London: Prentice-Hall.
- Kottak, C.P. (2004). *Window on humanity*. New York: McGraw Hill.
- Langacker, R.W. (1991). *Concept, Image and Symbol: The Cognitive Basis of Grammar*. Berlin: Mouton de Gruyter.
- Lave, J. and Wenger, E. (1991). *Situated Learning. Legitimate peripheral participation*. Cambridge: Cambridge University Press.
- 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.
- Macken-Horarik, M. (2003). Appraisal and the special instructiveness of narrative. *Text*, 23(2), 285-312.
- Mackway-Jones, K. & Walker, M. (1999). *The Pocket Guide to Teaching for Medical Instructors*. London: BMJ Books.
- Maguire, P. & Pitceathly, C. (2002). Key communication skills and how to acquire them. *British Medical Journal*, 325, 697-700.
- Marco, M.J.L. (2000). Collocational frameworks in medical research papers: a genre-based study. *English for Specific Purposes*, 19, 63-86.
- Mendez-Cendon, B. (2009). Combinatorial patterns in medical case reports: an English-Spanish contrastive analysis. *Journal of Specialised Translation* 11, [http://www.jostrans.org/issue11/art\\_mendez.php](http://www.jostrans.org/issue11/art_mendez.php)
- Mishan, F. (2004). Authenticating corpora for language learning: a problem and its resolution. *ELT Journal*, 58(3), 219-227.
- Moon, R. (1998). *Fixed Expressions and Idioms in English: A Corpus-Based Approach*. Oxford: Clarendon Press.
- Parkinson, J. & Brooker, C. (2004). *Everyday English for International Nurses*. London: Churchill Livingstone.
- Patvardhan, C. (2005). Tips on breaking bad news. *British Medical Journal*, 330, 1131.
- Powell, J. & Biggs, S. (2000). Managing Old Age: The Disciplinary Web of Power, Surveillance and Normalization. *Journal of Aging and Identity*, 5(1), 3-13.
- Pratt, D.D. (1998). *Five perspectives on teaching in adult and higher education*. Malabar, FL: Krieger Publishing Company.
- Precht, K. (2003). Stance moods in spoken English: Evidentiality and affect in British and American conversation. *Text*, 23(2), 239-257.
- Psathas, G. (1995). *Conversation analysis: The study of talk in interaction*. Thousand Oaks, CA: Sage Publications Inc.
- Ribes, R. & Mejia, S. (2008). *Cardiovascular English*. New York: Springer.
- Ribes, R. & Ros, P.R. (2005). *Medical English*. New York: Springer.
- Ribes, R. & Ros, P.R. (2006). *Radiological English*. New York: Springer.
- Rietveld, T., Van Hout, R. and Ernestus, M. (2004). Pitfalls in Corpus Research. *Computers and the Humanities*, 38, 343-362.
- Roberts, C. & Sarangi, S. (2002). Mapping and assessing medical students' interactional involvement styles with

- patients. In Spelman Miller, K and Thompson, P. (Eds.) *Unity and Diversity in language Use: BAAL Studies in Applied Linguistics 17* (pp. 99 - 124) London: Continuum.
- Rutherford, W. and Smith, M. (1988). Consciousness raising and universal grammar. In Rutherford, W. and Smith, M. (Ed.) *Grammar and Second Language Teacher: A Book of Readings* (pp. 107-116) Boston: Heinle and Heinle.
- Salomon, G. (Ed.) (1993). *Distributed Cognition: Psychological and educational considerations*. Cambridge: Cambridge University Press.
- Shipengrover, J.A. & James, P.A. (1999). Measuring instructional quality in community-orientated medical education: looking into the black box. *Medical Education*, 33, 846-853.
- Shomoossi, N. & Ketabi, S. (2007). A critical look at the concept of authenticity. *Electronic Journal of Foreign Language Teaching*, 4(1), 149-155.
- Simpson, M.A. (1985). How to use role-play in medical teaching. *Medical Teacher*, 7(1), 75-82.
- Sinclair, J.M. (1997). Corpus evidence in language description. In Knowles, G., Mcenery, T., Fligelstone, S. & Wichman, A. (Eds.) *Teaching and language corpora* (pp. 27-39). London: Longman.
- Skelton, J. (2008). *Role play and clinical communication*. Oxford: Radcliffe Publishing.
- Skelton, J.R. & Hobbs, F.D. (1999). Concordancing: use of language-based research in medical communication. *The Lancet*, 353(9147), 108-111.
- Skelton, J.R., Wearn, A.M. & Hobbs, F.D. (2002a). A concordance-based study of metaphoric expressions used by general practitioners and patients in consultation. *British Journal of General Practice*, 52(475), 114-118.
- Skelton, J.R., Wearn, A.M. & Hobbs, F.D. (2002b). "I" and "we": a concordancing analysis of how doctors and patients use first person pronouns in primary care consultations. *Family Practice*, 19(5), 484-488.
- Stillman, P.L., Swanson, D.B., Stillman, A.E., Ebert, T.H. & Emmel, V.S. (1986). Assessing clinical skills of residents with standardized patients. *Annals of Internal Medicine*, 105, 762-771.
- Swales, J. (1990). *Genre analysis: English in academic and research settings*. Cambridge: Cambridge University Press.
- Swan, M. (1985). A critical look at the communicative approach. *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* (pp. 92-109), London: Longman.
- Varley, S. (2009). 'I'll just look that up in the concordancer': integrating corpus consultation into the language learning environment. *Computer Assisted Language Learning*, 22(2), 133-152.
- Vygotsky, L.S. (1978). *Mind in society: The development of higher psychological processes*. Cambridge, MA: Harvard University Press.
- Wallerstein, N. (1992). Powerlessness, empowerment, and health: Implications for health promotion programs. *American Journal of Health Promotion*, 6(3), 197-205.
- Wang, J., Liang, S. & Ge, G. (2008). Establishment of a Medical Academic Word List. *English for Specific Purposes*, 27, 442-458.
- Wass, V., Roberts, C., Hoogenboom, R., Jones, R., and Van der Vleuten, C (2003). Effect of ethnicity on performance in a final objective structured clinical examination: qualitative and quantitative study. *British Medical Journal*, 326, 800-803.
- Watermeyer, J. & Penn, C. (2009). The organization of pharmacist - patient interactions in an HIV/Aids clinic. *Journal of Pragmatics*, 41, 2053-2071.

- Widdowson, H.G. (1998). Context, community and authentic language. *TESOL Quarterly*, 32(4), 705-716.
- Widdowson, H.G. (2000). On the limitations of linguistics applied. *Applied Linguistics*, 21(1), 3-25.
- Wiggins, S. (2009). Managing Blame in NHS Weight Management Treatment: Psychologizing Weight and 'Obesity'. *Journal of Community & Applied Social Psychology*, 19, 374-387.
- Woodward, T. (1996). Paradigm shift and the language teaching profession. In Willis, J. & Willis, D. (Eds.) *Challenge and Change in Language Teaching* (pp. 4-9) Oxford: Heinemann.
- Wu, S., Franken, M. & Witten, I.H. (2010). Supporting collocation learning with a digital library. *Computer Assisted Language Learning*, 23(1), 87-110.
- Yakhlef, A. (2010). The three facets of knowledge: A critique of the practice-based learning theory. *Research Policy*, 39, 39-46.

\*Corresponding author. Address: Nottingham University, Division of Nursing, Derby Centre, London Road Community Hospital, London Road, Derby, DE12QY, UK.

Email addresses: paul.crawford@nottingham.ac.uk (Paul Crawford)