A corpus based, lexical analysis of patient information for radiography

Richards Golini, Catherine

A corpus based, lexical analysis of patient information for radiography

Catherine Richards Golini

Submitted to Swansea University in fulfilment of the requirements for the Degree of Doctor of Philosophy

Swansea University

2018

DECLARATION

This work has not previously been accepted in substance for any degree and is not being concurrently submitted in candidature for any degree.

Signed (candidate)

Date 27 November 2018

STATEMENT 1

This thesis is the result of my own investigations, except where otherwise stated. Other sources are acknowledged by footnotes giving explicit references. A bibliography is appended.

Signed (candidate)

Date 27 November 2018

STATEMENT 2

I hereby give consent for my thesis, if accepted, to be available for photocopying and for inter-library loan, and for the title and summary to be made available to outside organisations.

Signed (candidate)

Date 27 November 2018

Acknowledgements

I would like to express my gratitude and thanks to my supervisor Cornelia Tschichold, to other staff members in the Applied Linguistics department at Swansea, and all associates of the former Vocabulary Acquisition Research Group who have given me encouragement, support and friendship over the last 7 years.

Special thanks go to my parents for always showing interest, and to my husband, Stefano, for his immense support and super-human levels of patience.

Abstract

Despite the importance and the ubiquity of medical patient information in many healthcare systems in the world, we know very little about the lexical characteristics of the register. We do not know how patients perceive the information in the leaflets or whether the messages are transmitted effectively and fully understood. How a medical authority instructs and obliges patients in written information is also unclear.

While the number of radiographic examinations performed globally increases year on year, studies consistently show that patients lack basic knowledge regarding the commonly-performed exams and show very poor understanding of the concomitant risks associated with radiation. There is, then, a pressing need to investigate radiography patient information in order to better understand why, and where, it is less effective.

This thesis applies three approaches common to the field of corpus linguistics to uncover some of the lexical characteristics of patient information for radiography. The approaches used in this thesis are a keyword extraction, a lexical bundles analysis and an investigation of modal verbs used to express obligation.

The findings suggest that patient information for radiography possesses characteristics more common to academic prose than conversation, although the high informational content of the register goes some way to explaining this and suggests that the reliance on these structures may, to a certain extent, be unavoidable. Results also suggest that the reliance on *should* to oblige and instruct is problematic as it may cause interpretation problems for certain patients, including those for whom English is not a primary language.

Certain other characteristics of patient information revealed by the analyses may also cause comprehension, and while further research is needed, none of these characteristics would be evaluated as problematic by standard readability measures, furthering doubts about the suitability of such measures for the evaluation of medical information.

Table of Contents

List of Tables

List of Figures

1. The case for a corpus-based, lexical analysis of patient information for radiography

1.1 Background

Medical patient information is the cornerstone of the policy of shared decision making in medicine, a key aspect of patient centred healthcare in many countries, particularly the UK, in the 21st century. Written information plays an increasingly vital role in the relationship between practitioner and patient and this relationship has been of growing interest to researchers from a variety of disciplines, including applied linguistics, for the past three decades. And not only the practitioner-patient relationship: research into medical discourse more generally continues to increase, and as the number of studies has grown, so the scope of inquiry has widened. The term ‘practitioner’ now references a wider range of healthcare professional, including nurse, dentist, surgeon, nutritionist or radiographer, and not solely the much-studied, general practitioner (GP) (i.e. family doctor) who had dominated earlier studies. Areas of interest to language researchers have spread beyond the academic and the relational to include a great variety of topics such as public health campaigns (Buckton, Lean & Combet, 2015; Zarcadoolas, 2010), online healthcare information-seeking (Harvey et al., 2008; Mullany, Smith, Harvey & Adolphs, 2015), metaphor in medical talk (Nerlich & Halliday, 2007; Semino, Demjen, Hardie, Payne & Rayson, 2018) and patient narratives (Moore & Hallenbeck, 2010).

The availability of user-friendly corpus software, such as WordSmith Tools (Scott, 2017) and Sketch Engine (Kilgarriff, Rychly, Smrz & Tugwell, 2004) is also a factor in the growth of studies of healthcare discourse. The ease of use of software has made the building, analysing and comparison of corpora relatively straightforward with much of the work being automated. This is particularly the case for written corpora, of course, as collecting spoken data in a medical context is still a complex process with ethical issues and a raft of permissions that need to be granted before the data collection, and transcription, begin.

An additional factor behind the increase in healthcare discourse studies is the huge rise in the use of the internet to search for healthcare information, to talk about healthcare both as a patient and as a professional, and to practise medicine. The e - health or digital health market, which has a projected growth rate of more than 15% in the US, is driven by the prevalence of chronic diseases and government initiatives to deploy e-health (Grand View Research, 2018). In the UK, more than 80% of adults use the internet regularly with just over half reporting that they use the internet to search for health-related information, a figure that has grown by more than 30% over the preceding decade. (Office of National Statistics, 2018). The use of the internet for health reasons is of particular interest to language researchers who have studied patient-professional exchanges, including medical advice sites (e.g. Harvey, Locher & Mullany, 2013) as well as patient-patient communication via forums (e.g. Angouri & Sanderson, 2016; Seale, Ziebland & Charteris-Black, 2006).

In spite of the growing interest in healthcare discourse studies, however, and despite its ubiquity and importance, written patient information has received scant attention from applied linguists.

The lack of attention paid to medical patient information may, in part, be explained by the fact that spoken medical discourse receives more attention in the literature, which is understandable, says Clerehan (2014, p 212) as spoken discourse is 'arguably the most salient, significant and principal mode of healthcare communication'. The lack of attention may also be explained by the fact that the use and visibility of patient information in many countries is a relatively recent phenomenon. In Switzerland, where I live, there is currently very little printed information available of the kind that is ubiquitous in the UK. Switzerland is a country where medical care is still paternalistic in nature, however, (Lucassen, 2015) along with many countries in eastern Europe (Simek, Krizova & Zamykalova, 2012) and Asia (Claramita, Nugraheni, van Dalen & van der Vleuten, 2013). The value of an informed patient (and thus by extension, the availability of patient information) is likely to be perceived less favourably in such countries.

In the UK and the US, however, patient information has been visible, considered important and discussed for many decades, making the lack of research from language researchers puzzling. The first official advice in the UK for writers of patient information first appeared in 1962. This booklet, the result of research by the King Edward's Hospital Trust, and presented as a style guide, contained advice on the writing of information for patients who were being admitted into hospital. (King Edward's Hospital Trust, 1962). Interestingly, the booklet referenced earlier reports from the early 1950s on the importance of giving patients information and what patients needed to know. The reasons behind informing patients and the advice regarding the tone in which the information should be presented looks very similar to contemporary guidelines issued by the NHS and other authorities on patient information such as the Patient Information Forum:

> For many people, admission to hospital is counted amongst the major events of their lives. Though most patients later recall their stay in hospital with gratitude and relief, beforehand they all too often view the prospect of admission with uncertainty and apprehension. It is largely to help dispel these fears, and to prepare patients for the unfamiliar hospital world, that information booklets are issued. It is clear too that some authorities consider the booklet is not just a means of giving information but also a way of helping to establish rapport with a patient in a manner that is warm and understanding, rather than patronizing or pompous. (King Edward's Trust, 1962, p 3)

The booklet does not refer to language directly, though does suggest that pictures and cartoons are included, along with a map, contact information and a friendly foreword, saying that 'this feature (or its absence) usually sets the tone of the booklet as a whole.' (p 6). The emphasis is very much on making the text attractive and readable for the patient. It is also interesting to note that the practice of asking patients and staff (not only department heads) to comment on the patient information before it was printed was being practised by a number of hospitals in the UK in the early 1960s. In 2018, however, patient involvement in the production of patient information is still not universal.

Patient information is not, then, a recent phenomenon in the UK, or the US. Given the importance of information for health outcomes, and the on-going concerns relating to its effectiveness and comprehensibility, the relative lack of interest from applied linguists is surprising.

1.2 Patient Information for Radiography

This doctoral thesis concerns itself with a specific kind of patient information from a specific medical field: procedural patient information for radiography. This type of patient information is specific to a particular radiographic exam, such as computed tomography (CT) or magnetic resonance imaging (MRI), or it may relate to a medical procedure that involves the use of a radiographic technique, such as angiography. There is a pressing need to explore the language used in radiography patient information, with the objective of improving the patient's understanding of the topic, as studies show that, in spite of the increasing number of examinations being performed, particularly CT scans (e.g. Thurley, Crookdake, Norwood, Sturrock & Fogarty, 2017), patients do not understand the differences between common radiographic examinations, are unable to say which exams use radiation and, almost certainly as a result of this lack of knowledge, they show an alarming lack of concern for radiation risk. (e.g. Singh, Mohacsy, Connell & Schneider, 2017). The lack of understanding of the technology and associated risk may be compounded by the complexity or even non-availability of printed information materials.

I have taught English to radiographers in Switzerland for a number of years, and in this time I have often used procedural patient information from the UK and the US as teaching material. Patient information lends itself well to English as a Second Language instruction as the text is generally written as if it were a conversation, with questions and responses relating to a patient's experience of a range of radiographic examinations. CT, MRI, ultrasound and plain X-ray are the most common. This written conversation is assumed to resemble what occurs in the radiography suite in a hospital, and thus is useful for presenting vocabulary and themes student radiographers will need in the workplace. At the beginning of this doctoral process, then, I was very familiar with patient information for radiography. I was also aware that writers were expected to follow certain guidelines relating to vocabulary and structure, although my reading of research had also shown me that there was, in fact, great variation in healthcare materials relating to the content and accuracy of the information, and also to the perceived complexity of the information and thus its comprehensibility.

1.3 Medical vocabulary vs terminology

In this thesis I use the terms medical vocabulary and (medical) terminology synonymously. I believe that referring to the kinds of words used in medical interactions as either general or medical is an unhelpful and overly simplistic dichotomy. It is more helpful to view this vocabulary as being situated on a continuum, with medical nomenclature (i.e. technical terms characterised by Latin and Greek terms and medical abbreviations) at one end, and very general vocabulary at the other. In between we find other categories that we may refer to as semi-technical vocabulary and lay technical, and general items at the other extreme. These are very useful terms to categorise vocabulary, particularly if our attention is on the vocabulary itself, rather than the use or understanding of it in medical interaction. How a word is used and understood in medical interaction varies from person to person however, suggesting that a word can belong to a number of categories - fully technical, semi-technical and even lay technical - depending on how the word is understood by the person using it.

Patients often use medical vocabulary when they are talking about their condition, though they sometimes use it in different ways to the professional. *Chronic* is a good example of this, meaning ‘long-lasting’ for a medical professional and ‘bad’ for some speakers of English. Lay people may also use terms differently because they emphasise the experience of the condition or symptom. Additionally, it may also be the case that patients, with their individual subjective experience of a condition always use medical terms differently to the professional, and that just as personal experience of a condition varies from one person to another, so do the meanings attached to words.

Professionals, too, may use the same word but their understanding of the word can differ, depending on their specialism and understanding of the field: a consultant psychiatrist who researches the condition will understand *schizophrenia* in a different way to the GP, while the lay person with schizophrenia has an understanding of the issue of a very important, but different, kind. (J Skelton, personal communication, 23 October 2018).

Given the complexity of medical vocabulary, and the importance of improving the comprehensibility of published healthcare materials, it is no surprise that the use and comprehension of medical vocabulary has, to date, been the principal area of interest for language researchers.

1.4 Language studies in healthcare

Studies from applied linguistics of healthcare communication are very valuable because they ´offer a means of making sense of some of the complexities of healthcare: exposing beliefs and practices that might be taken for granted or overlooked altogether´ (Harvey and Koteyko, 2013, p 2). Most language studies (from applied linguistics and other fields) have focused on the use of specialised, medical vocabulary (e.g. Bourhis, Roth & Macqueen, 1989; Skelton & Hobbs, 1999a; 1999b), while the differences in lexical style of doctors and nurses (Bourhis, Roth & Macqueen, 1989; Collins, 2005) and the lexical challenges facing medical professionals trained abroad (Bosher & Smalkoski, 2002; Cameron, 1998; Dahm, 2011) have also been of interest.

Earlier studies of spoken medical discourse, many using the approach of conversational analysis, tend to focus more on the practitioner, often a doctor rather than any other health professional. Initial findings that doctors use a lot of complex medical vocabulary with patients, in spite of their perception otherwise (e.g. Bourhis, Roth & MacQueen), contrast with other, later studies, e.g. Skelton & Hobbs, 1999a, who find that doctors are, in fact, generally aware of their use of medical vocabulary and are able to explain relevant terms for their patients (though it is notable that in Skelton & Hobbs, 1999a, it is they, the doctors, who decide what needs explaining). The narrow focus on the practitioner in healthcare communication studies began to widen to include the patient, and many studies since are interested in the comprehension of, and the impact of, medical vocabulary on the patient. Most research comes to the conclusion that patients understand far less than doctors think they do (e.g. Chapman, Abraham, Jenkins & Fallowfield, 2003; O’Connell, Hartridge-Lambert, Din, St John, Hitchins & Johnson, 2014). Even terms as significant in meaning as *benign* and *malignant* fail to be understood by as many as a third of cancer patients (Chapman et al., 2014). It is possible, conclude Chapman et al. (2014), that a substantial proportion of the public do not understand the language used in consultations, nor can their knowledge of even basic anatomy be assumed, with serious implications for the success of many medical interactions. An early study by Boyle (1970) found similar results to Chapman et al. (2014) regarding patients' ignorance of anatomy, which suggests that internet use has not resulted in an increase in this particular type of knowledge. Ignorance of medical vocabulary in general has been reported in many other studies and for other languages (e.g. Blackman & Sahebjalal, 2014; Hayes, Dua, Yeung & Fan, 2017; Pieterse, Jager, Smets & Henselmans, 2013).

While studies in disciplines other than applied linguistics and health communication continue to focus on what a patient does not understand, (e.g. Cherla, Sanghvi, Choudhry, Liu & Eloy, 2012; Hansberry, John A, John, E, Agarwal, Gonzales & Baker, 2014), a handful of more recent studies by linguists and healthcare communication researchers have focused on what patients do understand, and how they use the terms that they know (Fage-Butler & Jensen, 2016; Koch-Weser, de Jong & Rudd, 2009; 2010). These studies reveal a considerably more complex picture than the standard understand/do not understand dichotomy.

There is growing evidence that many patients, particularly those living with a chronic condition, use medical terms, and as they become better informed about their condition, these terms are used with greater frequency (e.g. Fage-Butler & Jensen, 2016). The identity of the 'expert' patient, a phenomenon that reflects a growing confidence among health consumers, a technological society and the ready availability of health information, has been the focus of a number of studies (Fox & Ward, 2006; Fox, Ward & O'Rourke, 2005).

Koch-Weser, de Jong & Rudd (2009) focused on the words used by both patient and professional, finding that while some patients use medical terms as often as their doctor during a consultation they rarely use the same words. The study found that patient medical vocabulary tends to cluster in the history-taking section of the consultation, which is not surprising as this also tends to be the moment when the patient has an opportunity to speak. Professionals, on the other hand, use medical vocabulary throughout the consultation. Further research may reveal whether patients do not use medical vocabulary elsewhere in the consultation simply because they are not speaking, or whether they are only able to use it when they are explaining to the doctor the reason for the appointment. If this is the case, doctors should not assume that patients understand everything that is being said during the consultation, irrespective of their accurate use of medical words in the history-taking part. Additionally, patients who are not able to use terminology in the history-taking part of the consultation may well have a lower health literacy level, which the doctor will need to be aware of and adapt to. As Koch-Weser, Rudd and DeJong (2010) say:

> For their part, patients must express themselves clearly to participate actively in decision-making. Patients' success in describing their symptoms accurately depends in part on the sophistication of the vocabulary they can call on. Thus, measures of word use can offer insight into their "expressed literacy level." By extension, such measures may also indicate the vocabulary that patients are likely to comprehend. (p 591)

These findings make it all the more important to train doctors to allow the patient to speak during the history-taking phase without interrupting as such an approach can yield very useful information.

As we have seen, while some patients do use medical vocabulary, studies also show that patients do not always use words with the same meanings as doctors. There are different explanations for this, though one reason is the different understanding practitioners and patients have of a word, as discussed above. (Dahm, 2018; 2011; Hadlow & Pitts, 1991; Ong, de Haes, Hoos & Lammes, 1995). Professionals and patients can respond differently to words, which is likely a response to the connotations of the words. This seems particularly the case with words that reference mental illness such as depression, along with words for body size, weight and the concept of obesity (Mullany, Smith, Harvey and Adolphs, 2015; Ogden et al., 1999; Tailor & Ogden, 2009). The emotional response to a word's connotations are not straightforward, however. Tailor and Ogden (2009) found that while doctors prefer to use euphemisms rather than the term *obese* with their patients, patients who were truly obese found the euphemism more upsetting to hear. On the other hand, patients who were not obese (when their BMI was measured) felt more anxious and upset on hearing the term *obese* than they did when the euphemism was used. The authors conclude, perhaps a little glibly, that a 'GPs choice of term therefore needs to reflect whether they want the patients to be upset or whether they want them to accept the seriousness of their problem.' (2009, p260).

The fact that different meanings and connotations can be attached to words is further evidence that the traditional dichotomy of technical and semi-technical - and the simplified view of medical jargon versus general language - fails to capture the very varied characteristics of medical vocabulary. In a recent study, Fage-Butler and Jensen (2016) developed five categories of medical terms, using them as a framework to evaluate online forum interactions by patients with chronic conditions. Patients in the study were found to use a great variety of terms from all of the categories, leading the authors to cautiously recommend that doctors use an adaptive approach to medical terminology during consultations, adaptive to the knowledge of their patients.

The idea of adapting to the knowledge and health literacy of the patient - which may be higher or lower than the professional initially believes - relates to the idea of tailoring written information to the patient, an approach that studies show is generally more effective than a one-size-fits-all approach to information (Jensen, King, Carcioppolo, Davis, 2012; Lustria et al., 2013.) The idea of tailoring information will be discussed in greater detail in Chapter 4. As it currently stands, however, much of the written procedural information produced by and for hospital departments in the UK is not tailored. General patient information guidelines are applied to all texts and the aim seems to be the simplification of information, for all patients, irrespective of their status.

1.5 Health literacy vs literacy

The significant implications of not understanding, and the negative impact of poor understanding on patient outcomes has motivated the very many studies of written health materials (all from fields other than applied linguistics) that focus on patient comprehension and the readability of materials. While different readability tests use different measures, they generally count word length in syllables and/or sentence length to assess the complexity of a text. However, most readability tests cannot differentiate between a monosyllabic or two syllable medical word and a general term: *sacral*, *benign* and *ructus* could be assessed as more readable than *operation* or *corridor*. Another problem with the concept of readability is that it relates to literacy. Literacy, however, is not the same as health literacy. As Zarcadoolas says (2011), health literacy is complex and multifaceted, requiring much more than the ability to read simplified text. Just how complex and multifaceted is made clear in the following paragraph from the US Department of Health and Human Services, who define health literacy as:

> The degree to which individuals have the capacity to obtain, process, and understand basic health information and services needed to make appropriate health decisions. Health literacy is dependent on individual and systemic factors: Communication skills of lay persons and professionals; Lay and professional knowledge of health topics; Culture; Demands of the healthcare and public health systems and the Demands of the situation/context
>
> Health literacy affects people's ability to: Navigate the healthcare system, including filling out complex forms and locating providers and services; Share personal information, such as health history, with providers; Engage in self-care and chronic-disease management [and] understand mathematical concepts such as probability and risk. (Department of Health and Human Services. n.d)

To close the comprehension gap between health messages and the public, Zarcadoolas (2011) calls for `a richer, more theory-based understanding of text structures and functions, along with other powerful constructs, including cultural appropriateness, relevancy and context'. (p 338). This thesis shares the view, expressed by Zarcadoolas (2011) and a handful of other researchers (e.g. Clerehan, Buchbinder & Moody, 2005; Rubin, 2014), that standard readability measures are inappropriate tools for the development and evaluation of patient information materials. The ‘theory-based understanding of the structure and function of text’ that Zarcadoolas (2011, p 338) refers to is, I believe, an area of knowledge that applied linguistics research, including the research reported in this thesis, can certainly contribute to.

1.6 The importance of information in healthcare

Turning now to the final motivating factor for this doctoral research: the importance of information in healthcare. The ultimate objective of studies in healthcare communication must be to improve patient satisfaction and health outcomes, and there is plenty of evidence of the role that being or feeling informed has in improving clinical and non-clinical outcomes (Sheard & Garrud, 2005; Coulter & Ellins, 2007;

on knowledge (Maggs, Jubb & Kemm, 1996) and of the positive relationship information has with compliance and adherence to treatment programmes (Boundouki, Humphris & Field, 2004;). I believe that studies of the language used in healthcare communication, such as those reported in this chapter, can help make information effective, whether it be written information or spoken.

Information must be comprehensible in order to be effective. And while a better understanding of the complex nature of medical vocabulary is fundamental to healthcare communication studies, we should not forget about general language. General language makes up a sizeable amount of the language of patient information - after all, most complexity and terminology should have been removed if guidelines are followed - and general language is usually perceived to be transparent in meaning and to cause few comprehension problems. Is this really the case? Until we investigate the characteristics of general language in patient information how can we be sure that it is a valid assumption?

It is general vocabulary, then, as used in medical patient information that is the primary focus of this doctoral thesis. My aim is to uncover the lexical characteristics of patient information for radiography through the application of corpus linguistics approaches. The lexical analyses I conduct may well reveal the kinds of hidden beliefs and practices that Harvey and Koteyko (2013) refer to, they may reveal the underlying discourses and, additionally, they may also uncover linguistic aspects that may play a significant role in the comprehension of the text.

A variety of applied linguistic approaches have been used in the investigation of healthcare communication, and two of the more common for studying spoken medical interaction have been critical discourse analysis (e.g. Mishler, 1984), a mode of discourse analysis concerned with power relations and ideologies in language, and conversation analysis (CA) (e.g. Drew, Chatwin & Collins, 2008; Heritage & Stivers, 1999; Jones, 2003) which is an approach that seems particularly suitable for the analysis of patient-provider interaction as it has been described as `a direct research embodiment of patient-centredness` (Maynard & Heritage, 2005, pp. 433-434). In CA, the focus is particularly on the structure of the utterance and turn-taking in the development of the interaction. Written medical discourse, most especially research papers and their abstracts, has often been investigated from the approach of genre analysis (GA), which, broadly speaking, is an approach to text analysis that defines the lexical parameters inherent in a particular genre, while some studies have combined genre analysis with the use of corpus analysis techniques (Hill Davies, 2015). An overview of the approaches used in the investigation of written medical discourse will be presented in Chapter 2.

For the analyses described in this doctoral work, I draw on the field of corpus linguistics, and use the following corpus techniques: a keyword analysis and a lexical bundle analysis. I chose to use the keyword method as it has already been well-used in healthcare communication research and has established itself as an effective tool for revealing the underlying themes in healthcare communication. Some key papers from healthcare communication that utilise this technique, e.g. Adolphs, Brown, Carter, Crawford, & Sahota (2004), Harvey et al. (2008) and Harvey & Atkins (2010) are discussed in Chapter 2. As we saw earlier in this chapter, the first style guides produced by the health service in the UK referred to the need to reassure the patient and to 'establish rapport with a patient in a manner that is warm and understanding, rather than patronizing or pompous' (King's Trust, 1962). A keyword analysis can provide a way into the discourse and can show us whether these aspects are still present in the text, or whether there are other issues that have replaced them in importance.

The second technique I use is a lexical bundles analysis. Lexical bundles are multiword strings which are generally non-idiomatic, e.g. *let's turn to* and *at the end of*. They have been referred to as `characteristic features of language use in particular settings` (Hyland, 2008, p8) and are very common in both spoken and written discourse. Lexical bundle analysis has already shown itself to be very useful in uncovering the lexical characteristics of a register (e.g. Conrad & Biber, 2004; Biber & Barbieri, 2007). A better understanding of the functional characteristics of patient information can be arrived with an analysis of its lexical bundles. We can see, for example, whether the information is written in the conversational style that guidelines encourage writers to adopt, or whether the style of the text might reasonably be expected to develop a rapport with the reader.

One of the primary reasons for choosing corpus approaches to the analysis of healthcare discourse is that they allow the investigation of very large datasets. Irrespective of the size of the corpus, the use of corpus techniques allows reliable generalisations about language use to be made based on the statistical analyses of a large dataset; comparisons of the frequency of items in one corpus relative to another can be carried out and key words generated, which can give a sense of the nature of specialised discourse. Word lists can be produced for English for Specific Purpose teaching purposes and lexical relationships, such as collocations, investigated. All of these tasks would be difficult, time-consuming or even impossible to carry out without the aid of software.

Frequency information alone, however, is highly unlikely to tell us much about medical communication (Skelton & Hobbs, 1999a; 1999b). Lists of keywords, for example, can only suggest areas of communicative interest that may be interesting: they are `pointers which suggest to the prospector areas which are worth mining, but they are not themselves nuggets of gold`. (Scott, 2010, p 51). The studies reported in this thesis, then, employ first a quantitative analysis using corpus software followed by a qualitative analysis of the item of interest in the context of the corpus.

In addition to the two approaches described above, a third corpus analysis is described, that of modal verbs in patient information. The modal verbs I use in my analysis are those that are concerned with obligations and instructions and are often referred to as deontic modal verbs. This is a different kind of corpus analysis to those described in the previous paragraph. This analysis involves the prior selection of lexical or grammatical items of interest which are then searched for within the corpus. I was motivated to investigate deontic modal verbs as procedural patient information does not just inform but it also instructs, though how written information uses language to instruct, we know little about. This analysis, and its results, are described in Chapter 6.

1.8 Outline of thesis

The research in this thesis has been informed by different fields of study: studies from applied linguistics, particularly those that use corpus techniques to describe a language variety or register, and corpus studies of healthcare communication. My research has also been informed by studies from fields outside of applied linguistics, such as health literacy, patient education, and social psychology. Health communication is a multidisciplinary field, and applied linguistics interdisciplinary, but it has not always been straightforward to pull together all of the strands that have informed this study.

The literature that has informed my research appears in Chapter 2. This chapter contains key studies of written medical discourse, and studies of healthcare communication, both spoken and written, that have utilised corpus approaches healthcare communication. As we shall see, there have been far more studies of spoken medical interaction utilising corpus techniques than there have been of written registers. The chapter also includes studies that call for a new approach to healthcare communication research and that refer to the methodological difficulties of such cross-disciplinary research. Finally, I present studies that call specifically for a new, linguistic approach to the production and evaluation of healthcare information materials.

In Chapter 3 I describe the software that I used to build and investigate the specific corpora used in my analyses and I present the corpora and methodology used for each analysis. Chapters 4-6 are dedicated to the lexical analyses carried out of key words, lexical bundles and modal verbs for instructions in patient information. Each chapter contains the procedure, results, discussion and conclusions.

In Chapter 7, the thesis concludes with a discussion of the more significant findings from each of the analyses. The findings are first discussed with reference to the questions that motivated the research, namely what they reveal about the lexico-grammatical characteristics of patient information for radiography. I present my conclusions regarding the impact of these characteristics on the comprehension of the patient information, and present my conclusions regarding the appropriacy of standard readability measures to evaluate materials. I also present a summary of the findings that relate to the underlying beliefs and discourses of patient information for radiography that were also revealed by my analyses. The chapter ends with a discussion of the limitations of the research and with some suggestions for further research. Some of these suggestions for future research relate directly to the questions that motivated my study, while some relate to the themes that emerged during the research process.

It is inevitable, perhaps, that an exploratory register analysis such as the one described in these pages, will uncover unexpected themes of interest. These themes are not directly related to the research questions but nonetheless are relevant to the broader topic of healthcare materials and patient-provider communication. A number of such themes emerged during my research. They are presented and discussed generally in chapters 4-6. In chapter 7 they are signalled as areas worthy of further exploration. Any in-depth analysis is beyond the scope of my data, however.

2. Literature Review

There is a vast amount of research on medical language which, if we were to consider it all relevant for this literature review, would render impossible the task of selecting. This thesis, however, takes the view that medical language is not a homogenous genre, but a genre made up of a large variety of registers, each with its particular lexical and grammatical characteristics. It is the register of written patient information for radiography that is under investigation in this doctoral thesis, while corpus analysis is the methodological approach taken. Register in this thesis is used to mean a `specific language variety associated with a particular configuration of situational characteristics and purposes`. (Staples, Egbert, Biber & Conrad, 2015, p. 505).

This chapter begins with an overview of the literature on medical registers other than procedural patient information, beginning with academic medical English. The research focus here has mostly been on the characteristics of the medical research papers or its sections, on case notes (an academic-professional register), and on producing word lists of academic medical vocabulary.

The overview is followed by a summary of the few studies that have investigated the professional register of consent forms. The concerns central to these studies - readability and comprehension - are central to my investigations of procedural patient information. The few studies that look at patient information for sufferers of chronic conditions, such as rheumatoid arthritis and depression, is then summarised. These studies generally aim at understanding whether these healthcare materials are appropriate for the patients and whether their needs are met by information (e.g. Grimes and Ong, 2007) The focus on these studies is particularly on the content, the approach taken is often discourse analysis. The conclusion is that patient information needs to be far more accurate with the information it presents, and to involve the patient, their concerns and their experience far more than it currently does.

My focus then turns to a summary of the literature relating to pharmaceutical patient information leaflets (PILs). There has been a fair amount of interest in PILs from researchers (including applied linguists), possibly because these packaging inserts are obligatory in Europe, and their content regulated. I believe that many of the themes discussed in these studies are relevant for studies of procedural patient information, such as the one described in this thesis. These themes include readability, complexity and how risk is expressed and perceived.

In the second part of this chapter, the broad literature overviews described above give way to more detailed discussions of papers that have particularly informed my research. This section begins with two studies that focus on the importance of cross-disciplinary research. Roberts and Sarangi (2003) present some of the challenges that cross-disciplinary work can bring, while Candlin and Candlin (2003) highlight the importance of real-world outcomes of language research in healthcare. This is a theme that is also present in many of the corpus studies of medical discourse which I present in 2.4. Many of the papers in 2.4 utilise a keyword analysis and I include here studies that have focussed on both written and spoken discourse. Some of the earliest and most important corpus studies of medical language were of spoken communication, e.g Adolphs et al. (2004). Table 1 shows the areas of interest of the corpus studiesI discuss in 2.4.

Table 1 Corpus studies of medical discourse presented in section 2.4

Section	Authors & date	Register and focus	Mode	Corpus technique?
2.4.1	Skelton and Hobbs 1999(a)	Patient-provider communication - consultations	Spoken	Concordancing*
2.4.2	Adolphs, Brown, Carter, Crawford and Sahota (2004)	Patient-provider communication - advice phone lines	Spoken	Keyword analysis
2.4.3	Seale, Ziebland and Charteris-Black (2006)	Forum postings (gender and cancer) *plus* traditional interviews	Online written *plus* spoken	Keyword analysis
2.4.4	Seale and Charteris-Black (2008)	Illness narratives (age and cancer)	Spoken	Keyword analysis

2.4.5	Harvey et al. (2008)	Adolescent health communication	Online written	Keyword analysis
2.4.6	Grabowski (2015)	Pharmaceutical registers (x4)	Written	Keyword analysis and lexical bundle analysis
2.4.7	Vihla (1999)	Modal verbs in medical registers	Written	Corpus-based analysis
*refers to the analysis of lines of text from a corpus. Areas of interest choosing by the researchers and subsequently examined in the corpus				

The literature review concludes with two studies that focus specifically on healthcare materials and health literacy: Clerehan, Buchbinder and Moody (2005) and Zarcadoolas (2011). Both studies call for a different approach to the development and evaluation of health materials, for a better understanding of what we mean by health literacy and for an approach to the development of healthcare materials that is informed by both the expressed needs of patients and by research from the field of applied linguistics.

2.1 Academic medical registers in the literature

In medicine, the variety of written registers includes academic research papers, case reports, consent forms, care plans, discharge notes, dietary and lifestyle advice, medical device instructions, handovers, progress notes, imaging reports, injury prevention information, medication information and procedural information. Very few of these have been the subject of any linguistic interest, however. Academic medical writing has been the focus of most interest, with a handful of studies on the language of case reports and consent forms. This are summarised in the following sections.

2.1.1 Research papers and case notes: genre analysis

In the literature, studies of academic registers predominate, particularly research papers (and their abstracts) (Atkinson, 1995; Nwogu, 1997; Salager-Meyer, 1990;

1992; 1994) and case reports (Nissen & Wynn, 2014; Taavitsainen & Pahta, 2000). Many such studies have been carried out within the field of genre studies and its three research branches: English for Specific Purposes (Swales, 1990; Bhatia, 1993), New Rhetoric (Bawarshi and Reiff, 2010) and Systemic Functional Linguistics (Halliday, 1978; 1985; Halliday and Hasan, 1976).

Broadly speaking, these three research traditions all relate to the investigation of the language of discourse communities, though New Rhetoric (NR) is not concerned with L2, unlike English for Specific Purposes (ESP) which is generally concerned with L2 learning and teaching, and Systemic Functional Linguistics (SFL) which is concerned with both L1 and L2 contexts. While an exploration of genre studies is outside of the scope of this thesis, I will present a brief summary of the characteristics of each branch in the section that follows.

New Rhetoric, which developed in North America, is particularly focussed on the institutional contexts that surround genres, and the functions of genres within those contexts. SFL, on the other hand, emphasises the function of language in constructing meaning in society. The emphasis in SFL is very much on communication. While SFL has been used less often in Europe as an approach to genre analysis, it is better known in other parts of the world, particularly Australia (e.g. Clerehan, Buchbinder & Moodie, 2005; Clerehan and Buchbinder, 2006). Clerehan and colleagues take an SFL approach in their various studies of pharmaceutical patient information and in their development of a new framework for the evaluation of patient information. Clerehan et al. (2005) is an important study that will be presented and discussed in more detail later in this chapter.

ESP research is particularly concerned with the linguistic features of a text, its organisational, grammatical and stylistic features. Many of the studies overviewed in the section below fall into this research tradition (e.g. Li & Ge, 2009; Nwogu, 1997).

Genre studies within the ESP tradition has been a common approach taken to the analysis of medical research papers and of case notes, the two most academic of registers. Several studies have looked at the structure of the text, with reference to the sections or ‘moves’ (Swales, 1990) within it (Davies, 2015; Nwogu, 1997; Williams, 1999), while other have looked at the historical evolution of the research paper or case notes (Li & Ge, 2009; Salager-Meyer, Alcaraz & Luzardo, 2012). Several studies have focussed on specific linguistic aspects of the text, such as hedging (Salager-Meyer, 1994) or modality (Salager-Meyer, 1992; Vihla,1999). How medical research writing and abstracts differ from other disciplines has also been the focus of several studies (Fløttum, 2006; Giannoni, 2010), as have cross-linguistic comparisons (Maci, 2012).

Genre analysis within the ESP tradition is concerned with the relationship between the generic characteristics of a text, characteristics which are not fixed but 'highly dynamic and closely related to their socio- professional contexts' (Bhatia & Gotti, 2006). This dynamic variation of the characteristics of a text relates to the purpose of the text, the relationship of writer and recipient, the setting and the communicative events being enacted by the text. The communicative purpose of the text, therefore, shapes the text. The particular interest in academic medical English, indeed, any academic professional subject, relates to the need to train student doctors and scientists to be competent members of what Swales refers a 'discourse community' (Swales, 1990). Learning how to write academic research papers that will be accepted by that discourse community therefore has considerable importance.

Genre studies, as we have seen, have investigated the medical research article as a whole, along with its individual sections. Budgell (2011) narrowed the focus even more to investigate the titles of research papers in biomedical research, with a particular focus on randomised control trials (RCTs). A corpus of titles from RCTs was analysed using a comparative corpus made up of titles from 1,000 papers from four generalised medical research journals. Budgell (2011) found that the titles of RCTs are characterised by distinct conventions that include word choice, length and the use of recurrent phrases. Given the role of RCTs in the world of medicine, learning to craft a title according to the convention revealed by Budgell (2011) may well contribute to a paper being accepted for publication.

Corpus techniques have predominantly been used for the production of academic wordlists for medical study. These will now be summarised in the following section.

2.1.2 Medical academic wordlists

Still within the field of ESP, we move away from small-scale data analysis to corpus linguistic approaches, where large amounts of data are analysed by computer rather than by hand. Here we find a number of studies that focus on the identification of specific lexis for academic medical study. The increasing availability of corpus software over the last 20 years has made the task of producing word lists much easier and over time these word lists have become ever more specific. Coxhead's general Academic Word List (AWL) (2000) was the first such list. Questions relating to the AWL's suitability for specific disciplines, such as medicine, gave rise to an examination of the AWL's coverage of medical research papers (Chen & Ge, 2007). This in turn led to the creation of the Medical Academic Word List (MAWL) (Wang, Liang & Ge, 2008) which has been joined more recently by the New Medical Academic Vocabulary List (MAVL) (Lei & Liu, 2016). The MAVL is around half the length of its predecessor, but with greater coverage. Earlier studies in academic medical discourse (and it must be said, non-academic medical discourse in general) have favoured the physician/doctor over other health professionals and this is also seen in the kinds of word lists that have been produced, where medical students, i.e. doctors in training, are the focus. Nonetheless, academic word lists have also been produced for pharmacology students (Fraser, 2007) and for nurses (Yang, 2015).

And what of other registers in written medical English? It transpires that there have been far fewer language studies of written medical registers, other than the academic and pedagogic. Consent forms are one exception. A summary of the literature relating to these important and wordy medical documents now follows

2.1.3 Consent forms

Consent forms in medicine are extremely important for they are declarations by the patient that they understand the procedure they will be having and, more pertinently, they understand the concomitant risks. Consent forms have a legal status and their primary purpose is to prove that the patient did understand the risks associated with the procedure and thus the hospital or medical provider are protected from legal recourse. These important documents have been the focus of a handful of language studies (e.g. Ilić, Auchlin, Hedengue, Wenger & Hurst, 2013; Mayberry & Mayberry, 2002; Sterling, 2015) with their comprehensibility a primary focus: consent forms, after all, need to be understood before they are signed.

It is the presumption of understanding that is the focus of Mayberry and Mayberry's study (2002). The authors, both medical professionals, suggest new approaches (new to the world of medicine, that is) to the testing of comprehension of consent forms, such as a cloze test, and raise as an issue of primary importance, that of the comprehension difficulties faced by patients with functional illiteracy and by those with a language other than English as an L1. The difficulties of comprehending consent forms by L2 speakers of English is also the focus of Sterling (2015).

Many studies highlight the length and complexity of the consent form (e.g. Ilić, Auchlin, Hedengue, Wenger & Hurst, 2013; Pandiya, 2010) and the issue of complexity is central to studies of other medical registers, including patient information leaflets. Readability measures have also been used to assess the comprehensibility of the consent form, though as we will see, literacy measures are problematic when they are applied to medical discourse.

2.1.3.1 Readability measures of consent forms

The limitations to standard readability measures include the fact that these measures do not assess health literacy but literacy. Health literacy and literacy are not the same thing, with the latter relating solely to the ability to read and write. Health literacy, on the other hand, relates to the capacity an individual has to obtain, process and act on health information. Being health literate involves a number of different skills, including social and cognitive skills. An individual can be literate but not health literate. Being health literate is vital for the successful transmission of the message: evidence exists that health literacy and numeracy are far stronger predictors of comprehension of medical information (in the form of a leaflet, for example) than literacy (Hibbard, Peters, Dixon & Tusler, 2007).

Another issue with readability evaluations is the means by which word complexity is measured. Most readability measures cannot distinguish likely-to-be-known long words form unlikely-to-be-known words in the text, distinguishing between them solely by their length. However, some long medical words will be recognised by many patients as they are part of most people’s lexicon, e.g. *antibiotic* or *diarrhoea*. These terms are lay technical. It should be remembered, however, that a patient may not use or understand the terms in exactly the same way as a medical professional.

Conversely, plenty of rare, technical terms exist that are monosyllabic and thus judged very readable by a standard readability test. Acronyms, e.g. DEXA (dual energy x ray absorptiometry) are a good example of this: used and understood by medical professionals (and often only from the field or speciality) or by the expert patient who is considerably more familiar with them, they are impenetrable to everybody else.

Pandiya (2010) underlines the fact that, while readability contributes to comprehension and promotes a willingness to read the text, readability is not the same thing as comprehension. Pandiya (2010) makes a number of proposals to improve the comprehension of consent forms -which can run up to 20 pages in her experience. Many of the approaches she proposes are those suggested for the development of patient information leaflets, such as those analysed for this doctoral study, and for pharmaceutical information leaflets. They include the use of bullet points and diagrams; the use of general vocabulary in place of medical terms; shorter sentences and frequent paragraphs and the use of active verbs rather than passive. An interesting observation from the study is that in India, where Pandiya works, a standard consent form is translated from English into the very many languages spoken in India. This has resulted in very long documents with literal translations (which cause comprehension problems) and with language that is considerably more sophisticated than the level of many of the readers. While Pandiya (2010) does not refer to the concept of consent being problematic, this may also be a factor: medical consent giving is not common practice all communities and cross-cultural research shows that translations into languages that do not share the concept can result in confusion, embarrassment and even mistrust. (McCabe, Morgan, Curley, Begay & Ghodes, 2005).

While the problem of the (in)appropriacy of the level of the language in consent forms remains a concern, there does seem to be some sign of a move away from sole reliance on the standard readability measure (e.g. Pandiya, 2010; Villafranca, Kereliuk, Hamlin, Johnson and Jacobsohn, 2017). Villafranca et al. (2017) is a particularly good example of this. This study assesses the linguistic complexity of a consent form using a traditional readability measure, in this case the Flesch-Kincaid tool, and also with the Coh-Metrix v3.0 (McNamara, Graesser, McCarthy, Cai, 2014). The Coh-Metric v3.0 software uses the measures of word familiarity and imageability to evaluate the coherence and cohesion of research consent forms. Word familiarity pertains to how familiar the involved words would be to a lay population while average word imageability to how easy it would be for a lay population to visualise the involved words (McNamara et al., 2014). The tool also provides linguistic norms for different types of writing, including science writing, and these norms are sub-divided by grade level, making it a far more sensitive assessment of specialised text than a standard readability measure. Results show, however, that while the measurement tools may change, the appropriacy of the language of healthcare materials continues to be problematic: the researchers considered the language level a problem in all of their 94 research consent form templates, taken from a number of English-speaking countries. All measures exceeded recommended limits and all countries had material which exceeded their own local recommendations for readability.

In addition to consent forms, there have been several language studies of health materials that relate to chronic health conditions, such as asthma and rheumatoid arthritis. An overview of these studies is presented in the following section.

2.1.4. Chronic condition information and biomedical discourse

The content and language of health information materials for patients living with chronic conditions, such as osteoarthritis (Grime & Ong, 2007) and depression (Grime & Pollock, 2004) have also been investigated. One of the key findings of these studies is that there exists a disparity between the content of the materials and the information that patients say they want to have. Grime and Ong (2007) used a discourse analysis approach to look at the thematic content of six leaflets on osteoarthritis, finding, among other things, inconsistent advice, an emphasis on patient responsibility to prevent the progress of the disease, and that the leaflets were dominated by a ‘disease-discourse’ (p. 33). The experience of living with osteoarthritis was backgrounded, if it appeared at all.

The finding that written information for chronic conditions often sidelines a patient’s experience of living with that condition, prioritising instead a biomedical discourse, was also evident in Grime and Pollocks’s study (2004) on a leaflet containing information on depression that accompanied anti-depressant medication. The researchers found that the leaflet used six questions to deliver an overwhelmingly biomedical discourse, while people with depression interviewed by the researchers had a number of questions that remained unanswered by the leaflet. These questions related to self-help, stigma, and to a sense of self, none of which were dealt with at all by the leaflet. The authors conclude that patient information needs to be consistent and accurate with its information and involve the patient, their concerns and their experience far more than it currently does. This is the conclusion of an increasing number of studies (Halkett, Short & Kristjansen, 2009; Mathers, Chesson & McKensie, 2009; McCartney, 2013).

Involving the patient, as we have seen in Chapter 1, was proposed by the first handbook for patient information produced in the UK nearly 50 years ago, though is a policy that, while frequently recommended, is not always followed by patient information writers. Wright (1999) remarks that involving the patient means collaborating during the production of a leaflet, not simply asking patients at the end of the process whether they can understand the information contained within it.

That much patient information concerns itself solely with the biomedical is also seen in studies about studies of patient information: Dixon-Woods (2001) found that of the two principle discourses she identified in over 1,000 papers on patient information, it was the biomedical model that predominated. The idea that patient information, in containing primarily biomedical content, may be falling short of delivering the information that is needed, is relevant for the study described in this thesis. How much procedural patient information for radiography might be described as biomedical or biotechnical in nature? Or, if not biotechnical, what discourse predominates? The linguistic analyses I perform on my corpus, the results of which are reported in chapters 5-7, may well provide answers to these questions.

2.2 The PIL: Pharmaceutical patient information leaflet

To date, the pharmaceutical patient information leaflet, usually referred to as a PIL, has been the subject of considerably more interest from language researchers than has the procedural patient information leaflet (e.g. Fage-Butler, 2013; Haw & Stubbs, 2011; Hirsh, Clerehan, Staples, Osborne & Buchbinder, 2009). This may be because, while the content of pharmaceutical patient information in Europe is regulated by law along with the order in which that information is presented, how it is written and what non-obligatory content is included varies considerably, making it an interesting area for the language researcher. It is also the case that, by law, no pharmaceutical product can be marketed without containing an information insert. The fact that medication can kill as well as cure makes it vital that people take their drugs appropriately, which means they need to comprehend the accompanying information leaflets.

Many of the findings of language studies of PILs are relevant for this thesis as the healthcare material shares similar objectives: to inform and to instruct. The intended recipient is the patient, rather than the professional, and certain themes such as risks and benefits, are common to both types of information. Some of these studies are overviewed in the next section, while those that have directly informed my research will be presented and discussed in detail later in this chapter.

Findings from many studies of PILs suggest that, in spite of the efforts to improve the patient experience through legislation, many consumers do not feel fully informed by them (Haw & Stubbs, 2011; Raynor, Savage, Knapp and Henley, 2004). This is the case particularly with regard to interactions with other medication along with information regarding side-effects (Dickinson, Raymor & Duman, 2001).

Comprehensibility is the focus of many studies and is a constant concern. In common with patient information in general, many consumers say that they find the language of PILs too dense and complex (Askehave & Zethsen, 2014). There are even signs that leaflets are more complex now than they were in the past: Askehave and Zethsen (2014) reproduced a survey from 2000 to discover Danish consumers’ views on PILs. They found that fewer respondents read PILs in 2014 compared to 2000, and fewer respondents agreed that PILs are easy to read. Legislation, say the authors, may well be the explanation for the findings as it has rendered the PIL too regulated and too complex, a view echoed by Wright (1999), who says that listing drug ingredients at the beginning of the leaflets in over-the-counter medications, as per European directives, may well put off some people from continuing to read.

2.2.1. Risk in patient information

Along with the complexity of the vocabulary and the difficulty reading the text, another central theme in the PIL literature concerns risk: how it is expressed and how it is interpreted. Risk is frequently overestimated by health consumers (Berry, Knapp & Raynor, 2002; 2006; Pander Maat & Klaassen, 1994; Pander Maat, 1997), while risk expressed in writing is particularly problematic if imprecise frequency descriptors are used. Imprecise descriptors relate to adjectives and adverbs of frequency, such as *seldom, rare* or *common*. The problem with over-estimating risk in the context of healthcare is that patients may be less likely to take their medication as instructed if they feel that side-effects are too likely and too severe. As Berry et al. (2002, p 854) say, ‘If people are unable to estimate the risk of occurrence of side-effects, they cannot be expected to make informed decisions about medicinal drug taking.’

EU guidelines issued by the EC Pharmaceutical Committee in 1998 encourage manufacturers to express risk using five frequency descriptors as *very rare, rare, uncommon, common, very common.* rather than using numeric information. Other adjectives and adverbs expressing frequency, e.g. *rarely, sometimes, often*, are also used in packaging inserts. Using a word-only approach seems sensible in the light of consumers’ (and health professionals’) reported difficulty with numeric information and statistics that report risk (Gaissmaier & Gigerenzer, 2008; Keller & Siegrist, 2007) and the evidence that shows numeracy to be a strong predictor of comprehension of health materials (Hibbard, Peters, Dixon & Tusler, 2007). Patients with poor numeracy skills will find comprehension more difficult when the risk of side-effects is expressed statistically or in percentages. Nonetheless, research has shown that using the frequency descriptors mentioned above instead of numbers is also highly problematic. The findings of studies investigating the use of qualitative descriptions of risk, i.e. those using adverbs and adjectives of frequency in place of ratios or percentages, show that patients not only frequently over-estimate risk when they read frequency descriptors (Blalock, Sage, Bitonti, Patel, Dickinson & Knapp, 2016; Berry et al., 2004) they also show great individual variation in their interpretation and express this variation with irregular consistency (Pander Maat & Klaassen, 1994; Knapp, Gardener & Woolf, 2015).

It is not just the possibility of risk that is over-estimated, however, but how serious that risk is perceived to be: Berry et al. (2002) found that patients reading only verbal descriptors of risk not only over-estimated the likelihood of risk but also the severity of risk, and the risk to health of reported side-effects, while, at the same time, their reported intention to follow the treatment was considerably lower than the control group who were given numeric values only.

If both the use of words to express risk and the use of numeric information to express rick can be problematic, what is the effect of using the two approaches together? Using both approaches together is, it transpires, the current advice from the European Medicines Agency (EMA). However, the first study to examine the effect of this approach on estimations of risk finds that this combined approach is also problematic (Knapp, Gardener and Woolf, 2015) as patients consistently, and greatly, over-estimate the risk of side-effects when numeric and frequency descriptors are used together. The authors also confirm earlier findings that numeric information alone contributes to over-estimation which, they conclude, suggests that not only is further research needed on how best to represent risk, particularly on the verbal descriptors which seem to be hold greatest potential, but that government agencies and professional bodies should be very cautious when giving recommendations about the representation of risk in patient information, particularly when they recommend a combined approach.

2.2.2 Risk in radiography patient information

The concept of risk is relevant to my investigations of patient information for radiography, as radiation risk is a topic of central importance, as we saw in chapter 1. While radiography patients express a desire to have information (Mathers, Chesson & McKensie, 2009; Singh, Mohacsy, Connell & Schneider, 2017), it seems that most receive little or no information, including that relating to risk (Ukkola, Oikarinen, Henner, Haapea & Tervonen, 2017). Ukkola et al. (2017), in their study of the provision of patient information (oral and written) in a radiography department in Finland, found that the quality of information that patients received was poor, and that more than 90% of patients visiting hospital for a range of radiographic examinations received no information whatsoever about risk, with a similar percentage reporting that they had received no information regarding the benefit of the examination, either.

However, even when patients are informed of the risk for computed tomography (CT), (a very common radiographic exam that uses x-rays and thus radiation), and for other radiographic examinations, they show a general tendency to under-estimate the risk (Baumann et al., 2010; Rosencrantz & Flagg, 2015) and also show a poor understanding of the concomitant risk of radiation exposure. (Baumann et al., 2010; Singh et al., 2017). Singh et al. (2017) found that the majority of patients attending a Melbourne hospital for a range of radiographic examinations did not know that a mammography was a radiation-emitting examination. Disconcertingly, more than half of the patients in the same study expressed no concern about dose and radiation, though the majority also under-estimated the radiation risk of a range of radiation-producing examinations. This finding is in contrast to that regarding risk in pharmaceutical information, where, as we have seen earlier in this chapter, patients generally over-estimate the chance and severity of side-effects. Patients undergoing radiography seem poorly informed about the mechanisms of radiographic modalities and lacking in knowledge about radiation risk. The minimal information a patient receives prior or during a hospital visit may explain the large number of people who express little or no concern about risk (Singh et al., 2017); that is, they are completely unaware that there is any need to be concerned about such a thing. In pharmaceutical information, on the other hand, patients scanning the leaflet see a list of side-effects, common, occasional and rare, bulleted and bolded. Remaining unaware of risk is far less likely. The rising number of CT examinations ordered by doctors in the UK and many healthcare systems, however, makes ignorance about radiation risk a great concern. The rate of CT scans performed in hospital in the US doubled between 2000 and 2017, while in the UK, in the same period, there were nearly three times as many CT scans performed (OECD, 2018).

Just as with medication and pharmaceutical advice, patients cannot make informed decisions about their healthcare in the absence of comprehensible information. Clearly, patient information must be available as a matter of priority, whether it be oral or written. The inclusion of risk information for radiography is something that patients say they welcome (Singh et al., 2017) though there have been few studies that focus on how risk information is currently presented in radiographic patient information and how patients understand it.

I believe that many of the studies undertaken of pharmaceutical patient information could be undertaken on procedural information, including an assessment of the manner in which risk and benefit - the latter something that patients and professionals are reported to over-estimate (Hoffman & Del Mar, 2017) - is presented to the patient. Establishing the lexical characteristics of procedural patient information for radiography, as I do in this doctoral study, is the first step.

In this chapter thus far, I have presented an overview of the literature on PILs, reasoning that much of the findings of these studies will be relevant to other sub-registers of patient information, including procedural. I have also presented an overview of the literature concerned with other medical registers. As we have seen, many studies have been pedagogic in nature and focussed on the medical research paper or case notes, with a handful of studies related to the consent form and to information for chronic conditions. Given the sheer number of medical research papers published annually (indexed citations at MEDLINE for 2017 number more than 800,000, more than twice those in 1995 (National Library of Medicine, 2018)), and when one considers how many of these papers will be written by people for whom English is an L2, the academic focus in the literature is understandable. Other written registers, however, including procedural patient information, have yet to be explored to any approachable degree by language researchers.

I now turn to the second half of this chapter, a presentation and discussion of the studies that have particularly informed this thesis. The literature review that follows is divided into three sections. I begin with a relatively short section where I step back from both the subject of patient information and corpus linguistic studies of healthcare communication in order to consider the importance of collaboration between applied linguistics and medicine. Some of the methodological issues raised by cross-disciplinary collaboration will also be the focus. Two papers are discussed in this section, both from the same year: Candlin and Candlin (2003) and Roberts and Sarangi (2003).

The second section contains studies of the language of medical discourse, both spoken and written, that have used corpus linguistic techniques in their investigations. These studies begin with a discussion of Skelton and Hobbs (1999a). Skelton, an applied linguist, collaborated with a medical professional in this and in subsequent papers (e.g. Skelton & Hobbs, 1999b; Skelton, Wearn & Hobbs, 2002) which were written for the medical profession and all published in medical journals, e.g. The Lancet, the British Medical Journal (BMJ) and Medical Education. The literature review continues in this section with a presentation of several studies of healthcare communication that utilise the keyword approach, a corpus technique which has shown itself to be a valuable means of accessing areas of interest in the data, and one applied a keyword approach and a lexical bundle analysis to the investigation of a range of pharmaceutical registers. This latter is one of the very few published studies that have investigated lexical bundles in non-academic medical registers. Here, too, I present one of the very few studies of modal verb use in a range of medical registers.

The literature review concludes with two studies that are related specifically to patient information and the way in which it is developed and evaluated. Clerehan et al. (2005) is followed by a study by Zarcadoolas (2013) who calls for a new approach to healthcare materials, one that shows more understanding of how language works, how text is read and the role of pragmatics in making meaning. These papers take the view that using readability measures as the standard evaluation of the comprehensibility of patient information is misguided; both put forward arguments for an approach to the development and appraisal of patient information based on linguistic and communicative principles. I include these studies here because I believe that this doctoral thesis, with its investigation of the underlying linguistic characteristics of the register, can contribute to the kind of knowledge that Zarcadoolas (2013) believes should inform the production and evaluation of healthcare materials.

2.3 The applied linguist and healthcare communication

The papers reviewed in this section underline the value of research by applied linguists in the field of healthcare communication while, at the same time highlighting the importance of looking beyond the field of applied linguistics to other areas - social sciences, medical education, medical communication - where a lot of the research into language and healthcare is being conducted. These papers also refer to the importance of multidisciplinary research and of collaborating with healthcare professionals from the field when conducting research.

I found these papers convincing, and, as I have stated previously, much of the reading that has informed this thesis has come from fields other than applied linguistics and throughout the process of conducting research and writing up, I have sought information and advice from medical communication writers and medical professionals. The concerns I have had regarding my ability to conduct self-directed research into medical communication are expressed in 3.2 in my commentary on Roberts and Sarangi (2003). First, though, I present Candlin and Candlin (2003), a key paper that makes clear the value to applied linguists of research - and reading - outside of their field.

2.3.1 Candlin and Candlin (2003)

Summary

In this paper, Candlin and Candlin (2003) present an overview of medical language research and call for applied linguists to broaden their reading to include other disciplines, where considerably more research was being undertaken. The authors highlight some of fields outside of applied linguistics where medical communication research was being conducted, and reference certain journals where it is commonly published (e.g. *Social Science and Medicine*). They continue by summarising the key themes in the literature (e.g. risk; narratives and interpreting) and by assessing some of the methodological approaches used (e.g. grounded theory; semi-structured interviews and questionnaires). Referencing a number of well-researched studies from various fields, the authors also caution against some of the less methodologically-sound approaches taken, where data is dealt with superficially. The quality of the data is as important as the methodology used to process it, and it is here, in particular, that applied linguists can make a contribution.

The principal problem referred to in the paper's title, `Health care communication: a problematic site for applied linguists research`, is the outsider status of applied linguistic researchers; it is very rare to find researchers who are both medical practitioners and applied linguists. If applied linguistics really is problem-driven then it should also, say the authors, be outcome-focussed, and thus research should be undertaken wherever possible in collaboration with practitioners and with the end-users in mind. The end-user here is the medical professional or the patient.

Achieving a more inclusive and collaborative approach is not easy, though adopting a more open methodological stance and not attempting to fit health data to existing linguistic theories is a start, say Candlin and Candlin (2003). Results, too, can be presented in the language of the practitioners and end-users, while the research questions should not simply address the how, but also the why and to what purpose, echoing Roberts and Sarangi (2003), Adolphs et al. (2004) and Skelton and Hobbs (1999a), all discussed in this chapter, in highlighting the importance of the communicative context and purpose in healthcare language analysis.

Commentary

Candlin and Candlin (2003) is a key paper for benchmarking the developments of research into medical discourse over two decades, both in terms of methodologies used, the disciplinary areas in which the research is conducted and the themes of interest. It is no-longer the case that doctor-patient studies proliferate and there are an increasing number of studies and book-length works on communication in nursing (e.g. Bosher and Stocker, 2015; Henderson, 2016; Lu, 2018; Staples, 2015) though there are still very few studies of the discourse in other healthcare professional fields, such as radiography. It is also the case that the professional voice still predominates. Anton and Goering's book (2015), `Understanding Patient Voices` is unusual in that it focuses exclusively on patient discourse.

There have also been slower developments. Candlin and Candlin (2003) reference just one paper from the *English for Specific Purposes* Journal, that by Frank

(2000), expressing surprise at the lack of coverage of medical discourse in the journal. In 2018, the journal still devotes comparatively little space to the theme of healthcare discourse, and the papers that are published are generally concerned with academic medical language. Of course, it may be that this is a reflection of the priority given to academic research, which I have referred to earlier in this chapter.

To find out more about where studies in the field of medical language are published, I conducted a Scopus search that reveals the majority of studies continue to be published in journals from disciplines other than Applied Linguistics, e.g. *The Lancet; Social Science and Medicine; Journal of Advanced Nursing; Patient Education and Counselling*. This highlights the need for researchers to continue to head the advice given by Candlin and Candlin (2013) to read broadly. Some progress has been made, however, as there is now a targeted journal dedicated to the subject of communication in healthcare, *Communication and Medicine*, and a journal dedicated to the field of applied linguistics within the world of work, *Journal of Applied Linguistics and Professional Practice*. Both of these journals were launched in 2004, a year after the publication of this paper.

The question is raised in the paper of whether quantitative, questionnaire-based approaches or qualitative, narrative approaches are more appropriate for collecting healthcare data. The latter, say Candlin and Candlin, has certain advantages in that it can reveal `personal constructions of cultural relevancies` (p. 139). Elsewhere in the paper the authors concede that a combined approach can be ‘productive and explanatory’ (p. 143), referencing studies that combine discourse analysis with interviews (e.g. Dijkstra et al., 2002). What is surprising, however, is that the authors do not mention corpus linguistics or the use of frequency data in healthcare language studies. Corpus linguistic studies into healthcare discourse were in their infancy, and this may well be the reason for omitting referencing any of them, though there were a number of earlier studies into medical discourse, predating Candlin and Candlin (2013), that make use of both corpus linguistic techniques and qualitative analyses include Salager-Meyer (1994), Skelton and Hobbs (1999a; 1999b), Skelton et al., (2002) and Ferguson (2001). Only Skelton’s research, however, was concerned with oral interaction and was carried out in collaboration with medical professionals, something that Candlin and Candlin (2003) feel is essential. This may explain the exclusion of Salager-Meyer's (1994) paper and that of Ferguson (2001) (though not, of course, the exclusion of the work of Skelton and colleagues.)

Roberts and Sarangi (2003), the next paper to be discussed, published in the same year as Candlin and Candlin (2003), also foregrounds the status of the applied linguist who conducts research in medical language, and the value and difficulty of collaboration between language researcher and medical professional, though Roberts and Sarangi (2013) is primarily a description of the challenges that cross-disciplinary collaboration can bring.

2.3.2 Roberts and Sarangi (2003)

Summary

In this paper, Roberts and Sarangi's (2003) focus was the relations between two applied discourse analysts (the authors) and medical professionals from the Royal College of Physicians who had approached the authors for a consultancy. While collaboration was very much welcomed by all involved in the project and the authors echo Candlin and Candlin's (2003) call for more outcome-focussed collaborative work between linguists and professionals, they also concede that this kind of project is not without difficulty for both parties. Referencing the feedback and comments received during the write-up of their research for a medical journal, the British Medical Journal (BMJ), Roberts and Sarangi (2003) describe some of the theoretical and methodological challenges they experienced, and the reflection that these challenges prompted.

At the heart of the issue is the 'outsider` status of the sizeable majority of discourse specialists and applied linguists working in healthcare research, say the authors. By claiming relevance for their research in a professional field other than their own, these outsiders face issues of `identity, roles, authority and credibility` (p 339). All too often, however, this status results in research that is not of direct practical relevance. What's more, say Roberts and Sarangi (2003), `applied` in applied linguistics is generally used to refer to real-world settings (e.g. the workplace) rather than real-world outcomes. The traditional relationship between applied linguistic researcher and their subject is also questioned. Roberts and Sarangi (2003) believe that to produce research with useful, practical outcomes, a less imbalanced approach to collaborative language research is needed, one where every aspect of the study is developed in conjunction with the research informants: study design, presentation and dissemination included. This is not as easy as it sounds, however.

A problem occurs when the type of research undertaken, and the language used to present and disseminate its findings, is perceived to be lacking in rigour and relevance. Medical research is still very much based on `hard` science models, and where clinical trial research is held up to be the gold standard of scientific enquiry. The kind of research that discourse analysts engage in, with its qualitative and ethnographic basis, are ‘largely treated with suspicion if not contempt’ (p 341). The emphasis in medicine is still very much on the ‘bio-technical modal’ (citing Becker et al. (1961)), notwithstanding the fact that journals such as *Medical Humanities* point to the existence of more humanistic models of medicine. In scientific research, everything begins with a hypothesis; the open-ended research that the authors were engaged in did not convince many of their medical collaborators - who, otherwise, were wholly supportive of the authors’ work. Roberts and Sarangi (2003) struggled to convince their colleagues, for whom hypothesis proving or refuting was the norm, that `understanding` as a research outcome was a valuable one. Altering their preferred style of research reporting was also necessary, as the researchers’ hedged claims were considered to be lacking in authority by reviewers more comfortable with the categorical style of medico-scientific writing. Other issues encountered in the collaborative process concerned the use of language from the field of discourse analysis that was not understood by medical professionals, and as a result was perceived to be inclusive and institutional, creating quite the opposite effect from that intended. This, and the other issues reported, highlighted the difficulty of disseminating research findings to an audience unfamiliar with, or even unappreciative of the methodology and language considered appropriate in the field of the researchers.

Commentary

There are many issues raised in this interesting paper about the methodological and discursive differences that collaboration across disciplines may encounter. I also experienced difficulties due to the cross-linguistic nature of the research, though these difficulties were not evident to me at the beginning of the process. Unlike the other papers in this chapter, and perhaps unconventionally, I include Roberts and Sarangi (2003) not as a study that informed my research design or methodology, but as a paper that provides an explanation for some of the issues I experienced during the research process, particularly when writing up this thesis. These issues are undoubtedly common to many, perhaps all, cross-disciplinary studies.

Roberts and Sarangi (2003) also encouraged me to reflect on the assumptions that I may have had concerning the usefulness of my research, and on my research outcomes. One of the motivations for this research, as I stated in chapter 1, was my belief in the importance of patient information for the patient. This research, however, was self-directed and, while medical professionals were consulted for information and advice during the research process, the study was not carried out in collaboration with medical professionals. As I mentioned earlier in this chapter, many studies of medical discourse that are published for a medical readership have a lack of linguistic methodological detail that an applied linguist may find a frustrating omission. These details are omitted as they would be a barrier to comprehension for a medical readership. It is a fact, however, that as an applied linguist I am the ‘outsider’ that Roberts and Sarangi (2003) refer to. I am neither radiographer nor medical information writer. My findings and my conclusions may not have the value in the professional world that I would like them to have.

But the reverse is also true. It may be that in exploring studies from outside of the field of applied linguistics, from social science, medical education and medical communication, to name but a few, the value of my research may be questioned by applied linguists. Roberts and Sarangi (2003) found that their discourse analysis methodology and the qualitative, ethnographic research culture failed to impress their medical collaborators, whose research culture was hypothesis-based and biotechnical in nature. Roberts and Sarangi’s research culture was even ‘viewed with suspicion if not contempt’. (p. 341). My research has been viewed in a similar way from time to time, particularly by corpus linguists and those more interested in computational linguistics, some of whom have stated that applying corpus techniques to language analysis does not make one a corpus linguist. I am not, however, a discourse analyst, nor a sociolinguist, and while the corpus methodologies I have used are well established methods in a research area that has been referred to as ‘applied clinical linguistics’ (Adolphs, Brown, Carter, Crawford & Sahota, (2004)), it is not a name that has had much of an impact, perhaps because clinical linguistics is quite a different field of investigation, with its focus on language disorder. I have suggested that corpus analysis of healthcare discourse may be part of the field known as medical humanities, though not all agree. The upshot is, I believe, that being part of a named research tradition with its own discourse and methodologies (such as keyword extraction), would help establish the validity of corpus approaches to healthcare communication.

2.4 Corpus approaches to healthcare discourse analysis

We now shift our focus from the methodological issues surrounding language research in the field of medicine to the use of corpus techniques in the analysis of medical discourse.

The first use of corpus techniques for the analysis of medical communication is generally regarded to be the study by Thomas and Wilson (1996) who compared the three different approaches of discourse analysis, the use of questionnaires and ‘computer content analysis’ to the investigation of the interaction between people with cancer and healthcare professionals. Their findings were that the computerised approach (a programme called ACAMRIT) did not tell them anything they did not already know; it did, however, speed up tremendously the investigative work that analysing the corpus entailed. ACAMRIT is described as an automated content analyser with a number of additional modules that sets it apart from simpler content analysers, including parts-of-speech tagging; semantic tagging and conceptual analysis. This is not a study, however, that foregrounds the practical outcomes of corpus techniques: its aim was to show what could be done with cutting edge, computerised analyses of transcribed data, and both researchers were linguists. The study that follows, however, does foreground medical concerns, and unlike Thomas and Wilson (1996), Skelton and Hobbs (1999a) was written for medical practitioners and published in a medical journal, the *Lancet.*

2.4.1 Skelton and Hobbs (1999a)

Some of the earliest work using a corpus of doctor-patient interaction was carried out by applied linguist John Skelton and medical colleagues at the University of Birmingham using Cobuild software.

Summary

Skelton and Hobbs (1999a) is essentially a descriptive paper on the value of corpus techniques - the authors refer to it as concordancing - and of combining quantitative and qualitative approaches for the study of professional language. At that time, such an approach was entirely new in the context of medicine. The utility of quantitative data according to the authors is its capacity to capture linguistic aspects of the consultation that may have been taken for granted and thus remain under-researched. The importance of qualitative analysis, on the other hand, is that it provides information about meaning that frequency information alone cannot. As the authors say 'if words were like numbers, it would be hard to understand why we bother with both'. (p109)

The corpus was made up of 40 primary care doctors conducting 373 consultations. Skelton and Hobbs (1999a) performed three type of analysis in order to present to their medical readership the different ways that concordancing can be used in the analysis of a spoken consultation. The first was a quantitative study of doctors' use of jargon, the second applied a partial, quantitative operational definition to the hitherto qualitative concept of doctors' power, while the third analysis aimed to show the value of the approach in an investigation of the way threats are diminished in medical consultations.

The first analysis, to investigate the doctors' use of jargon, found that doctors did not use unexplained words that were considered technical or medical with their patients. The assumption that doctors did routinely use technical language with their patients without explanation was simply wrong, say the authors. In the study, when doctors did use such words, they were often lay technical in nature, such as *paracetamol*, which the authors considered non-problematic for the patient. When more complex words were used, a definition was often provided by the doctor. The word was often prefaced with a phrase that signalled the word was about to be defined or that it would not be, as it was not worth defining. The second analysis concerns power relations. Operationalising one definition of power imbalance, that of markers of diffidence connected with social inferiority, the authors select the use of past tense to reference present concerns. Skelton and Hobbs (1999a) analyse the patients' use of the past tense in these contexts. 11 uses were found when patients were expressing worry. The phrase *I was wondering* was also highlighted as it appears with some regularity in the corpus, as in *I was wondering if you could just give me the sick note (p. 110).* The third analysis found, among other things, that minimisers such as *just* and *little* are used fairly frequently by doctors to encourage the patient and to diminish a threat, which may be the threat of a potentially serious health condition or the threat of a potentially embarrassing examination, e.g. *can I just have a quick look*.

Commentary

Skelton and Hobbs (1999a) is a study that introduced the potential of corpus analysis to medical practitioners. Its value is found particularly in the questions that are raised by the authors in the discussion section, and the suggestions for future research. Before I come to these, however, a word about the limitations of the paper.

Three different analyses are reported in Skelton and Hobbs (1999a) though each is short on methodological detail. For an applied linguist this is a little frustrating. We do not know the size of the corpus, only that it contains 373 consultations. Where detail is given, it is principally related to the statistical calculations the Cobuild software uses, particularly the Mutual Information (MI) scores. MI scores relate to the probability of two variables - in this case words - appearing together. It measures the strength of association between words and is often used in corpus software programmes such as Sketch Engine to extract collocations. A t-test was also used by the software. The emphasis on explaining the statistical processes underpinning the quantitative analysis may well be explained by the need to explain clearly (avoiding any linguistic jargon) what corpus software does - remember, such an approach to the analysis of a medical consultation was entirely new to the world of medicine. The focus on the statistics may also be an acknowledgement of the attitude to research of the largely medical readership. As we saw in the discussion of Roberts and Sarangi (2003), medical and scientific research is hypothesis-driven, values observable proof and is generally categorical in its claims; a medical readership can be dismissive or even contemptuous of research approaches that are not seen to have academic rigour. Statistical information is foregrounded in this paper, perhaps, to counteract any such suspicion. It would be a more satisfying paper for an applied linguist, however, if there was more information given about the methodology and more data examples. We are told, for example, that past tenses were used to express present worries 11 times, but only one example is given. This is an interesting area and more examples would have been interesting. We do not know, either, if one person or 11 people were responsible for the utterances, nor can we appreciate whether 11 times is to be understood as frequent, as the information needed to make that decision is not given.

The issue of academic rigour also relates to hedging. Hedging is uncommon when discussing research findings in medical and scientific research (Roberts and Sarangi,2003); this may well explain the categorical claim in Skelton and Hobbs (1999a) which some applied linguists may find overly strong though a medical readership may not. The claim relates to the first analysis, that of doctors' use of jargon. The authors claim that the assumption that doctors do use medical jargon in their consultations is wrong, adding 'for the group of practitioners we analysed' (p. 110). The addition suggests that the authors also found the claim overly strong. Nonetheless, my reading of the study tells me quite the opposite. Skelton and Hobbs (1999a) does show that doctors use jargon. Not only do the practitioners use medical terms but they often seem to be aware of it, adding a preface containing a signal to indicate that a complex word is about to be used and that a definition will follow, or that a definition will not follow, as the practitioners believe that there is need to translate the term.

The authors also disregard the use of terminology that was considered (by the authors) to be unproblematic, on the basis that these words were very likely known by the patient, e.g. *paracetamol*. There are several issues raised here, some of which the authors point to in their discussion. The principal issue of whether a term is known by the patient is a complex question. We have seen already that both professionals and patients over-estimate the latters' understanding of many words (e.g. Chapman et al., 2003). How a word is understood by a patient - or by a professional of course - is the central issue here. When practical, real-word outcomes are the focus, the question of whether the patient possesses sufficient understanding of a medical word to comprehend the information being relayed (and to comprehend it more or less as the professional intends) is more important than how the word itself should be classed i.e. lay technical or semi technical or fully technical.

The importance of understanding is raised by the authors in the discussion section who propose further investigation of what we mean by comprehension in a medical context, and how the training of doctors can be informed by such research. While I disagree with their conclusions that jargon use in consultations does not cause miscomprehension (this study does not provide sufficient evidence to reach such a conclusion), I think Skelton and Hobbs (1999a) convincingly demonstrate that further investigation of the concept of understanding in medical interaction is warranted. I do not know whether the readership was convinced by their presentation of concordance for medical language analysis, however.

The authors propose several areas for further research, including the use of the language of uncertainty by doctors and patients, of advice and instruction, and of opportunistic health advice, all of which can be researched using corpus techniques, while underlining the fact that corpus investigations also lend themselves to the possibilities of investigating medical interactions from a range of demographic variables such as age, gender and social class. Some of these variables have been investigated in the studies that follow, including gender and age. These later studies are primarily corpus-driven, using comparative data to generate keywords and phrases that reveal patterns in discourse that may otherwise go unnoticed. Comparative data was not used in Skelton and Hobbs, 1999a and items were pre-selected for analysis. While some of these studies that follow were written for a medical readership (e.g. Seale & Charteris-Black), others were written primarily for applied linguists (e.g. Atkins & Harvey, 2010). The level of methodological detail, therefore, and the emphasis on real-world outcomes, varies accordingly.

Establishing the role and validity of corpus approaches to healthcare is also the focus of Adolphs et al. (2004), the study that we turn to now. Adolphs et al. (2004)

functions as a bridge between the papers already presented, and the later, corpus-driven studies that follow. Adolphs et al. (2004) echoes some of the concerns of the earlier papers regarding the status of an applied linguistic researcher in the field of healthcare, but also promotes the value of corpus-driven analysis and the keyword method for investigating a healthcare corpus.

2.4.2 Adolphs, Brown, Carter, Crawford and Sahota (2004)

Summary

In this multi-authored paper, Adolphs et al. (2004) sought to establish the role of corpus linguistics and data driven learning as a means to better understanding the language of healthcare and healthcare interactions. The authors propose a convergence of various approaches from the fields of health and social science with those from applied linguistics, namely conversational and discourse analysis and corpus linguistics, referring to this new convergence as `a kind of applied clinical linguistics` (p. 25). The primary motivation for this was to bring new insights and discoveries that would benefit medical practitioners and patients. The authors’ view is that if language is both interpersonal and transactional, then it becomes vital to understand how the language in a medical encounter is tailored to the recipient.

The paper presents an analysis of a small corpus of spoken interaction between NHS helpline staff and callers (researchers playing at being patients), to show what kind of information can be revealed, and why that information might be significant within the context of healthcare. Using a keyword analysis to begin with (for an explanation of this approach, see chapter 3), items of interest were categorised into the following groups: negatives, imperatives, pronouns, vague language, affirmations/positive backchannels, directives. Medical terminology was ignored. These linguistic features were then investigated in greater detail within the corpus and examined to see where and when they were most frequent. This investigation established a link between certain linguistic features and particular phases of the consultation and highlighted what the authors describe as `an overarching tendency` for the health professionals to use politeness strategies and the language of convergence when speaking with the callers. This was often seen as a downplaying of the impact of the advice. The results, say the authors, illustrate just how useful an exercise uncovering linguistic features of discourse is as it can uncover communicative patterns that can then be linked to subsequent outcomes, desirable or otherwise. This is particularly related to issues surrounding compliance and concordance in healthcare.

Commentary

Adolphs et al. (2004) makes a convincing case for the need for corpus-based research into the language of healthcare, research that is both theoretically interesting and that has a focus on practical outcomes. Drawing on research findings from a good many non-linguistic areas (the vast majority of references in this paper come from journals outside of applied linguistics), this was not the first paper to call for a more inclusive approach to research into the language of healthcare, as Candlin and Candlin (2003) and Roberts and Sarangi (2003) had done so a year previously, but it was one of the first to give a clear, practical example of the kind of data that can be uncovered, even in small corpora, through the application of the keyword method followed by a closer, qualitative analysis.

The quantitative and qualitative approach, as demonstrated in this study, goes beyond word frequency and word distribution information and permits access to subtler, more complex linguistic patterning. The analysis of patient information in this thesis was undertaken for the same purpose: to uncover the kind of hidden, linguistic features of the discourse which may have a practical relevance to practitioners and patients. It is important, too, that Adolphs et al. (2004) were interested in a range of linguistic features but not in medical terminology/vocabulary, showing that general vocabulary is as useful an area of study in medical communication as is medical/technical vocabulary.

Small, in comparative terms, at 35,014 words for the health professionals' contributions, the corpus used in Adolphs et al. (2004) was sufficiently specialised and coherent for the size not to be of much concern. It was, say the authors, a 'preliminary vignette' into the entire NHS Direct corpus and served as a means to show what can be achieved using the techniques described. Concerns relating to corpus size are considerably less relevant if the corpus is specialised and targeted and as Biber (1990, p269) said, `descriptive linguistics should not be intimidated by the ‘need’ for larger corpora`.

The findings of the study show that health professionals use a range of politeness strategies and are highly likely to mitigate the advice they give to callers, even when quoting an authoritative voice such as the British Medical Association, who were almost certainly more categorical than the health professional referencing them. The findings also reveal a high frequency of modal verbs *can* and *may* which is significant for this doctoral thesis. *Can* and *may* are also the most commonly used modal verbs in patient information, which, along with mitigators such as *just* are used to reduce the threat of the advice or instruction; *can* and *may* offer the advice as an alternative that the patient is free to follow - or to ignore. As I discuss in Chapter 6, and in the context of my study into the use of deontic modal verbs (those used for instruction and obligation), this linguistic behaviour seems to be particular to healthcare advice and not to consumer or legal advice where we find considerably less mitigation.

Adolphs et al. (2004) is one of the earliest studies to show the utility of using corpus techniques, particularly a keyword analysis, in revealing important aspects of spoken medical interactions that may otherwise have remained unnoticed. The paper was published in the *Journal of Applied Linguistics and Professional Practice*, a journal which aims to advance the interdisciplinary nature of applied linguistics research and which had been launched in the same year as Adolphs et al. (2004). The paper’s authors came from three different disciplines: applied linguistics, health communication and medicine. The next paper I discuss, Seale, Ziebland and Charteris-Black (2006) is also a good example of interdisciplinary research, with authors from the fields of medical sociology and applied linguistics. The study, published in *Social Science & Medicine*, and which utilises the keyword technique, is an analysis of the impact of gender on the language used to talk about cancer.

2.4.3 Seale, Ziebland and Charteris-Black (2006)

Summary

In this study, Seale, Ziebland and Charteris-Black (2006) use a keyword analysis to investigate the impact of gender on patients' experience of cancer. The study involved a keyword analysis of forum postings relating to prostate and breast cancer, and of 97 interviews with people with cancer. After an overview of the literature, the authors (who are careful throughout the paper to reject the notion that gender behaviour is fixed), suggest that a summary of the literature thus far of gender difference in language use, communication preferences and illness behaviour might be that 'men tend to focus on information and women on emotional support. Women draw on wide informal social and family networks when ill, whereas men deal with things on their own more, perhaps with the support of their wives, or in collaboration with doctors' (p. 2580).

The study is in two parts. The first part compares the results of a keyword analysis of a corpus of interview data (at 727,100 words and including only the patients' words) with the results of an analysis of the transcripts using the more conventional thematic approach that is common to social science. In the second part of the study, the keyword approach is applied to a corpus of over 12,000 forum postings by people with cancer (PWC). The corpus contained over 1.6 million words. The objective was to further validate the keyword approach, and to extend the existing evidence base of gender differences in the experiences of these two common cancers. The thematic content analysis of the interviews found men seem to prefer to seek information, and women prefer to engage in social and emotional support when online. These findings are not to be considered absolute difference, however, say the authors. The results of the keyword extraction of the interviews paint a far richer picture. Semantic categories were devised by examining the keyword in context (KWIC), and words categorised accordingly. Words that could be associated with two of the coding categories were entered into both categories and marked as a 'split' word. All 300 keywords were examined, though words that were used less than 10 times were disregarded.

Keywords related to three semantic categories of Treatment, Procedures and Tests, and Symptoms and Side effects were very common in the interviews with men. 74 keywords fell into these three categories, compared to just 18 for the women interviewed. There were a number of other categories where the difference was just as significant. Men used more keywords relating to the treatment, to specialist medical staff and to medical procedures. Women, on the other hand, were more likely to refer to themselves and named family members, and, unlike men, they frequently used superlatives, e.g. *lovely*, *amazing*, *wonderful*. Of particular note in this study are the differences in the keyword category Feelings. 20 keywords were extracted from the interviews with women; in contrast, only two words appear in this category from the men: *concerned and embarrassment.*

The results of the keyword analysis of the web forums confirmed the differences found in the analysis of the interviews, but also revealed new differences not seen in the interviews. The authors conclude that the keyword approach has a number of significant advantages to the analysis of large datasets, such as internet postings. It is fast, is independent of the views of the researchers, at least in the extraction stage, and sensitive to context. The technique has an advantage over traditional qualitative methods in its capacity to reveal areas of interest to the language researcher that would otherwise remain unnoticed.

Commentary

This is a very interesting study, and the first of the studies discussed to be concerned solely with patients. It is also the first study discussed in this chapter to investigate internet forums, and the first that relates to gender and language use. This paper introduces the keyword technique to an audience of social scientists, and as we have seen, different disciplines necessitate different approaches to the disseminating of results. This is a study that is satisfying for an applied linguist to read: there is a good level of detail relating to the methodology, and better examples of the data given than we saw in Skelton and Hobbs (1999a). Nor is there an overly long discussion of the statistical measures involved in the keyword analysis, presumably because social scientists are less interested in this kind of detail.

Seale et al. (2004) were interested in learning how forum conversations by women about breast cancer differed from those by men about prostate cancer. The comparative corpus used was not the standard general corpus, but the other dataset: each set of forum conversations was used as the reference corpus for the corpus under investigation, i.e. the corpus of breast cancer conversations had, as a reference corpus, the conversations about prostate cancer, and vice versa. A reference corpus is usually a general corpus and is chosen to be representative of general language. Forum postings from people with cancer are highly specialised, however, and the researchers were intent on showing how each forum differed from the other. The idea that specialised language requires a specialised corpus has been expressed by several researchers: there are several studies that use either a dual reference corpus or a single, domain specific reference corpus (e.g. Baker, 2004; Goźdź-Roszkowski, 2011 and Grabowski, 2015). Goźdź-Roszkowski (2011) in justifying his choice not to use a general corpus says, 'comparing a range of specialized genres[1] from the same domain against a general reference corpus would inevitably lead to obtaining finding which may be highly homogeneous and probably valid for legal language in general but would not help identify features unique to a particular genre' (p36). The same can be said for any specialised register under investigation, of course.

In this thesis, I use a specialised corpus of general radiography in my keyword extraction, in addition to an extraction using the BNC, reasoning that each will give me different kinds of information relating to the differences between patient information and general information on the one hand, and general radiography on the other. This is discussed in more detail in Chapter 3, where I present my corpora and methodology, and in Chapter 4 in my presentation of the keyword analysis and its findings.

The corpus of postings in Seale et al. (2006) contained 1.6 million words and, while we do not know how many postings came from each forum, if we assume they were fairly evenly split, the size of each corpus was around 800,000 words. Opinions differ regarding an appropriate size for a corpus should be: given the nature of the

[1] Gozdz Rowkowski uses genre as I use register in this thesis.

topic and the number of postings (around 12,000 in total), I believe that 800,000 is a perfectly acceptable size. The authors make no claim about the generalisability of their findings and are keen to make clear their belief that the concept of gender is not fixed, adding that expressions of illness may also be highly individual, varying from individual to individual. Nonetheless, this study's findings are significant, and make a substantive contribution to the literature on the use of keyword techniques in the analysis of medical discourse. The findings related to gender differences in the way people experience a disease or health condition are important and relevant to this thesis.

Gender difference in information-seeking behaviour is a topic that is well-researched (e.g. Bidmon & Terlutter, 2015; Ek, 2013; Rice, 2006; Rutten, Squiers & Hesse, 2006). These studies confirm not only the existence of different information seeking behaviour, and a difference in the type of information being sought, but also that women are generally more likely than men to engage in online health information seeking. As Ek (2013) says 'When it comes to health, women seem to be more engaged, more involved, more attentive and apparently better-informed decision-makers.' (p742). Seale et al. (2006) found slightly more men than women reported using the internet for health information seeking, though I think we can disregard this as the difference was not significant (38% to 33%), not all interviewees were asked about their internet usage and this was not the theme of their study.

Surprisingly, however, in spite of the studies showing how gender relates to the frequency of health information-seeking online and the differences in the information being sought (and wanted), gender is not always considered to be a determinant of healthcare and is very rarely considered when producing health information. In an analysis of five internationally recognised health promotion frameworks, Gelb, Pederson and Greaves (2011) state that 'although gender was at times mentioned as a determinant of health, gender was never identified and integrated as a factor critical to successful health promotion.' (p. 445).

There is nothing fixed about the linguistic performance of gender (Seale and Charteris-Black, 2008), but if men and women tend to look for different information and focus on different aspects of health, illness and treatment, it seems reasonable to consider whether existing patient information, such as that in the corpus used in this thesis, is appropriate to these different needs. Additionally, given that sex and gender are rarely considered either as determinants of healthcare or when producing healthcare information, the question is raised of how inclusion criteria for information is decided?

We have already seen earlier in this chapter that many of the studies of PILs and information for chronic conditions find that they fail to deliver what patients say they want to read. Studies also show that a biomedical discourse predominates in much patient information. A biomedical discourse relates to the body, the disease, treatment, medical procedures, symptoms and side effects, precisely the areas that Seale et al. (2006) find foregrounded in the men's discussion of prostate cancer.

With regards to the field of radiography, we do not know what kind of discourse predominates in patient information, or whether it contains the kind of information patients want to read, though we have already seen that patients attending radiography departments often seem remarkable uninformed regarding radiography examinations and associated risk. The keyword analysis that I describe in Chapter 4 can contribute to our understanding of the first question, while further studies are needed in order to answer the second question.

Keyword studies of healthcare discourse have also investigated age as a variable, and the interaction of age with gender. The latter was the focus of Seale and Charteris-Black (2008), the next study that will be discussed.

2.4.4. Seale and Charteris-Black (2008)

Summary

Seale and Charteris-Black (2008) investigated the linguistic expression (as illness narratives) of the experience of a range of cancers in men and women, across three age groups, older, mid and young. The corpus of interviews amounted to just over 1 million words, made up of 102 interviews. A keyword analysis using Word Smith tools (Scott, 2005) was carried out, followed by semantic tagging using WMatrix (Rayson, 2005). This permitted further analysis of the frequencies of the semantic groups. These two analyses informed the choice of keywords that were then investigated qualitatively in the corpus.

The results showed that older men living with a serious illness viewed the medical system as an expert system and saw their condition as a problem to be fixed, alone, and with ‘medico-scientific solutions’. (p. 1025). This finding supports that of previous studies (e.g. Seale, Ziebland & Charteris-Black, 2006). Older women used more words related to social networks (including the absence of one), using more keywords from the categories Pets, Living alone, Death and Religion than men, though women in the Mid-age group spoke slightly more frequently about death than either younger women and older women. Death was a key topic for men in the Older and Mid-age groups, too. Findings also showed that women used considerably more words related to feelings and self-revelation, whereas men used more words from the categories Strange/Weird and Worry/Concern.

One semantic category, Young Style, related to the discourse of younger people, which contained a large number of adverbs (e.g. *basically*) and minimisers (e.g. *just*) and differed quite markedly from the other two age groups. Older men in this study did not use any of the words that appeared in this category. The authors conclude that older men bring considerable social confidence to their interactions in the medical system, a confidence that young men do not yet possess. In contrast to the literature that finds older men increasingly dependent on their wives and lacking in social support, Seale and Charteris-Black (2008) suggests that older men wield a social power and confidence that does not require the maintenance of strong social networks.

Commentary

This keyword study raises some very interesting questions regarding the role of information in the experience of illness. One particularly relevant finding in the study, which will be picked up again in Chapter 4, relates to the differences in the naming of healthcare professionals in men and women’s illness narratives. Seale and Charteris-Black (2008) found that men, in general, were significantly more likely to refer to *specialists*, *general practitioners* and *consultants* than were women. Older men in particular were more likely to mention these professionals as well as other named medical specialities. *Doctor* was used by all subjects more or less equally, suggesting that men, and older men especially, give particular importance to the specialist knowledge of the medical professionals they interact with. *Nurse*, in contrast, was referred to significantly more often by women.

The study’s findings also have implications for our understanding of young people’s experience of serious illness, suggesting that young men in particular lack the social power and confidence that their fathers and grandfathers possess, and, as a result, are unable to talk about their illness in the same way. The category Young Style contained the following words, the vast majority adverbs: *basically, obviously, actually, probably, really, gonna, stuff, yeah, yeh, like, sort of, to be honest, you know, just, kind of.* It is of particular note that young men used these words much more than young women, and the difference was statistically significant. Many of these adverbs in speech are associated with indirect strategies that reduce the illocutionary force of face- threatening speech acts (Brown and Levinson 1987). Even those adverbs that initially appear to be used as intensifiers - *basically, obviously* - were found in the text to be used by young men to minimise the threat of something that was potentially frightening and very serious:

> um well I was diagnosed when I was just about 15-and-a-halfum with leukaemia which we later found out I had the Philadelphia chromosome as well which can’t be treated just by uh radiotherapy it has to, you have to have a bone marrow transplant basically (YPC10 male, aged 19 years). (Seale and Charteris-Black, 2008, p. 1033)

The finding suggests that the hedges and minimisers used by young people in relation to illness can often hide anxiety and fear, rather than being demonstrations of disinterest or lack of worry. This has important implications for our understanding of how young people, and perhaps particularly young men, cope with a serious diagnosis like cancer and for the training of healthcare staff working with young people with cancer. It also underscores the importance of targeted healthcare information and healthcare advice forums for young people. Indeed, healthcare advice forums for young people have been investigated in several key corpus studies of healthcare discourse, some of which I shall now present.

2.4.5 Harvey et al. (2008)

Harvey at al. (2008) was one of the first studies to look at young people's e-communication about healthcare, with the aim of investigating the communication difficulties experienced by young people when discussing their health concerns. A corpus of around 1 million words was used, made up of over 62,000 messages from young people to Teenage Health Freak, a doctor mediated, health advice website, (which, at time of writing, appears to be offline). The messages were collected over a period of 24 months. The corpus was analysed using the keyword approach, an approach common to several of the studies already presented in this literature review. The comparative corpus used for the keyword extraction was the general spoken part of the British National Corpus (BNC), resulting in 1160 keywords. Keywords that related to communication, verbal interaction, and to advice and information seeking were highlighted, and a collocational analysis was carried out on these words. This involved the calculation of the Mutual Information (MI) score of the common collocates of the selected keywords. MI is a measure of the strength of association between words, and for general language, is considered very reliable. The third and final stage involved an examination of these keywords and collocations in context.

The results of the analysis were that young people not only use a higher frequency of words related to communication, e.g. *tell*, *say*, *talk*, *answer*, but that these words were overwhelmingly used with negative collocates, Adjectives *afraid*, *scared*, *worried*, *embarrassed* and *stressed* were very common collocates of the communication verbs. A second finding revealed that the young people commonly experienced difficulties in communicating their concerns to doctors and family members, with one reason being the fear that doctors would not respect confidentiality. A general ignorance of the medical system and the roles of healthcare professionals were also revealed by the keyword analysis.

Commentary

This study was the first to investigate adolescent health communication concerns using corpus linguistic techniques. Authenticity was retained, too, as the messages (referred to as emails in the study) from the young people had not been edited by the website and thus the original wording and the nuances of individual expression were retained.

The authors are keen to point out that theirs was not an ‘outsider’ perspective, a remark that references the concerns voiced in both Candlin and Candlin (2003) and Roberts and Sarangi (2003) regarding the role of the applied linguist when undertaking research of healthcare discourse. Harvey et al. (2008) did not decide what issues needed to be researched; the issues were revealed in the adolescents’ communication, and the study was unusual in that it focussed solely on patient-initiated action. This is one of the obvious advantages of the keyword technique in language analysis. There is, of course, researcher-bias in the decision to focus on certain keywords and not others, or to focus on one semantic area, those related to communication in this case, but this is not the same as deciding *a priori* what will be investigated in the corpus.

The findings of this study were an important contribution to the literature on adolescent health and have obvious implications for the training of people who work with adolescents - and not only healthcare professionals. The study revealed great anxiety on the part of the adolescents about confidentiality, and their fears about their health concern being disclosed.

The study also contributes to the literature related to e-health, in particular to that on computer-mediated, doctor-patient communication. In spite of the fact that a very large percentage of young people had used the internet to access health information in 2008 (75% according to Harvey et al. (2008)), and notwithstanding the fact that the benefits of email consultation were obvious to some health practitioners (Car & Sheikh, 2004a; 2004b), Harvey et al. (2008) make reference to the resistance they perceive from many health professionals to the application of the internet and email to healthcare, and the slow response to the changing world of medicine that they feel many medical general practitioners are guilty of.

A decade has passed since this study was published, and a google search for the terms *e health* and *adolescent* result in many information sites, apps and research studies that relate particularly to depression and mental health issues in young people. E health has also been joined by M health: healthcare information, apps and tests on the mobile phone or as wearables. In terms of public healthcare, there is evidence that the e health revolution faces several barriers that slow its progress. A 2015 study by Currie, Philip and Roberts found that while age-related disparities in internet use had declined in the UK, and that a number of e health NHS projects had already been successful rolled out nationwide, barriers that slow the progress of e health included capacity issues relating to digital infrastructure, budgets, and technological literacy of both staff and patients.

Adolescent healthcare, and the importance of understanding the specific healthcare needs of young people, is a theme taken up by a number of subsequent corpus linguistic studies, three of which I will overview briefly in the following section. All three studies involved the Teenage Health Freak website, referred to above in Harvey at al. (2008). The studies all concern online communication in healthcare forums, where professional advice is given to young people, and all use the keyword method.

Harvey, Locher & Mullany (2013) investigated forum postings in both a UK and US healthcare advice forum, aimed at young people. The researchers carried out a keyword analysis and looked specifically at discourse related to HIV/AIDS. The study painted a rather bleak picture of the effectiveness of public information strategies, as it found that young people were confused and misinformed about a number of important aspects, including how the virus is transmitted and the difference between HIV and AIDS. The results may go some way improving information initiatives directed at young people,

Weight, eating and body discourses were the themes of a study by Mullany, Smith, Harvey & Adolphs (2015). The corpus contained just over 2 million words and was made up of 113,480 advice requests sent to the site over a 5-year period between 2004–2009. The findings are that weight and eating is a persistent concern among girls, in particular, between the ages of 11-16. The researchers saw a peak in advice requests related to weight and eating at the age of 12, with a second, smaller increase in advice requests on the topic at the age of 16. This finding, the data that shows knowledge gaps at different ages, and the alarming attitude that some young people show towards anorexia, which is not always viewed as an illness, has potential implications for medical professionals and educators.

The final summary in this literature review of keyword studies of adolescent healthcare interaction is a study by Harvey and Brown (2012), who investigated adolescents' experiences of self-harm in a corpus of messages taken from the Teenage Health Freak website between 2004-2008. The study follows the same steps as Harvey et al. (2008) outlined above: a keyword extraction was performed using the BNC as comparative corpus, which was followed by semantic categories being assigned to the keywords. The words specific to self-harm were identified and their collocations explored in the corpus. Harvey and Brown (2012) find that their analysis reveals several patterns and commonalities in adolescents' accounts of self-harm, information which provide important insights for health professionals into what drives rising numbers of young people to self-harm.

I now conclude the summaries, in this literature review, of studies that have used a keyword analysis in their investigations of healthcare communication. All the studies I have presented thus far on the topic have promoted the keyword technique as a very effective tool in the language researchers kit bag, as evidence that it is:

> an effective means of identifying the 'incremental effect' (Baker, 2006, p. 13) of patterns across large quantities of text, allowing the researcher and the language learner to discover linguistic routines which are liable to remain submerged in extensive data sets. (Harvey & Brown, 2012, p. 333).

I will now move on from keyword studies to present a study that utilises two different corpus techniques in the comparative analysis of four pharmaceutical registers. Grabowski (2015) is included because it uses a keyword and a lexical bundle analysis and because pharmacy is sufficiently related to medicine to be deemed relevant. While there are a handful of lexical bundle studies that relate to medicine (e.g. Jalali & Moini, 2014; Jalali, Moini & Arani, 2015; Samar, Shokrpour & Nasiri, 2018) they are all related to research papers and thus solely from the area of academic writing. There have been no published studies of lexical bundles in any other written medical register.

2.4.5 Grabowski (2015)

Summary

Grabowski (2015) is an analysis of keywords and lexical bundles from the pharmaceutical field. A corpus-driven, descriptive study, it used the perspective of register outlined by Biber (2006), Biber and Conrad (2001; 2009) and others (e.g. Halliday and Hasan, 1976) to investigate linguistic variation across different registers. The hypothesis explored in the paper is that language can vary considerably within a single discourse - in this case pharmaceutical - and this variation relates to the situational contexts, function and intended users of the texts in question. The great variation in the types of written discourse in pharmacy (i.e. registers) had previously been unexplored.

The registers under analysis in this paper were patient information leaflets (PILs), summaries of product characteristics (SPCs), clinical trial protocols (CTPs) and chapters from academic textbooks on pharmacology (ATs). The PILs were accessed from the pre-built PIL Corpus (Bouayad-Agha, 2006), while the other texts were accessed variously form other pre-built corpora and registers or collected by Grabowski. The Corpus of English Pharmaceutical Texts (CEPT) with approximately 2.2 million words was the result. It was this corpus that was also used as the reference corpus, with the register under investigation removed from it for the analysis.

Grabowski’s first focus for analysis was keywords, which was followed by a second analysis of 4-word lexical bundles. The additional focus on phrase frames from the 2013 study was dropped in the 2015 paper and appeared as a separate study in the same year. The results of the analysis showed great variation in the number of keywords present across the registers, not only in the type of word considered key but also with the number extracted: academic textbooks having almost 3 times as many keywords as the clinical trial protocols. The results of the lexical bundle analysis showed similar variation, finding that while all registers made considerable use of lexical bundles, CTPs were the most repetitive and formulaic and ATs the least. Following the taxonomy described in Biber et al. (2004) and Biber (2006), the lexical bundles were categorised into three types: Referential, Discourse-Organising and Stance. Referential lexical bundles, while evident across all four text types, were used least frequently in PILs. On the other hand, stance bundles - more typically found in spoken discourse - dominated PILs while being entirely absent from the 50 most frequent bundles in CTPs.

Commentary

Grabowski's (2015) investigation of four pharmaceutical registers was an important contribution to literature on register analysis, notwithstanding some methodological weakness. It is the only study to date to compare some of the different registers within the domain of pharmacy, drawing the link between the use of keywords and of lexical bundles in specialised discourse with their communicative purpose. This was also the first study to undertake a corpus-driven analysis of keywords and lexical bundles in patient information leaflets. That the lexical characteristics of patient information, beyond considerations of complexity, might contribute to its function (and to its readability and effectiveness) has rarely been the subject of much investigation. A similar situation exists with procedural medical information such as the patient information for radiography analysed in this study.

A key finding is that stance bundles predominate in PILs. There were examples of epistemic stance in the use of words like *sure, probable, may,* as well as deontic stance, relating to obligation and instruction, and desire bundles. This result is surprising, as stance has been found to be more a feature of spoken discourse than written (Biber, 2006, pp. 157–160); and was also seen to be frequent in patient-provider interaction (i.e. *may*) (e.g. Adolphs et al., 2004). Grabowski concludes that the use of stance bundles, along with the use of advisory keywords, is related to the function of the leaflets, namely to instruct and advise patients regarding their medicine.

Grabowski (2015) chose not to use a table to present the categories of the keywords extracted in the different registers. A table would have improved the readability of the study, I feel. Keyword findings are reported in writing only which makes for a lot of dense text. With regard to the lexical bundle analysis, there is also a lack of precision in deciding what can be categorised as 4-word lexical bundles. Grabowski (2015) makes reference to a phrase *If-clause + ask your doctor or pharmacist* when referring to the PIL corpus. This is clearly not a 4-word lexical bundle but something else entirely. There are a number of examples, too, of phrases that contain primarily content words and/or are complete, e.g. *at the desired level* and *special precautions for storage*. Lexical bundles are usually transparent in meaning, they tend to be structurally incomplete, and often bridge two structural units, i.e. a clause or phrase, (Biber and Barbieri, 2007, p8.) Some of the 4-word-bundles in this study are not, in my view, 4-word lexical bundles Nonetheless, Grabowski (2015) is a relevant study that provides further evidence of the existence of register variation within domains of language (e.g. pharmacy, law and medicine) and provides further evidence of the relationship between the communicative purpose of the text and its situational function and its keywords and lexical bundles.

Given the primary importance of healthcare in society and the drive to improve the patient experience, further studies of the lexical characteristics of these registers are long overdue. The register analysis studies described in this thesis provide insights into the register of procedural patient information, and differences worthy of investigation are likely to exist in other sub-registers of medical patient information such as information produced for sufferers of chronic or terminal conditions and that produced by charities or non-profit associations.

I now turn to a brief presentation of a book-length work that investigated modal verbs in medical writing.

2.4.6 Vihla (1999)

Summary

This book length work, which came out of a doctoral thesis, was an analysis of epistemic and deontic modal verb use in a number of medical registers. Vihla, who trained as doctor, compiled a 400,000-word corpus of American medical writing, which included handbooks and clinical textbooks, research papers, scientific textbooks, editorials and consumer health articles. These registers - Vihla refers to them as genres - are categorised into three types: Directive, Argumentative and Expository, according to their communicative functions. The different communicative functions predict, says Vihla, a difference in distribution and frequency of modal verbs. This position is in line with applied linguists (e.g. Biber) who take a register analysis perspective to the study of language. Her analysis finds that the highest frequency of both epistemic (especially *may* and *might*) and deontic modal auxiliary verbs (*must* and *should*) are found in directive texts, which include clinical textbooks and manuals and what are referred to as guidebooks, the closest thing to patient information in her corpus. These findings suggested that the pattern of use of modal verbs and their frequency in medical writing does not follow that seen in general discourse, where *will*, *can*, *could* and *would* are reportedly most frequent. (e.g. Biber et al., 2002)

Commentary

Vihla (1999) is an unusual book for two reasons: its focus is a range of medical registers, rather than a single register (usually the research paper, as we have seen), and unusual because it used corpus techniques to investigate a variety of modal verbs, including deontic modals, when it is often epistemic verbs that get most of the attention. Only American writing was included, with the justification that American medical books and journals had (and continue to have) a wide, international distribution. The texts were complete texts of varying lengths though no mention is made of how different text length may affect frequency results.

Vihla did not include any semi-modals, however, and her deontic modals are restricted to *should* and *must*. Notably absent from her corpus, too, were patient information leaflets. While there were certainly leaflets in 1999, they were certainly less common than they are today and, perhaps, their utility had yet to be appreciated. The finding that *may* is the most common modal verb overall is interesting, as is the finding that *should* and *must* are most common in professional directive discourse and also fairly common in the popular guidebooks.

The limitations to the study are that only two core modals were investigated. Investigating a range of deontic modal verbs, including semi-modals, in patient information would answer the question of whether more informal semi-modals, such as *have to* or *need to,* are also frequent in this kind of medical writing.

It is to the topic of patient information specifically, and to the third and final section of this literature review that we now turn.

2.5 Linguistic approaches to patient information

The following two studies call for a different approach to healthcare materials, both in their development and their assessment. These studies highlight the unsuitability of traditional readability measures and make the case for a new approach that incorporates knowledge from the field of Systemic Functional Linguistics, in the case of 2.5.1, Clerehan, Buchbinder and Moodie (2005), and in the case of 2.5.2, Zarcadoolas (2011), knowledge from pragmatics and text linguistics.

2.5.1 Clerehan, Buchbinder and Moody (2005)

Summary

Clerehan and colleagues developed and applied a linguistic framework to assess the quality of medical information leaflets, in this case for the drug methotrexate used to treat rheumatoid arthritis. This was the first time such an analysis had been carried out. The linguistic framework is based on Systemic Functional Linguistics (SFL) (Halliday, 1994), a theory that considers language to be ‘a pattern of interlocking systems, from the smallest unit (e.g. words or phrases) up to the largest (e.g. a paragraph or a longer piece of text)’ (Clerehan et al. 2005, citing Halliday, 1994). In SFL, meaning is constructed by the reader via the interaction of text (in written discourse) and its context, which is key.

Clerehan et al. (2005) take the view that typical readability assessments, which use statistical averages of sentence and word length, are an inappropriate tool for evaluating patient information materials and do not consider the very many skills needed by a reader of healthcare materials, or the role the organisation of information has in comprehension. An assessment framework containing nine communicative elements was developed, based on the theories of SLF and the concepts of health literacy referred to above. These elements were then used to investigate 18 leaflets of varying length, ranging from half-a-page through to a six-page folded brochure. Words/token data was not given.

Clerehan et al. (2005) found that while all the drug information leaflets possessed a generic structure, only two of the moves outlined in the framework were obligatory (dosage and side-effects). In Clerehan and Buchbinder (2006), which is a different write-up of the same study, this claim is hedged by the addition of ‘possibly’. Inclusion of the other moves appears optional, while all of the other items evaluated showed variability across the texts. Evaluating patient information with a linguistic tool is, say Clerehan et al. (2005), is a far more successful tool towards consistency of the register than commonly-used statistical analyses (i.e. readability measures). The authors proposed further work to evaluate the utility of the framework for writers of patient communication (Clerehan, Hirsh & Buchbinder, 2009) and to develop and test a protocol based upon the framework for reader-focussed evaluation. (Hirsh, Clerehan, Staples, Osborne & Buchbinder, 2009)

Commentary

Clerehan et al. (2005) and Clerehan and Buchbinder (2006), with a longer, more considered conclusion and more methodological detail, proposed for the first time a linguistic and communicative approach to the development and evaluation of patient information materials.

While Clerehan and colleagues took the view that all bar one of their elements were linguistic in nature, it may be more accurate to say that all elements were communicative in nature, in that they contributed to the communicative success of the leaflet. Five elements related to the physical organisation of the text, to its appearance, including visual content and layout, and validity (e.g. accuracy) of the information. The remaining four elements related to the language used by the writer and were thus linguistic: the technicality of the vocabulary; the role relationships in the text, demonstrated through either the use of pronouns or nouns for the patients and medical professionals (i.e. *the patient* or *you*); meta discourse, i.e. the language about the text, and rhetorical elements, i.e. the linguistic function of each move.

In the 2006 study, Clerehan and Buchbinder give a little more methodological detail and organised the elements above into what they refer to as levels of analysis: The first is discourse semantics, under which we find the technicality of lexis (also described as specialization), role relations, and organisational aspects (including visuals and headings); the second level of analysis is lexicogrammar, under which mood and theme is grouped. The theme, say the authors, is generally found in the initial part of a clause and contains known or familiar information. Mood, as described by Clerehan and Buchbinder (2006) is what is usually referred to as stance in the literature.

Some of the findings are of direct interest to this study. In both the 2005 and the 2006 studies, the authors go beyond the concept of technicality of vocabulary to consider other sources of lexical confusion, such as parsing sentences where terms are presented as synonyms with the addition of the word *or*, e.g. ‘Methotrexate may cause a reduction in the number of white cells or platelets in the blood’. This can be doubly-confusing for patients if they are unfamiliar with either or both terms as they do not know if the second word is a synonym or a new, additional word.

Another area of relevance in Clerehan and Buchbinder (2006) is the reference to the blurring of ‘shouldness’, (the authors cite Iedema (2007)), where the language of obligation, probability and suggestion are mixed, e.g. *taking more than the prescribed dose could be dangerous*’, increasing, the authors say, the likelihood of patient confusion. The analysis conducted for this doctoral thesis on modal verbs, reported in Chapter 6, will reveal whether the blurring of ‘shouldness’ is also an issue in procedural information or whether obligations and instructions are presented directly.

We have seen already in this literature review the problems associated with frequency markers and patients’ over-estimation of risk. The use of frequency markers, very common in drug information, also receives attention from Clerehan and colleagues. *Seldom, rare, sometimes* and *usually* are often used in patient information without quantifying more precisely, say the authors. We have already seen, however, that neither quantifiers or a combined approach to the presentation of risk information is without problems. Clearly this is an area that requires more investigation.

In the Clinical Contact move, Clerehan et al. (2005) find a variety of linguistic structures, including imperatives. These structures, they conclude, are either offering a service (*please consult your doctor*) or instructing (*inform your doctor*) (p339). The addition of *please* does not always make an imperative an offer, however. This observation was acknowledged in the revised paper, Clerehan and Buchbinder (2006). Here, the example of an offer, *please see your doctor*, from the 2005 study (which is an imperative and not an offer), is replaced by *please do not hesitate to contact me*, which clearly invites (e.g. offers) contact.

Neither study summarised above was a corpus study - just 18 documents were selected from 91 about rheumatism (as we learn in the 2006 study, though this information is absent in the 2005 paper): 15 produced by individuals/hospitals; one from the Arthritis Association (of Australia), one from a medical journal and one from a pharmaceutical company. All documents relate to the same treatment drug but in some documents the intended recipient is not a patient but more likely a doctor. Clerehan and Buchbinder (2006) suggest that this may mean that the role relationships can be unclear and status relations can vary within a document, leading to patient confusion. My view is that it would have been better to exclude these documents from the analysis as it is quite possible that many more of the elements analysed will differ when the recipient is a professional. not just role relationships and status relations. As the documents were all different lengths and produced by different people (in many cases unknown), controlling for recipient and topic would provide more consistent information.

The lack of methodological detail in the 2005 study was, to a certain degree, rectified in the 2006 study, though we still do not know how the ‘technicality’ of vocabulary was decided, or whether verbs were chosen to mark rhetoric function and if so, how were they decided upon and counted. The size of the study also limits its applicability but, nevertheless, the linguistic focus taken in both papers is ground-breaking, going far beyond the usual reference to everyday terms or simple language that many studies of the utility and comprehensibility of patient information leaflets refer to. A subsequent study (Clerehan, Hirsh and Buchbinder, 2009) applied the framework to a new set of leaflets, showing the usefulness of the framework.

What is surprising, however, is that the impact of these studies have not been felt more strongly: more than a decade on, standard readability indexes, e.g. Flesch-Kincaid and Flesch Reading Ease, continue to be used in studies that evaluate the comprehensibility and utility of patient information (e.g. Lampert et al., 2016; Paz et al., 2017) and there are still very few studies that consider or investigate the linguistic and communicative features of patient information. One important study that does consider the linguistic features of patient information is Zarcadoolas (2011), who discusses the usefulness of pragmatics and text theory in healthcare message simplification. It is the final study presented in this literature review.

2.5.2 Zarcadoolas (2011)

Summary

The problem of the low health literacy of millions of Americans - which is referred to as the ‘silent killer’ - is the motivation for the call presented in this paper for a new approach to the production of health materials for the 21st century. Zarcadoolas (2011) questions the role and efficacy of text simplification, particularly as healthcare materials become increasingly complex, and our understanding grows of the multifaceted nature of health literacy. A discussion of the utility of applying two models from sociolinguistics - pragmatics and text theory - to develop a ‘richer, more theory-based understanding of text structures and function’ (p. 338) is presented.

Simplified text and readability measures - which we have already seen are considered by some researchers to be poor tools in the assessment of healthcare materials - seem an inadequate response to the complexity of healthcare information and the multifaceted, complex nature of health literacy, with its social, cultural and environmental links. Simplified text can affect coherence, and yet we know, too, that cohesion and content are vital in order to decode meaning. When text is overly simplified, information can be missing, and the natural inference that takes places when we read is affected, says Zarcadoolas (2011), adding that most meaning takes place in ‘beyond-the-sentence chunks in the text and the interaction between the text and the reader/listener (pragmatics)’ (p. 342). Unwittingly, says Zarcadoolas (2011), simplified messaging may be making it more difficult for individuals to get the information they need to make informed healthcare decisions.

A ‘health literacy load analysis’ (Zarcadoolas and Pleasant, 2008) is presented as a possible solution to the problems described above. The aim of the analysis is to assess the difficulty of a text by using linguistic text models in conjunction with an ecological model of health literacy. The analysis ‘unpacks’ a text and requires an identification of the aspects that are likely to affect comprehension at the surface-level, text level and at the pragmatic level. The author concludes with a partial load analysis example and an acknowledgement of the need for further research into their proposed health literacy load analysis.

Commentary

Zarcadoolas (2011) is one of very few studies to propose that knowledge from the field of applied linguistics - specifically pragmatics and text theory - be applied to the evaluation of healthcare materials. Another researcher who also calls for applied linguistic involvement in this area is Rubin (2014), who states that ‘the symbiosis between health literacy and applied linguistics is strong’. Both are pragmatic disciplines, Rubin says, adding that perhaps the greatest contribution that applied linguistics can make to health literacy is to develop ‘modality and context-sensitive parameters for characterizing and then mitigating health message complexity’. (p. 161).

Zarcadoolas and Pleasant’s (2008) health literacy load analysis may be useful in this regard. Along with the knowledge of how we process meaning, and the role of cohesion and content in making meaning, Zarcadoolas and Pleasant (2008) present other literacies they say are required to comprehend many healthcare messages: basic functional literacy; science literacy, which relates to basic scientific and technical knowledge; civic literacy, which relate to media literacy, the capacity to assess the source of the information, knowledge of governmental and civic systems, and awareness of personal responsibility; and finally, cultural literacy.

The framework developed by Clerehan et al. (2005) seems more helpful for the developers of healthcare leaflets perhaps, as it focusses clearly on the communicative and linguistic aspects of the leaflets, permitting a step-by-step approach to both development and evaluation. Zarcadoolas (2011) refers to the need to ‘unpack’ the text at the surface-level, but there is no detail given, and only a partial example of a health literacy load analysis is provided. The surface-level of the text I take to mean the structure of the text and the lexis used, though ‘unpacking’ is a vague term if the writer of the text fails to see how the words they choose, or the way in which they express certain ideas, can be problematic for the reader. Zarcadoolas (2011) seems to relate more to the application of applied linguistics knowledge to the definition of health literacy, than it does to the production and evaluation of healthcare materials.

Nonetheless, Clerehan et al. (2005) and Zarcadoolas (2011) are significant contributions to the very small, but steadily growing, body of literature that sees the value of applied linguistic approaches to the study, development and appraisal of healthcare materials. My own research, I hope, will also be considered a useful contribution to the literature.

2.6 Conclusion

In this chapter I have presented a literature review of key studies relating to a variety of academic medical registers, including research papers and case notes. I have included, too, an overview of the literature relating to patient information leaflets for pharmaceutical products, commonly known as PILs, and consent forms, suggesting that the principal themes common to these studies may also be common to procedural patient information. These themes include readability and how risk is expressed and understood.

The second part of my literature review began with a presentation of two studies that highlight the challenges and the importance of real-world outcomes of language research into the discourse of healthcare and healthcare materials and concludes with two studies that underline the need to approach the development and evaluation of these materials in a new way. These studies highlight the importance of text cohesion in the comprehension of the materials, underlining the difference between literacy and health literacy and emphasising the insights that linguistic research can bring to the discussion.

The importance of coherence and cohesion in text comprehension is referred to by Zarcadoolas (2011; 2013) and Clerehan et al. (2005) though it is an area that, to date, has not been sufficiently investigated in relation to healthcare materials. Simplifiied text, which is the standard approach to the production of healthcare materials, may even be at odds with the concept of cohesion as ‘text that is highly cohesive maintains continuity of ideas. If there are few or no connections between ideas/sentences in a given text, readers need to bridge the cohesion gap through inferences (Singer & Ritchot, 1996). As Liu and Rawl (2012) and Zarcadoolas (2011) say, health literacy issues can mean that patients are unable to bridge the cohesion gap through inferences as they have none to draw on. This is likely to be the case when the healthcare information pertains to rare or unusual conditions, or to conditions that are not commonly discussed, but whether a health topic is commonly discussed will vary from one individual to another and is dependant on social, cultural and religious appropriacy. Even relatively common conditions may be entirely unfamiliar to some people. The same applies to novel experiences, such as a radiographic exam, the focus of the material investigated in this thesis.

In Liu and Rawl's study (2012), the researchers conclude that higher text cohesion facilitates the reading speed and comprehension of colorectal screening information, (but not retention of vocabulary), while a study from Finnish researchers (Kaakinen, Salonen, Venäläinen, & Hyönä, 2011) on the relationship between cohesion and expository text - and healthcare information materials often have an expository role - found that high cohesion text was not only more persuasive than low cohesion text, but that attitude after reading predicted successful recall of the message arguments. In healthcare materials, it is not necessarily the vocabulary that needs to be remembered but the message: if readers have understood the reasons why not eating before an operation is important, and this message has been presented persuasively, perhaps they are more likely to remember and act upon it. Further research in this area is very much needed.

The lexico-grammatical patterning of a register is linked to its cohesion, and a lexical bundle analysis reveals this patterning in a register (e.g. Conrad & Biber, 2005; Biber & Barbieri, 2007). There are very few lexical bundle studies of non-academic medical registers and Grabowski (2015), presented in this chapter, is the only such study that I am aware of. The gap in the literature is enough of a motivating factor to conduct such an analysis of procedural patient information. The other motivation, however, is that such an analysis may well reveal valuable information about the level of cohesion of the text, with corresponding insights relating to its predicted ease of comprehension and message retention. The methodology of a lexical bundle analysis and some key studies will be presented in the next chapter, with the results of the analysis presented in chapter 5.

In this chapter I have also presented a number of studies that use the keyword approach to investigate practitioner and patient language. The keyword approach is the most utilised corpus technique in the healthcare discourse literature, principally because it allows a more sensitive uncovering of areas of interest than traditional qualitative methods. Applying a keyword extraction to patient information for radiography may reveal some interesting information about its characteristics, both semantic and thematic, and allow, perhaps, a comparison between this register and what we know of other medical registers, as revealed by existing studies, some of which I have included in this chapter. The papers presented in this chapter are keyword studies of practitioner spoken language (Adolphs et al., 2004 and Skelton and Hobbs, 2009a) and the written and spoken language of the patient. In my study, I apply the approach to official written language of the healthcare provider, procedural patient information. The methodology of the approach is presented in chapter 3 while the results of the analysis are presented in chapter 4.

Vihla (1999) provides the springboard into my analysis of modal verbs for instructions, which I present in chapter 6. Patient information was notable by its absence in her research with a likely explanation being that in the age before universal access to the internet, patient information was neither produced or digested at the rate it is today. My analysis also extends the range of modal verbs under analysis to include the semi-modals *have to* and *need to*, in addition to *should* and *must*, the two modals Vihla investigated in her corpus. The ‘blurring of *shouldness’* referred to by Clerehan et al. (2005), citing Iedema (2007), where the language of obligation, probability and suggestion are mixed, e.g. *taking more than the prescribed dose could be dangerous*’ leading to possible comprehension issues, is also of interest. By extracting the modals and semi-modals used for instruction in the corpus, and examining their uses in the patient information, we can have a clearer idea of how they may help or hinder the comprehension.

We now turn to the detailed methodology of the corpus approaches I have referred to, preceded by a presentation of the corpora the software that was used for the analyses.

3. Corpora and methodology

We have seen in chapter 2 just how useful the keyword extraction is for uncovering hidden discourses, beliefs and attitudes. Applying the keyword technique to patient information for radiography is likely to reveal some interesting findings about the register. A lexical bundle analysis, on the other hand, has not been applied to healthcare materials, though the literature shows it can reveal important information about the communicative function of a register. Such an analysis can provide insights into the structure of patient information for radiography and the structure of a text is an important factor in how complex it is or is perceived to be, as we have seen in the previous chapter. The readability of patient information is an ongoing concern, but the inadequacy of readability measures for the evaluation of patient information is also an issue. A lexical bundles analysis may well contribute to our understanding of the underlying structure of the patient information and from this, we may have a better idea of how its cohesion relates to its readability. The third analysis carried out as part of this doctoral thesis is on the use of modal verbs to express obligations and to give instructions. Given that instructing is one of the primary functions of patient information, it is important that we evaluate the lexical means used to express these instructions. The concepts of adherence and compliance appear differently if patients have not understood that an utterance was intended to instruct, and thus did not realise they were expected to do something. The ‘blurring of shouldness’ that Clerehan et al (2005) refer to may well cause comprehension problems, as referred to in chapter 2, but this idea has not yet been explored in healthcare materials.

This chapter presents the methodologies and techniques used in this study, beginning with a description of the patient information corpus, the material it contains and how this material was selected. This description is followed by a similar description of the two other corpora that I built for the purposes of comparison: a consumer information corpus and a general radiography corpus. After a detailed description of the steps taken to construct the corpora, I present the software that was used: Sketch Engine. This, in turn, is followed by a detailed presentation of the methodology of the keyword and lexical bundle approaches, along with the methodology used for the third analysis of deontic modal verbs, those modals that are used for instructions and obligation.

3.1 The Patient Information Corpus

The patient information that was used in this research was procedural patient information for radiography. As I state in the preceding chapter, we can think of this as a sub-register of patient information. This kind of information is given to people who are attending hospital or medical centre for a diagnostic radiographic examination and given to patients who are undergoing a medical operation or intervention that involves the use of a radiographic technology, such as angiography or bronchoscopy. Patient information is also written for patients who are undergoing radiation therapy for cancer.

To put the topic of radiography into some context, it is helpful to know which exams are commonly carried out, how frequently they are performed and which use radiation. Table 1 presents a snapshot of imaging activity in the NHS over a 13-month period. The radiographic examinations most frequently performed by the NHS are x-ray, computed tomography (CT), magnetic resonance imaging (MRI) and ultrasound. The last, ultrasound, is also referred to as sonography or ultrasonography. CT and x-ray use ionising radiation. Table 2 gives a snapshot of the numbers and type of radiographic exams performed between January 2016 and January 2017 in the UK by the NHS.

Table 2 Count of imaging activity in England, on NHS Patients, January 2016 to January 2017

Modality	Number in 2016	Radiation used?
X-ray	22,398,045	Yes
Ultrasound	9,099,225	No
CT scan	4,655,065	Yes
MRI	3,234,690	No
Fluoroscopy	1,033,250	Yes

Nuclear Medicine	418,220	Yes
PET-CT scans	125,640	Yes
SPECT scans	31,015	Yes

Note: Data from https://www.england.nhs.uk/statistics/wp-content/uploads/sites/2/2016/08/Provisional-Monthly-Diagnostic-Imaging-Dataset-Statistics-2017-05-18.pdf

According to government figures (Baker, 2018), the number of CT scans carried out by the NHS has increased by 43% in the period 2013-2018, while MRI scans are up by 42%. Increases such as these have been reported in most countries in the world (IAEA, 2015). In spite of the frequency reported above, we have seen already that patients know very little about the examinations being performed in radiography suites (Singh et al., 2017; Ukkola et al., 2017) and they are even less knowledgeable about radiation risk (Hansberry et al., 2014; Singh et al., 2017). These findings, along with the increase in the number of scans being performed annually, strongly suggests that comprehensible patient information for radiography has never been more necessary.

3.1.1 Corpus contents

The corpus used in the analysis was made up of 221 written patient information documents for radiographic examinations and medical interventions that involved a radiographic procedure, such as angiography or bronchoscopy. The corpus was first compiled in 2011, with later additions in 2014 and 2016. Both diagnostic and therapeutic procedures were included, though the majority of documents relate to diagnostic exams and medical procedures involving the use of radiographic technologies. The patient information documents were all Word or pdf documents that were available online. The sources of the information were NHS hospitals in the UK (54 documents), the British Society of Interventional Radiology (37 documents) and a US radiology patient information website, www.radiologyinfo.com (130 documents), produced by the Radiologic Society of North America (RSNA). The RSNA is a non-profit association.

I chose to include patient information from both the UK and the US as the internet means that patients are not restricted to information produced in their own countries and may freely read various documents from a range of countries. Radiology Info is a very well-known and well-respected information website with, it is reported, around 700,000 visitors a month. (cited in Hansberry, John, A., John, E., Agarwal, Gonzales & Baker, 2014). A search in Google for 'CT information' returns Radiology Info as the first result. A search in Google for CT information UK' returns the NHS. With regards to the latter, it was particularly NHS hospital websites that were the source of the information materials, as it was important that only websites that offered printable, stand-alone documents were included, as I will now explain.Websites that offered information but not as a .pdf or a Word document were not included. Nor were forums or chat rooms. The focus of this thesis is on official patient information produced by hospitals and medical trusts, or developed by medical associations with professional authority, producing information for both patients and professionals. Peer-to-peer online communication, where patients give information and advice to each other, or general medical websites reporting on radiographic procedures, were both excluded. As the documents were available as pdf or Word documents, I assumed that the patient was expected to print off the information (although whether patients did or did not print off information was irrelevant to the study.)

The corpus contained 408,997 words and a total of 221 documents. There were 37 documents produced by the British Society of Interventional Radiology; these were published between 2010 and 2011. The 54 documents from the NHS hospitals were published or updated between 2007 and 2015 while the remaining 130 documents from RadiologyInfo.org were published between 2013 and 2014.

Table 3 shows the make-up of the corpus.

Table 3 The contents of the corpus of patient information for radiography

Source	N word	N docs
NHS (UK)	62,957	54
British Society of Interventional Radiology (UK)	51, 654	37
RadiologyInfo.org (US)	294,386	130
	408,997	**221**

3.1.2 Corpus size

Sinclair (2004), when referring to corpus size, asserts that ‘small is not beautiful; it is simply a limitation’ (p. 89), but while the corpus used in this study is small, it is also specialised. A specialised corpus has been described as a collection of texts delimited by a particular register, discourse domain, or subject matter (De Beaugrande, 2011). As a specialised corpus, its size is appropriate, and I included only patient information that met certain criteria, as described above in 3.1.1. Nor are small, specialised corpus studies that investigate medical language considered unusual. Indeed, it would be difficult to find a researcher working in the field of specialised registers who would argue for the need to use a corpus greater than 1 million + words. The handful of people who suggested such a thing in the course of my research were not people working with corpora on a regular basis or even at all.

Specialised corpus studies in the domain of medical language that use small datasets include Yang et al (2015), who look at epistemic modality in medical research papers in a corpus of around 80,000 (exact count not given), and Webber (2005), who considers interactive features in a 34,692-token corpus from medical conference presentations. Adolphs et al (2014) investigated a 61,981-word sub-corpus of the Nottingham Health Communication Corpus. The sub-corpus were transcribed telephone conversations between NHS Direct health advisors (NHS Direct was a telephone health advisory service run by the NHS, no-longer active) and patient callers. As Adolphs et al. (2004, p. 13) say: ‘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’.

Skelton and colleagues, who were among the first to utilise a combined quantitative and qualitative approach to medical discourse (e.g. Skelton and Hobbs, 1999a¸ 1999b; Skelton et al., 1999; Skelton et al., 2002a; 2002b), also used relatively small datasets - often around 500,000 words (exact totals were not given). While stressing the importance of quantitative data, Skelton and Hobbs (1999) also underline the need for context and qualitative information in interpreting the data, saying that

> A basic concept in the study of language is that meanings cannot be completely quantified: if words were like numbers, it would be hard to understand why we bother with both. Any quantitative analysis must, therefore, take place in a qualitative context. (p. 109)

Their view and that of many researchers working in the field of healthcare discourse is that mere frequency counts alone are unlikely to uncover the kind of patterns that are of interest, making a strong case for smaller corpora and combined approaches to data analysis.

3.1.3 Document type and variation

Documents varied considerably in length, from around 300-4388 words. The information from the US site, RadiologyInfo.org, contained the longest documents: nearly 60% of the documents were longer than 2,000 words, with around 25% longer than 3,000 words. The site was also responsible for a document on urography at 4388 words. In contrast, just 10% of the NHS documents and barely 3% of those from the Society of Interventional Radiography were longer than 2,000 words.

42% of the documents are from UK sources and 58% from a US source in the corpus: in terms of the number of documents, then, the corpus is balanced. However, the lengths of some of the US-sourced materials means that the ratio of words in the corpus is less balanced: 114, 611 words in the UK sourced materials compared to 294, 386 in the US sourced, which is a ratio of around 1:2.5.

It is interesting to note that the complexity of the patient information material on RadiologyInfo.org has been the subject of some criticism, and virtually all of their material scored poorly when subject to a raft of readability tests (Hansberry et al., 2014). I have, however, already referred to the shortcomings of readability measures to evaluate patient information, none of which, to my knowledge, include the length of the original document in their appraisal. Perhaps all that can be said about the materials in RadiologyInfo.org is that they are long, which may well put people off reading them.

The decision to use only UK- and US- sourced material was also taken because of the difficulty I had in sourcing material that fitted my criteria, and that was readily available on websites from other English-speaking countries. I had assumed that patient information is as ubiquitous elsewhere as it is in the UK and the US, or at least digitally ubiquitous, though it turns out not to be the case. The reasons for this vary, though technological advancement (i.e. making information available digitally), budgets (i.e. the cost of producing patient information) and how patient-centred a healthcare system is (i.e the perceived need for patient information) are likely to be primary factors. Due to the difficulties of finding suitable examples that fitted my inclusion criteria (see above), a decision was made to use British and American English examples only. As I have already mentioned, both RadiologyInfo.org and the NHS material is the first to be presented in a Google search, so we can be assured that these sources are likely to be the first that patients from the respective countries look at when online and searching for information about radiography.

3.1.4 Document lengths and sampling

But does the difference in length of the documents in the corpus matter? I believe that the length of the texts matters far less than the fact that all documents were complete and, as I explain below, the length of a document may even be a factor that is related to culture and to concepts of uncertainty avoidance. With regard to corpus building, Adolphs (2006) recommends that texts making up a corpus should be complete, though other researchers suggest sampling (McEnery, Xiao & Tono, 2006), and have stressed the need for documents of a similar size when building general corpora. However, sampling is not something that is necessarily easy or advisable to do, which I will explain in more detail in 3.1.5. Sampling is not only something that would have proved difficult to do, giving the enormous variation in length of the documents in the corpus referred to above, but would not have been a good idea, either, given that the documents were relatively short and all of the information contained in them, and where that information occurred, I considered important.

It may also be the case that the length of a document is a possible indication of cultural differences and thus a factor worth reporting on. As we have seen, patient information for radiography can be considered a sub-register of patient information, but it should be remembered that even within this sub-register there are likely to be differences that future studies could explore. These differences include those that relate to the variety of English, e.g. US and UK and those that compare English with other languages. There have been similar studies for the sub-register of the PIL (Biancho, 2016) with findings that suggest this is a fruitful area for further studies, as how information is presented and what is prioritised may well be culturally specific. How risk is perceived, for example, has been shown to be culturally specific (Gerritsen, Nederstigt & Orlandi, 2006; Van Berkel and Gerritsen, 2012) and information pertaining to side-effects in PILs can be presented differently depending on the language. Van Berkel and Gerritsen (2012) demonstrated this in their study of side-effect information in a drug leaflet produced for Flanders and the Netherlands, the latter a low uncertainty avoidance culture where people fear risk less and do not feel the need to have the details related to risk made clear, the former a high uncertainty avoidance culture. Five leaflets for an ibuprofen medication were compared. The leaflet produced for Flanders was presented in three languages: Dutch, French and German, though it was the Dutch content alone that was compared. Significant differences were found by Van Berkel and Gerritsen (2012) in the amount of risk information included and the number of medical terms used, and while the leaflet from Flanders was, in four cases, considerably longer than the version for the Netherlands, though the authors did not find the difference in length to be statistically significant.

Reducing texts in size by sampling them may remove important information that relates to the readership and production culture. This kind of information is increasingly important as new laws come into effect in Europe relating to the provision of lay clinical trial summaries. These lay summaries must also be translated into the languages of the countries where pharmaceutical products have been evaluated. Van Berkel and Gerritsen (2012) shows us that a one-size-all approach is not appropriate when discussing risk or when using medical terms. In addition to the reasons presented above, sampling patient information may unwittingly remove key sections of the information, as I will now explain.

3.1.5 Sampling

Sampling may not be appropriate when working with certain types of document. The corpus of patient information developed for this doctoral research is not a general corpus: it is specialised, and the register of patient information has an organisational structure that is subject to guidelines and convention (e.g.MHRA, 2012). Patient information for radiography contains similar information and instructions, i.e. advice/directions regarding preparation, descriptions of the procedure, reference to risks and benefits, and to follow-up care, presented in a similar order (e.g. information regarding preparation appears at the beginning of a document, while follow-up care generally appears towards the end of the document). The aim of this study is to describe some of the linguistic devices used to express that information and those that instruct, so it was important to have complete documents in order not to remove sections of information and thus skew the results. Risk, for example, will generally not be discussed at the beginning of a document but towards the end. Offers of more information and advice come at the end of a document, not in the middle. The legal disclaimers that, as we shall see in the following chapter, are a very significant part of the US-sourced information, always appear at the end of the document. A description of the procedure and the benefits are far more likely to come at the beginning of a document.

It is my view, then, that sampling would not have been appropriate for the documents in my corpus, irrespective of the difference in length between them. Nor is sampling something that needs to be done when texts are already brief in length. The idea that sampling is not suitable for certain types of corpora is shared by many researchers (e.g. Flowerdew, 2004; Kennedy, 1998) and it is worth remembering that the individual documents in the patient information corpus are already short or very short in most cases. ‘Frequent linguistic features are quite stable in their distributions and hence short text chunks (e.g. 2,000 running words) are usually sufficient. (Xiao, n.d.) If Xiao considers 2,000 words to be a short text chunk, the vast majority of the documents in my corpus can be considered short or very short.

3.2 Reference or comparison corpora

When conducting a keyword analysis, a reference corpus, also known as a comparative corpus, is used to calculate the ‘keyness’ or statistical significance of words in the target corpus. The default reference corpus in Sketch Engine is English Web 2013 (EnTenTen 13), a corpus of internet texts running at around 19 billion words. For my analyses, however, I chose to use the British National Corpus (BNC), a 100-million-word collection of British English, 90% from written discourse. The BNC comes pre-loaded in Sketch Engine. Not only is the BNC a very common choice of reference corpus but the preponderance of written discourse in the corpus made it a suitable reference corpus for this study. It is also the case that my corpus was made up of Word and pdf documents, often the same documents that are available in hospital radiography departments and GP surgeries. The materials downloaded from RadiologyInfo.org were also produced to be both read online and downloaded. The size, too, was a factor. EnTenTen 13 is a web-based corpus and is enormous, at 19 billion words. While 100 million words (the BNC) is still considerably more than my corpus of round 400,000 words, the comparison in size is more appropriate.

As we have already seen in the literature review, however, comparing a specialised corpus against a general corpus like the BNC, is not the only approach likely to yield interesting results. In fact, comparing a specialised corpus against another specialised corpus may well reveal discourses that remain hidden with the former approach. Because of this advantage, I also built two reference corpora to be used for more targeted analysis, both in the keyword analysis and in the modal verb analysis. These two corpora I will now present.

3.2.1 Comparison corpus 1: Consumer Information¨

This was a small, 104,670-word corpus of consumer information, with material from both the UK and the US. The inclusion criteria for the consumer information corpus was very close to that for the patient information corpus. All texts were available as Word or pdf documents on the Citizens Advice website, a recognised authority in the UK for consumer information, and its US equivalent, the Federal Trade Commission. This was a much smaller corpus than the patient information, however, with a word count of just 104,670 and the majority of texts came from the UK Citizens Advice site. The topics covered included housing, health, children, consumer topics and the law. The length of the documents included in this corpus also varied, from the longest at over 7,000 words to the shortest at under 300 words. The longest documents in this corpus were from the UK.

The rationale behind building a reference corpus of consumer information lay in the fact that patients are increasingly referred to as consumers, or as clients or service users. But are patients treated as consumers in written informational materials? Are they spoken to in the same way? I wanted to investigate any lexical similarities or differences between medical patient information and general consumer information, particularly with regards to the use of deontic modal verbs, those verbs we use for instruction and obligation. It was my hypothesis that irrespective of the fact that patients are sometimes referred to as consumers, they are not treated as such in procedural patient information. The way that healthcare materials instruct patients is quite different to general consumer information, which is far more likely to direct and instruct readers. The topic will be explored in more detail in Chapter 6.

3.2.2 Comparison corpus 2: General radiography

This was the first reference corpus that I built, and at 719,209 words it is considerably larger than the corpus of Consumer Advice. The corpus of General Radiography is made up of a radiographer handbook, *Clark's Positioning in Radiography*, a textbook, *Patient Care for Radiography* and research from *Radiography*, a peer-reviewed journal of the Society and College of Radiographers and the European Federation of Radiographer Societies. The research was included because it is written for and by radiographers, and not radiologists, and thus deals with the issues that are relevant for radiographers: patient safety, radiation dose, patient position and workflow, for example. Radiologists are doctors who specialise in radiology. Their job is to diagnose, primarily. Radiographers, on the other hand, are the healthcare professionals who carry out radiographic examinations.

The reason for using this corpus of radiography when extracting keywords has already been touched upon in Chapter 2. A specialised reference corpus is often the best choice when working with a specialised corpus, otherwise the keywords are likely to be solely the technical terms in the corpus, which may not be of much interest to the researcher. Patient information for radiography and radiotherapy contains many references to radiographic modalities (types of exam), (e.g. MRI, CT), radiographic procedures (e.g brachytherapy), and to medical or technical terms, even if they are then glossed for the patient information. I was not interested in these words for the purposes of my research, but more interested in what might be considered general language. Using the BNC as a reference corpus, however, naturally results in many of medical and radiological terms appearing as keywords. By conducting a second keyword extraction using a radiography reference corpus, and thus reducing the chance of medical and technical words being extracted as key, I felt that the analysis would be more fine-tuned and perhaps reveal more about the underlying characteristics and concerns of patient information.

I now turn to a description of the software programme used to build and analyse my corpora, describing its functionalities and tools before moving on to present the specific steps take for each of the three methodologies.

3.3 Software: Sketch Engine

Sketch Engine (Kilgarriff et al., 2004) is a corpus manager and text analysis software programme, developed by a company called Lexical Computing. It is designed primarily for lexicographers, translators and researchers, and along with the corpus building function it offers, it also contains a large number of pre-loaded multilingual corpora and a range of lexical analysis functions for use with either self-built or the pre-loaded corpora. These functions include the proprietary Word Sketch, which is a summary of a word’s grammatical and collocational behaviour; along with a concordance search; collocation search; word frequency lists; keyword and terminology extraction; diachronic analysis and n-gram extraction. N-gram is synonymous with lexical bundle, where n stands for any number.

For this research thesis, I used the concordance search to examine the lexis in context; I used, too, the keyword function to generate lists of keywords and also the n-gram function, which is a synonymous term for lexical bundle. I used Word Sketch to look at an item’s common lexical and grammatical collocations, particularly when I was investigating the deontic modal verbs and keywords. The first step, however, was to build the corpora. The steps will be illustrated below and were the same for all three corpora.

3.3.1 Building the corpus

Sketch Engine (www.sketchengine.eu) Kilgarriff et al., 2004) was used to compile the corpus and to perform the analyses described in the following chapters. The illustrations below relate to the original interface which I had used during my research. A new interface was introduced in 2018.

One of the advantages of Sketch Engine is that it does not require text to be converted to raw text files, which is standard to most or all other similar concordance programmes. In fact, Sketch Engine accepts a range of document types, including Word and pdf files, meaning no time-consuming conversion was needed. To my knowledge, the programme is unique among concordance programmes in its ability to process a wide range of file types. Sketch Engine supports corpora in many languages and also offers many pre-loaded corpora, in multiple languages, which can be used for analysis or for comparative purposes. I did not use any of the pre-loaded corpora as none of them was suitable for my purposes, however. While there is a dedicated medical English corpus, it is made up of data found on the World Wide Web and is enormous in size, at 33 million words, rendering it unsuitable for my aims on the basis of size and its contents, As I have stated, this study is not concerned with peer-to-peer communication, web-based chat programmes or online interaction, but, rather, available-for-printing healthcare materials, and thus the content of this medical English corpus on Sketch Engine - not to speak of its size - was unsuitable.

3.3.2 Adding text to the corpus

After selecting Create corpus on the menu, documents were uploaded as zip files for multiple texts, or individually. I had previously downloaded documents as Word or pdf files and placed many of them into zip files. It is also possible to download directly from an online location or to paste text, as you can see in Figure 1.

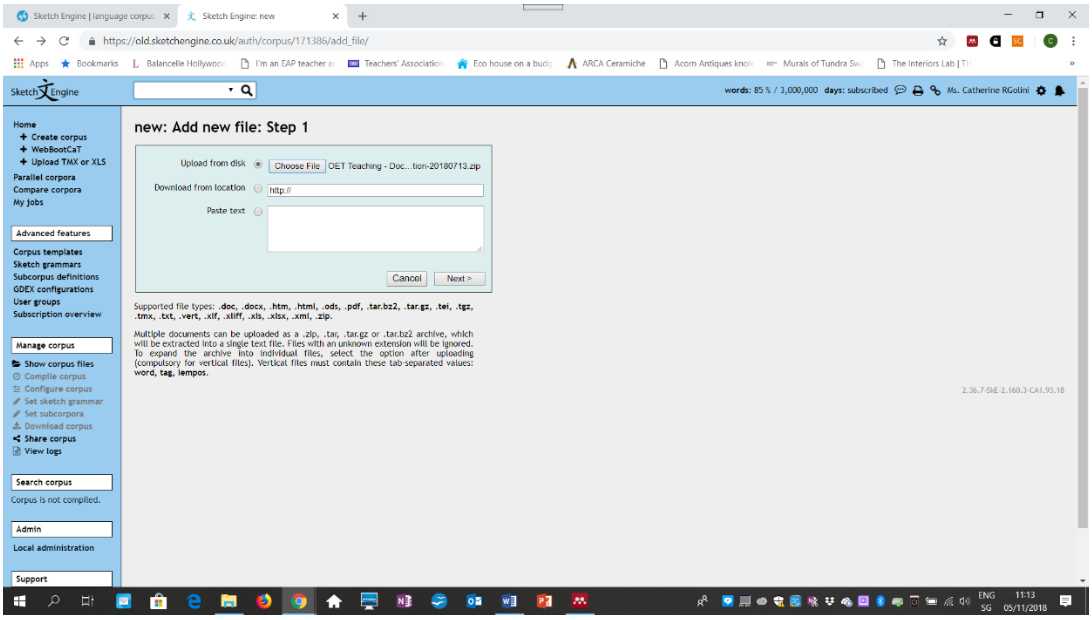


Figure 1 Adding text to the corpus

3.3.3 Compiling the corpus

The next step, illustrated in Figure 2 below, was to compile the data, processing it so that the various functions of the programme described above can be used. I accepted all default settings in Sketch Engine: for the Sketch grammar English 3.3 for TreeTagger pipeline v2 and for the term extraction, English (TreeTagger-PennTB) for terms extraction 2.3. TreeTagger refers to the part-of-speech tagging that the text files are subject to. Part-of-speech tags relate to the grammatical category of a token (i.e. verb, singular noun) and also, in some instances, to case and tense.

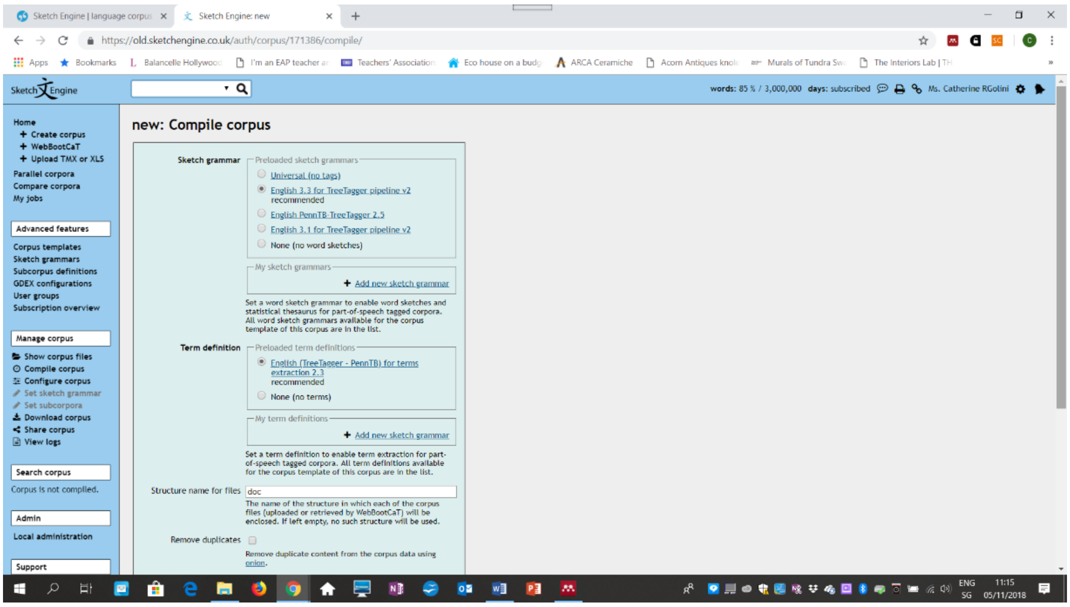


Figure 2 Compiling corpus

At this point, the user can select to make use of the 'Onion' programme, which removes all duplicate content at whatever level the user sets (i.e. the sentence, paragraph or document level). There is a lot of similarity in patient information for radiography; templates are used and the same phrases are repeated, removing duplicate content would have reduced my corpus considerably. However, the standard (i.e. repeated) phrases used in a register are part of that register's characteristics. One of the aims of this doctoral research was to reveal some of the lexical characteristics of patient information for radiography, included the oft-used sentences. Clearly, for my purposes, removing duplicate content was not an option that I wanted, so it remained unchecked.

Compiling a corpus is something that must be done whenever new content is added to the corpus. If the uploaded text is not compiled, it cannot be searched. Documents can be added to the corpus at any time after the initial compilation. Once the corpus has been compiled, it can be searched and the functionalities referred to previously, e.g. term extraction, n-gram extraction, can be applied.

3.3.4 Searching the corpus

Sketch Engine as I have said in 3.3, offers the user a range of options for corpus searching and analysis. Word Sketch, which gives the user the grammatical and lexical collocation information of a selected word, is proprietary. It is also possible to compare collocational information for two different words. As an example of the kind of lexical information presented in Word Sketch, a screenshot for the word *pain* can be seen in Figure 3 below.

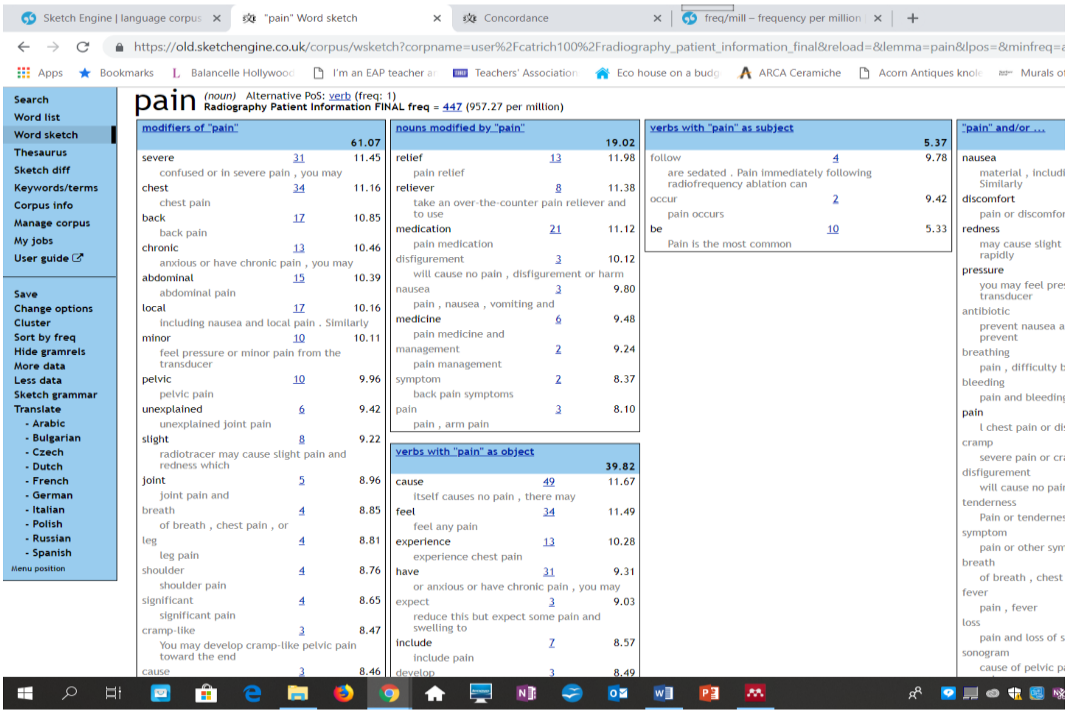


Figure 3 Word Sketch showing results for 'pain'

To the right of the word *pain*, at the top of the page, we can see the most common part of speech for the selected word (in the case of *pain* it is a noun) and any other part of speech, if it was found in the corpus. There is just one example for *pain*, of a verb. On examining this data in context (by clicking on the number 1), we see that it is not, in fact, a verb, but a noun, underlining the importance of context and the fallibility of computer programmes.

You may feel pressure or even pain when the needle is advanced into the joint.

The collocation and grammatical information are arranged according to frequency, in descending order. The count appears both as raw count (a simple frequency count of how many times the search term appeared in the corpus) and as a normalised count, which will be explained in the next section. The raw count is hyperlinked and when clicked on, takes you to the data in context. In the following section I will present raw and normalised frequency counts.

3.4 Issues of frequency reporting and distribution

3.4.1 Raw frequency and normalised frequency

When conducting my modal verb analysis, frequency counts were normalised, that is, expressed as per million words (pmw). Normalising frequencies allows comparisons to be made between differently sized corpora, as raw frequencies do not accurately reflect relative frequencies. A lexical item that appears 50 times in a corpus of 200,000 words is not less frequent than one that appears 500 times in a corpus of 2 million words. Per million words seems increasingly to be the standard, as corpus sizes increase, although some researchers have used 100,000 (Biber, 1998, p. 32) some 10,000, and sometimes even 1,000 is used. (e.g. Goźdź-Roszkowski, 2011). In this thesis, notwithstanding the fact that the corpus was a little over 400,000 words, I used pmw as it was a default setting in Sketch Engine, which could not be changed (though of course, the normalised frequency can also be calculated by hand).

The analysis where the normalised frequency is more relevant was the investigation of the modal verbs, reported in Chapter 6. Lexical bundles and keyword occurrences are not reported in terms of raw or adjusted frequencies. As I was more interested in the use of modal verbs in patient information, and this was not a study to compare registers - i.e. it was the frequency of use of a modal verb relative to other modal verbs in the patient information that interested me, rather than the frequency of use relative to another register - I felt that the setting for the normalised frequency was not a concern. As for the decision to report both raw and normalised frequency, I have followed the advice of McEnery and Hardie who say: ‘It is usually considered good practice to report both raw and normalised frequencies when writing up quantitative results from a corpus’ (McEnery & Hardie, 2012, p. 51).

The minimum frequency default setting in Sketch Engine is 5, which means 5 raw frequency counts in the corpus. 8 raw counts, when normalised in my corpus of patient information of around 400,000 words, represents about 20 pmw. 20 pmw was the cut-off point used in the lexical bundle analysis, the methodology of which will be discussed in more detail later in this chapter.

3.4.2 Distribution

Distribution is also an important factor when conducting a corpus analysis. A word may appear 50 times in a corpus, but if 49 of those uses are found in one document, and thus the work of one writer, its use is considered idiosyncratic and not generalisable. Establishing a minimum range is standard practice in corpus studies and five is both common in the literature and was the Sketch Engine default. In my research, too, five was set as the minimum distribution.

Having presented the corpora I built and the software I used to carry out my analyses, I now turn to the methodologies of these analyses: keyword extraction, lexical bundle analysis and an analysis of the modal verbs used for instruction and obligation.

3.5 Methodologies: Corpus-driven and corpus-based approaches

There are generally two approaches to corpus analysis, what Tognini-Bonelli (2001) refers to as corpus-based, where lexical items are pre-selected and then searched for within a corpus, and corpus-driven studies, where there are no preconceived lists of expressions and 'recurrent patterns and frequency distributions are expected to form the basic evidence for linguistic categories; [and where] the absence of a pattern is considered potentially meaningful' (Tognini-Bonelli, 2001, p84). The three lexical analyses that make up this study of patient information and the approach taken varies accordingly. I used a corpus-driven approach in my analyses of keywords and lexical bundles reported in chapters 4 and 5, while in chapter 6 and my analysis of a range of modal verbs, the approach was corpus-based. The methodologies used are quite different from each other and will be presented below. I begin by describing the methodology of a keyword extraction.

3.5.1 Keyword Extraction

3.5.1.1 Introduction

The simplest definition of a keyword is that it is a statistically significant lexical item (Scott,1997). Keywords are generated by a computer using statistical calculations and thus this stage of the analysis is quantitative. The extraction of keywords is then followed by a semantic categorisation in order to establish the underlying themes in the discourse. These categories - all or some depending on the researchers' agenda - are then examined in context, taking careful note of how the words are used, especially their collocational partnerships. A keyword analysis is, therefore, both a quantitative and a qualitative analysis.

A keyword analysis is a very useful tool in healthcare language studies, especially those where real-world outcomes are a key objective. We have already seen a number of these studies in Chapter 2, where these real-world outcomes relate to the end-users, who are generally professionals or patients. For example, a deeper understanding of a psychologically-motivated condition (e.g Harvey and Brown, 2012) or evidence of the severity of adolescents' anxiety concerning the revealing of confidences (e.g Harvey et al., 2008) can be used to inform healthcare professionals' training and to improve adolescents' experiences of healthcare. Many register studies have shown keywords to successfully reflect the characteristics of a register, and the approach seems to have greater sensitivity than some purely qualitative methods (e.g. open-ended interviews) (e.g. Seale et al., 2006). The approach can give important lexical information about the information priorities of the register under investigation and can also reveal the discourse and themes prevalent in a text which may be hidden when examining the text with a purely qualitative approach. I believe that a keyword analysis is also complementary to an analysis of lexical bundles, as while the former can give us more information about the themes and beliefs in a text, the latter can tell us more about the underlying communicative function of a text. Lexical bundles will be the subject of the next chapter.

This section begins with a definition of a keyword, an explanation of the reference corpus, and mention of the settings and statistical tests that are applied in

Sketch Engine when generating keywords. This will be followed by a section describing the second stage of the methodology: the semantic categorisation.

3.5.1.2 Defining a keyword

The simplest definition of a keyword is, as we have seen in 3.5.1.1, that it is a statistically significant lexical item (Scott,1997). The item is statistically significant because its frequency in a corpus is compared to that in a second, reference corpus. All keyword analyses, then, involve the use of a reference corpus which serves as a comparison corpus; a statistical analysis is carried out which produces a frequency list of lexical items in the corpus under investigation when compared to the reference corpus - usually, though not always as we shall see, a large-scale, general corpus such as the British National Corpus (BNC).

An item may appear as a keyword with both positive or negative frequency, that is, appear more or less frequently than might be expected by chance. In Sketch Engine, however, keywords have only a positive frequency: a keyword is listed when it appears more frequently in the corpus under investigation than might be expected by chance. Keywords are not synonymous with terminology, though Sketch Engine offers the user the possibility of extracting what they refer to as ‘terms’ in addition to carrying out a keyword analysis. Terms, as defined by Sketch Engine, are two-word noun phrases (collocations to a language teacher or researcher) that appear with a greater frequency when compared to a reference corpus. A keyword as it is used in this thesis and the studies reported in Chapter 2 is not solely a noun but can be any kind of word class, including pronouns or conjunctions.

In Sketch Engine, as I will now explain, the user has some control over the balance between content and grammatical words in their keyword list. When setting up the software to extract the keywords, Sketch Engine offers the option of varying the degree of ‘rareness’ of the keyword, what they refer to as the ‘smoothing parameter’ as Figure 4 below shows. Varying the smoothing parameter results in more, or less, content words being extracted. In contrast, the closer to ‘common' the setting, the greater the likelihood of grammatical words being extracted.

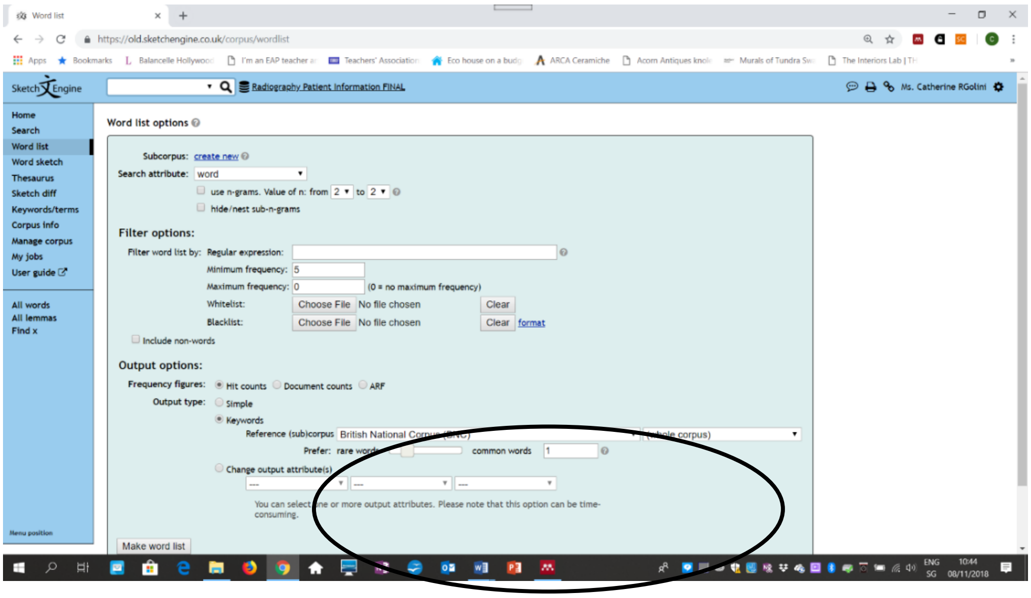


Figure 4 Screenshot of Sketch Engine; option to change word 'rareness'

For example, the setting at 1,000,000 (with the British National Corpus as the reference corpus) gives us the most common items, and we get the following in the first ten keywords: *the or, is, may, your, will, you, be, procedure, are.* As we see, there is only one content word when the setting is at 1,000,0000: *procedure*. *CT* appears in eleventh place (though placing or where a keyword is in the list is not necessarily important when dealing with keywords, as I explain in 3.5.1.8). In contrast, when the setting is at 1, the Sketch Engine default setting, we get the following words: *CT, MRI, imaging, x-ray, copyright, radiologist, RadiologyInfo.org, reviewed, physician, ultrasound*, which are all content words. When the smoothing parameter is set to 1000, on the other hand, we get the following: *procedure, may, your, CT, images, MRI, imaging, information, radiation, or;* a mix between content words and grammatical, with the former predominant.

What does this mean for the researcher? All of these lexical items are potentially interesting, of course, though there will be research questions that render content words more or less interesting and thus govern the choice of setting. I elected to set the smoothing parameter at 1 for this study, i.e. the default setting. I did not want to exclude grammatical words, but the content words, in particular, were my focus. There was no precedent for this, but given that this is the first study of keywords in patient information, it seemed appropriate. Later studies may wish to focus on grammatical as well as lexical items in the patient information, and thus choose a different parameter setting.

3.5.1.3 Statistical calculations

There have been a number of statistical analyses that have been carried out to generate keywords, though the most frequently used for keyword extraction within applied linguistics are Mutual Information (MI) (Church and Hanks, 1990), Log-likelihood (Dunning, 1993) and the t-test. In Sketch Engine, the statistical calculation used is what is referred to as 'Simple Maths', where the calculation is as follows:

$$\frac{fpm_{rmfocus} + N}{fpm_{rmref} + N}$$

where

$\mathrm{fpm}_{rmfocus}$ is the normalized (per million) frequency of the word in the focus corpus,

fpm_{rmref} is the normalized (per million) frequency of the word in the reference corpus,

N is the so-called smoothing parameter (N = 1 is the default value)[2]

3.5.1.4 Reference corpus

As I have stated already in this chapter, keywords are words that are statistically more frequent (or less frequent) in the corpus under-investigation than they are in a reference corpus, also referred to as a comparison corpus. The standard approach when investigating general vocabulary is to use a general reference corpus, such as the BNC. However, an increasing number of studies of specialised discourse are electing to use a specialised corpus as a reference corpus.

As we saw in chapter 2, Seale et al. (2006) and Seale and Charteris-Black (2008) did not use a general reference corpus in their study. In the 2006 study, the authors were interested in learning how forum conversations by women about breast cancer

[2] https://www.sketchengine.eu/documentation/simple-maths

differed from those by men about prostate cancer, each set of forum conversations was used as the reference corpus for the corpus under investigation, i.e. the corpus of breast cancer conversations had, as a reference corpus, the conversations about prostate cancer, and vice versa. I have suggested that such an approach is not usual, though there are other keyword studies, some of which have been referenced in the literature review in chapter 2, that use either a dual reference corpus or a single, domain-specific reference corpus (e.g. Baker, 2004; Goźdź-Roszkowski, 2011 and Grabowski, 2015). Explaining why this might be an appropriate step to take, Goźdź-Roszkowski (2011) says ’Comparing a range of specialized genres[3] from the same domain against a general reference corpus would inevitably lead to obtaining finding which may be highly homogeneous and probably valid for legal language in general but would not help identify features unique to a particular genre’ (p36).

In my study of keywords reported in chapter 4, I performed two separate keyword extractions and thus used two different reference corpora: the BNC and a domain specific corpus of general radiography. The BNC was used as it is widely available, comes pre-loaded in Sketch Engine and is often used as a reference corpus in the literature, meaning that it provides ‘a recognizable common ground for keyword […] comparisons' (Charles, 2009, p20). The domain-specific corpus of general radiography contains 719,209 words and is made up of a radiographer’s handbook, *Clark’s Positioning in Radiography*, a textbook, *Patient Care for Radiography* and research from *Radiography*, a specialised academic journal about radiography and thus with research by and for radiographers. The full details of this domain-specific corpus have already been reported in 3.2.2, while the reasons for using a domain-specific reference corpus in addition to the more-usual general corpus I have outlined earlier: a more targeted description of the register may be achieved which, alongside the results of the BNC comparison, can help build a more informed picture of the lexical characteristics of radiography patient information.

[3] Genre, as used by Goźdź-Roszkowski, is synonymous with my use of register in this thesis.

Having presented the software settings relevant to a keyword extraction, I now turn to the steps taken to extract a keyword list.

3.5.1.5 Extracting a keyword list

The first step in a keyword analysis is to extract a keyword list. In Sketch Engine, this can be done in one of two different ways: the first is found under the tab Keywords/terms while the other can be found under Word List. I elected to use the latter, though it is my understanding that both offer the same options, i.e. reference corpora, minimum frequency, search terms, etc. and both give the same results.

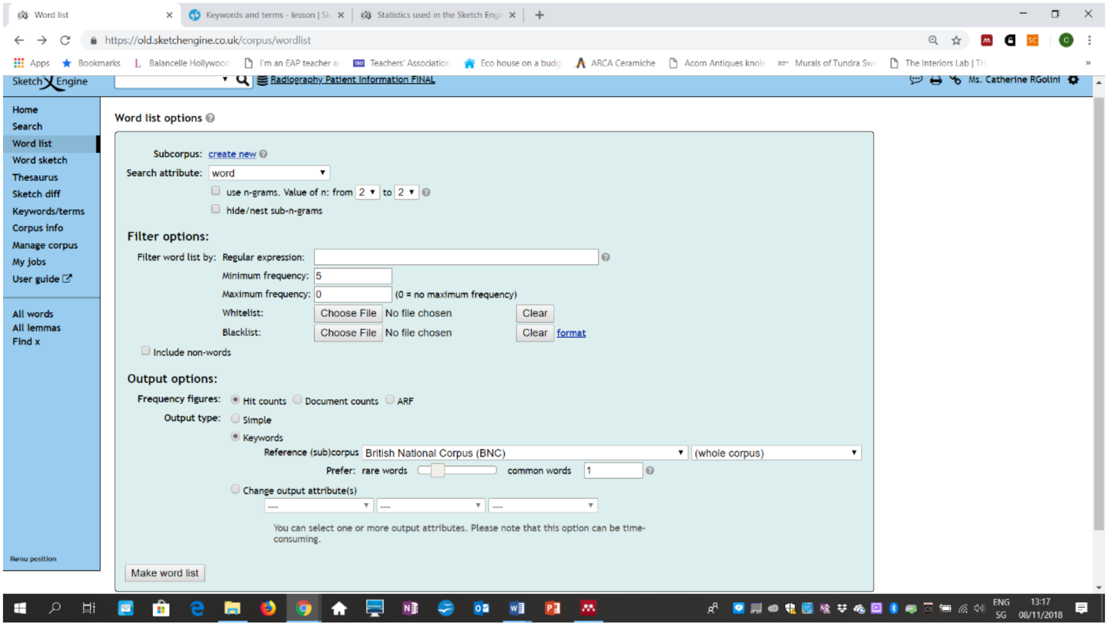


Figure 4 Extracting a list of keywords

3.5.1.6 Setting the options

The first setting to select is the Search attribute. I selected ‘word’ to be able to see unique examples of any form of a word in the final list. If the lemma option is chosen, the keyword *scan*, for example, would also include instances of *scans*, *scanned* and *scanning*, though clearly for a keyword list, these words need to be treated separately.

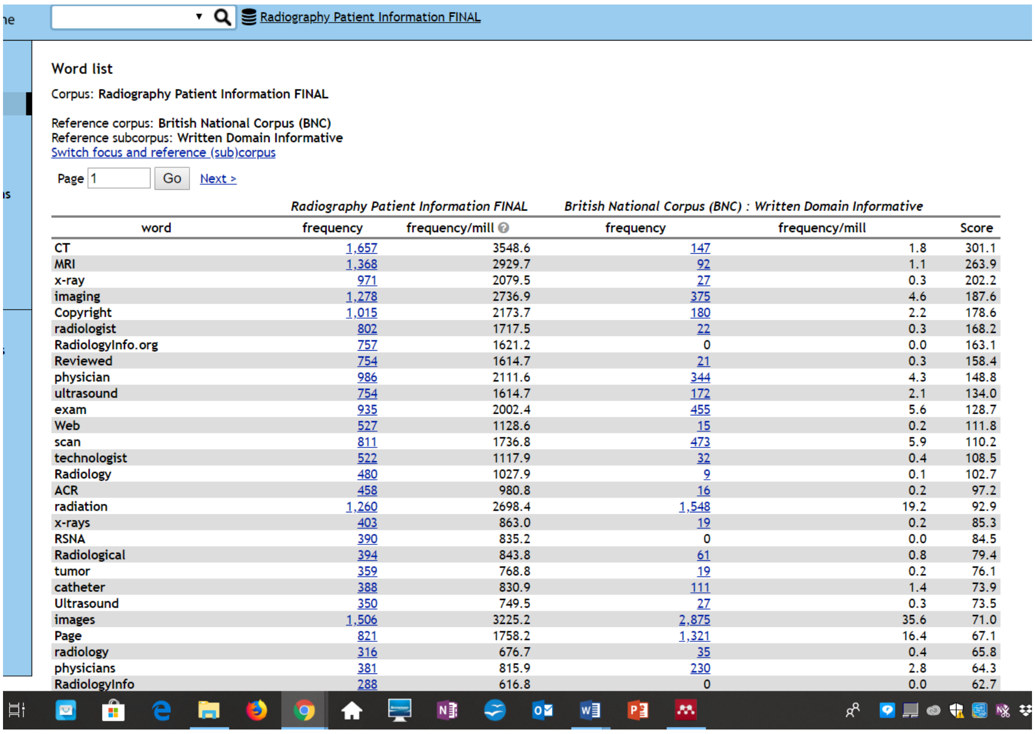
I did not use any of the settings in the Filter options as they are not, with the exception of the Blacklist, relevant for a keyword search. I did not change the minimum frequency as keywords are ordered by keyness and I intended to investigate the first 50, irrespective of how many appeared on my final list. A Blacklist is used to exclude words. For my keyword extraction, I did not wish to exclude anything. A Whitelist searches only for the words on the Whitelist - this is clearly not a keyword list if items are being included a priori.

The Output options are where we set the Keyword search (as opposed to a Simple search, which does not require a reference corpus and produces a frequency list instead), and where we select the reference corpus and adjust the smoothing parameter if required.

3.5.1.7 Keyword list

Once we have selected our parameters, the button ‘Make Word List’ will produce a keyword list. The results for a keyword extraction using a section of the BNC: Written Domain Informative can be seen below in Figure 6. (This was not the study carried out for this doctoral research but serves purely as an example).

There are 5 columns. The first two show the raw frequency and the per million words (pmw) normalised frequency of the keywords. As I have said earlier in this chapter, the pmw normalised frequency is the default setting in Sketch Engine and is not a parameter that can be changed. To the right are two further columns, showing the same figures for the reference corpus. The fifth and final column is the Score. This refers to the keyness score, which is a calculation of the frequency of a lexical item in the source corpus, compared to the reference corpus.



Word list

Corpus: Radiography Patient Information FINAL

Reference corpus: British National Corpus (BNC)
Reference subcorpus: Written Domain Informative
Switch focus and reference (sub)corpus

Page 1 Go Next >

	Radiography Patient Information FINAL		British National Corpus (BNC) : Written Domain Informative		
word	frequency	frequency/mill	frequency	frequency/mill	Score
CT	1,657	3548.6	147	1.8	301.1
MRI	1,368	2929.7	92	1.1	263.9
x-ray	971	2079.5	27	0.3	202.2
imaging	1,278	2736.9	375	4.6	187.6
Copyright	1,015	2173.7	180	2.2	178.6
radiologist	802	1717.5	22	0.3	168.2
RadiologyInfo.org	757	1621.2	0	0.0	163.1
Reviewed	754	1614.7	21	0.3	158.4
physician	986	2111.6	344	4.3	148.8
ultrasound	754	1614.7	172	2.1	134.0
exam	935	2002.4	455	5.6	128.7
Web	527	1128.6	15	0.2	111.8
scan	811	1736.8	473	5.9	110.2
technologist	522	1117.9	32	0.4	108.5
Radiology	480	1027.9	9	0.1	102.7
ACR	458	980.8	16	0.2	97.2
radiation	1,260	2698.4	1,548	19.2	92.9
x-rays	403	863.0	19	0.2	85.3
RSNA	390	835.2	0	0.0	84.5
Radiological	394	843.8	61	0.8	79.4
tumor	359	768.8	19	0.2	76.1
catheter	388	830.9	111	1.4	73.9
Ultrasound	350	749.5	27	0.3	73.5
images	1,506	3225.2	2,875	35.6	71.0
Page	821	1758.2	1,321	16.4	67.1
radiology	316	676.7	35	0.4	65.8
physicians	381	815.9	230	2.8	64.3
RadiologyInfo	288	616.8	0	0.0	62.7

Figure 5 Keyword list

3.5.1.8 Keyword order

Keywords have been ordered by keyness score, as we see in Figure 6. This does not mean, however, that the order of keywords extracted by the analysis is necessarily significant, as it would be in a frequency list. Scott (2010) says ‘the order of KWs [keywords] is not intrinsically trustworthy, because it depends not only on the frequency in the text we are studying…but also on their frequencies in the reference corpus’ (p50). The greater the number of keywords extracted, the greater the possibility that the inclusion of some is based on statistical chance, says Scott (2010, p50). Given this, it is wise to see keywords as suggesting ‘to the prospector areas which are worth mining but they are not themselves nuggets of gold.’ (Scott, 2010, p51). It is also why the keyword extraction is followed by a semantic categorisation of all or some of the keywords and their subsequent investigation in the context of the corpus.

The semantic categorisation, which is an important step in the methodology, will now be presented.

3.5.1.9 Semantic categorisation

Keywords are then grouped semantically, either using the researcher's intuition, with or without the aid of a dictionary or by using software. I categorised the first 50 keywords for both of the analyses manually, by inspecting their use in the corpus and by using the Word Sketch facility in Sketch Engine. I decided to inspect the first 50 keywords, rather than the first-100 or 150, as I was carrying out two separate keyword extractions: one with the BNC as reference corpus and the second with the general radiography corpus. These have been described in 3.2. Word Sketch, a proprietary tool, gives lexical and grammatical collocational information about a word. It was referred to earlier in this chapter in 3.3 and illustrated with a screenshot of the Word Sketch for the noun *pain*.

Inspecting a keyword in context is vital; sometimes a word can appear in two categories or the most common meaning of a keyword turns out not to be the way it is being used in the corpus. An example from my data is *up-to-date*. I initially thought it referenced the information, but closer inspection revealed that, while it does modify the noun *information*, it is used only in formulaic, legal disclaimers that are included in some patient information, and thus it was removed from the category Information (which was generic in nature and included *leaflet*) and added to the Legal category. There were no words in my keyword lists that could be placed in two categories, that is, there were no keywords used with two distinct meanings.

Once the semantic categories have been established, the words can be investigated in the corpus, and their connotations and their collocations examined. As I have already said, I used Word Sketch to do this and examined the sections in the corpus where these keywords were used. At this point, I was able to draw some conclusions about the use of the keywords in the corpus and what they reveal about the salient themes in patient information for radiography. These themes will be presented and discussed in Chapter 4.

This concludes the presentation of the keyword methodology. In the next section, I will present the methodology of the lexical bundle analysis, the second corpus-driven method used in this analysis of patient information for radiography.

3.5.2 Lexical bundle analysis

3.5.2.1 Introduction

Lexical bundles are multi-word lexical sequences that frequently reoccur in a register, e.g. *in the light of* and *at the end of*. They have been described as ‘characteristic features of language use in particular settings’ (Hyland, 2008, p8) and as ’text building blocks’ (Biber et al., 2004, p443). Usually transparent in meaning, they tend to be incomplete and often bridge two structural units, i.e. a clause or phrase, very often functioning as the pragmatic head of an utterance and acting as an interpretative frame for the discourse that follows (Biber and Barbieri, 2007, p8.)

Variously referred to in the literature as formulaic sequences (Wray 2002; Schmitt and Carter 2004), lexical bundles (Biber & Conrad, 1999), n-grams (Stubbs and Barth 2003) or lexical phrases (Nattinger & DeCarrico, 1988), lexical bundles have received increasing attention over the last two decades, though, as we have seen in Chapter 2, there have been very few investigations of lexical bundles in medical registers. For McEnery and Hardie (2012, p110), lexical bundles are, ‘methodologically and technically’, simply recurring sequences of *n* words, i.e. n-grams, though they add that the term lexical bundle has become associated the work of Biber and colleagues on register description, and on their focus on the structural and functional interpretation of lexical bundles. As it is the structural and functional interpretation that interests me, and thus it is Biber and colleagues’ terminology and approach that I have chosen to use, lexical bundle is the terms I used.

While earlier studies of chunks of language relied on intuitive lists of prefabricated expressions (e.g Pawley and Syder, 1983; Nattinger and DeCarrico, 1988; 1992), corpus software has permitted an evidence-based approach to studies of bundles, with Altenberg’s study (1998) of the phraseology of spoken English, being one of the earliest.

We have seen already that corpus studies fall broadly into two camps: corpus-based, where lexical items are pre-selected and then searched for within a corpus, and corpus-driven studies, where there are no preconceived lists of expressions and ‘recurrent patterns and frequency distributions are expected to form the basic evidence for linguistic categories; [and where] the absence of a pattern is considered potentially meaningful' (Tognini-Bonelli, 2001, p84). Such studies also establish cut-off points and dispersion requirements in order to identify the lexical bundles that are frequent and worth investigating in the corpus. A lexical bundle analysis is a corpus-driven study.

In the section that follows, I present the methodological steps taken to conduct a lexical bundle analysis, which, for this doctoral study, was an analysis of 4-word lexical bundles. These bundles are less common than 3-word bundles, which occur very frequently in both spoken and written discourse: Conrad and Biber (2004) claim that 25% of the words in conversation are found in 3-word bundles, while the most frequent 3-word bundle in conversation (*I don't know*) appears repeatedly at over 1,000 times per million words (Conrad & Biber, 2004). 4-word bundles, then, are less common than 3-word bundles but are not as rare as 5-and 6-word bundles, meaning an analysis of 4-word bundles results in a sufficient, but not overwhelming quantity, of data.

3.5.2.2 Identifying 4-word lexical bundles

Before the two classifications described above can be made, a list of lexical bundles must be extracted and identified. The first stage of this process is automated. In Sketch Engine, lexical bundles are referred to as n-grams. The function tab is Word list, where we also extracted keywords. The options that must be selected are seen in Figure 7.

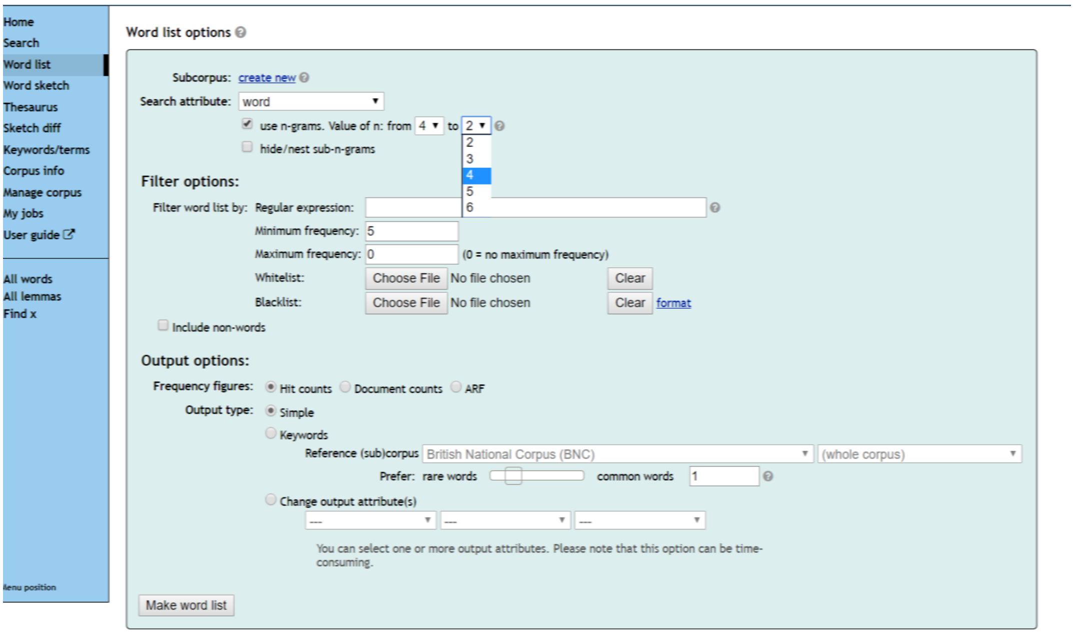


Figure 6 Options selected to produce 4-word bundles

The selection n-gram is made, with a value from 4 to 4. This will give us only 4-word lexical bundles. Had we also wanted to investigate 5-word bundles, we would have selected the value 4 to 5. Hiding or nesting sub n-grams was left unchecked as I was investigating only 4-word bundles, and thus did not need to see that there was also, in some cases, a 3-word bundle contained within the 4-word bundle. This is a feature of lexical bundles that can sometimes make their identification difficult. I will discuss this further in section 3.5.2.6.

There was no Blacklist or Whitelist, the Frequency figure was Hit, which counts each occurrence in a text, and the Output option was Simple. Unlike a keyword analysis, a lexical bundle analysis does not require a comparative or reference corpus but it does require that a minimum number is set, to establish a cut-off point. This will be discussed in the following section.

3.5.2.3 Cut-off point and dispersion

The minimum frequency default in Sketch Engine is 5. This setting sets a cut-off point, below which the lexical bundles will be ignored. 5 is low, however, and I was concerned it would result in a lot of data. Previous studies (Cortes, 2013; Csomay, 2013; Hyland, 2007) have set the minimum frequency to 20 per million words. While 40 is also a common cut-off in similar studies (e.g Biber and Barbieri, 2007; Goźdź-

Roszkowski, 2011) it is always fairly arbitrary and much depends on the size of the corpus - mine was small - the researcher’s preferences and the length of the bundle. A lower cut-off point is generally selected for the rarer, five-and si-xword bundles (e.g Cortes, 2013). It was also the case that a full description of four-word lexical bundles in procedural patient information was not the focus of this doctoral thesis but, rather, one analysis out of three. However, I examined the data extracted with the minimum setting at 20 and also at 40, finding many interesting bundles below 40 that would not be extracted if I did not lower the cut-off point. I decided, on this basis, to set the cut-off point to 20 to include these bundles.

To control for individual peculiarities - that is, one or two writers favouring a certain bundle - their dispersion is also a factor. Bundles need to appear in at least five individual texts in the corpus to be included in the final list.

3.5.2.4 Identifying the bundles

The first impression of the list produced, however, may be a little overwhelming, as Sketch Engine produces multiword units that are 4-words in length, though not necessarily 4-word lexical bundles. Reducing the size of the data necessitates the application of exclusion criteria, which is presented in 3.5.2.5. It also involves identifying true 4-word bundles, which, as I will now explain, is not always a straightforward procedure.

It has often been pointed out that 4-word bundles can include 3-word bundles and that 5-word bundles can contain 4-and 3-word bundles, etc. (Cortes, 2004; Hyland, 2007), which was indeed evident in my data, and the 2300+ bundles initially extracted by the corpus software also included many fragments or part bundles. It was not always straightforward to decide what was a fragment, or what could stand as a true four-word bundle, especially as some four-word bundles are better treated as a three-word bundle with a slot, e.g. *during and after the. During and after* can also be followed by an indefinite article or a noun with zero article, e.g. *during and after an operation* or *during and after childbirth*. The article *the* is dependent on the noun that follows and thus treated as a part of an optional slot. Likewise, *at the top of*, which is a four-word bundle. *At the top of the* is not, however, a fiver-word bundle, but a four-word bundle with a slot. The slot can be filled with a definite article, an indefinite article or a possessive, e.g. *at the top of the world*; *at the top of a tall building*; *at the top of his game.* The decision process was time-consuming and on occasion, I turned to others for advice.

3.5.2.5 Exclusion criteria

The exclusion criteria applied to the list is presented below in Table 4.

Table 4 Exclusion criteria applied to the extracted lexical bundles

Exclusion Criteria
1. Fragments of other bundles i.e. *eat or drink any; tip of the part*
2. Topic/Name specific e.g *in X Plain-T; University College Central Clinic*
3. Bundles with random or meaningless numbers or symbols e.g. *know page 40 if;*
4. Web noise e.g. *at www.radio.com*
5. Complete phrases e.g. *do not copy this; contact us for information*

The largest category excluded was the first: fragments of other bundles, along with phrases that were considered to be complete. In Chapter 2, I questioned some of the 4-word bundles Grawbowski (2015) presents in his study, as I believe that some of these bundles are examples of complete phrases or of fragments e.g. *special precautions for storage*; *be used with caution*; *(the) dose should be reduced*. There were many such examples in my data that needed to be removed. Structurally complete bundles that are classed as lexical bundles, such as *on the other hand* (Conrad & Biber, 2005) do exist. What differentiates these from phrases that are not considered bundles is that these are formulaic in nature, unlike Grabowski's (2015) examples above.

Once the exclusion criteria have been applied, and 4-word bundles identified, the remaining bundles are then classified structurally before they are assigned a discourse function. These two steps are presented below.

3.5.2.6 Structural types

A lexical bundle analysis applies two classification processes to the extracted bundles, and the first of these classifications is structural. The structure of a bundle is significant: while bundles are generally not complete lexical units (with just 15% found to be complete in conversation and 5 % complete in academic prose (Biber et al. (1999)), lexical bundles do possess clear, structural characteristics and different registers show preferences for different structural types.

Many bundles found in spoken discourse are made up of verbs and clausal components, such as *I want you to*, while 90% of bundles in conversation include a verb (Conrad & Biber, 2005). In contrast, many bundles found in written, more formal prose contain noun phrases and prepositional phrases, e.g. *in the middle of.* These bundles are also far more likely to contain passive structures. The taxonomy of structural categories for academic prose as presented in Biber et al. (1999) can be seen in Table 5.

Table 5 Structural classification of lexical bundles in academic prose adapted. (Biber et al. 1999, p. 1015–1024)

Structure	**Examples**
Noun phrase with *of*-phrase fragments	*the end of the, the base of the*
Noun phrase with other post-modifier fragments	*the way in which, the relationship between the, such a way as to*
Prepositional phrase with embedded *of*-phrase fragments	*as a function of, as a result of*
Other prepositional phrase (fragment)	*as in the case of, at the same time as*
Anticipatory it + verb phrase/adjective phrase	*it is possible to, it may be necessary to*
Passive verb + prepositional phrase fragment	*is shown in figure/fig., is based on the*

Copula be + noun phrase/adjective phrase	*may be due to, is one of the*
(Verb phrase +) *that*-clause fragment	*has been shown that, that there is a*
Verb/adjective +) *to*-clause fragment	*are likely to be, has been shown to be*
Adverbial clause fragment	*as shown in figure/fig., as we have seen*
Pronoun/noun phrase + *be* (+…)	*this is not the, this did not mean that*
Other expressions	*as well as the, may or may not,*

While a range of bundles types is found in both conversation and academic prose, not all bundles are used with equal frequency, in fact, there is commonly great repetition of just a few types. Conrad and Biber (2005) found that in their study comparing academic prose with conversation, just three bundle types accounted for 70% of the total number of 4-word bundles in conversation, and all three of these bundles included a verb. In academic prose, just two bundle types, both a noun-phrase type, represented over 60% of the 4-word bundles. These two bundle types were barely used in conversation. These marked differences, say Conrad and Biber (2005), are consistent with the differences seen between these registers at the word, clause and phrase level and are related to the communicative functions of the bundles.

The second categorisation that takes place in a lexical bundle analysis is to assign discourse function, which we will now turn to.

3.5.2.7 Assigning discourse function

A taxonomy of bundle meaning and purpose, first described by Cortes (2002) and later extended in Biber et al. (2003; 2004), categorised lexical bundles into three broad functions: stance, referential and discourse. An explanation of each was provided in Biber et al. (2004):

> Stance bundles express attitudes or assessments of certainty that frame some other proposition. Discourse organizers reflect relationships between prior and coming discourse. Referential bundles make direct refer

> ence to physical or abstract entities, or to the textual context itself, either to identify the entity or to single out some particular attribute of the entity as especially important. (p. 384)

Later studies expanded upon this initial taxonomy of discourse function (e.g. Cortes, 2004; Cortes, 2006; Cortes, 2013) while changes to it have also been made (Hyland 2008), as a result of the specific characteristics and thus discourse functions of the register being studied. In this doctoral study, I used the taxonomy set out by Biber et al. 2004 and shown in Table 6.

It is necessary at this point to underline the fact that assigning a discourse function is sometimes a straightforward process as the function is clear, though sometimes it is a process that is more complex and necessitates a careful examination of the context surrounding the bundle. A single lexical bundle can have multiple functions, even, as Biber et al. (2005) point out, in a single occurrence. *Take a look out* can function both as a topic introducer and a directive, while the bundles *the beginning of the* and *at the end of* can function as a time reference, place reference, or text deictic reference. (p. 384). Examples of these bundle types were also found in my data and will be discussed in Chapter 6.

Table 6 Functional classification of lexical bundles (Biber, Conrad & Cortes, 2004, p. 384–388)

Functional Classification of Lexical Bundles

1. Stance expressions	**2. Discourse organisers**	**3. Referential bundles**
Express attitudes or expressions of certainty that frame some other proposition	Reflect relationships between prior and coming discourse	Make direct reference to physical or abstract entities or to the textual context itself
A. Epistemic	A. Topic introduction/focus	A. Identification/focus
I don't know if, I think it was	*Now let's look at*	*that's one of the; of the things that*
B. Attitudinal/modality stance	B. Topic elaboration/clarification	B. Imprecision
B1) Desire	*what this means is*	*a little bit like; a bit more than*
If you want to; I like to go.		C. Specification of Attributes C1) Quantity specification
B2) Obligation/directive		*there's a lot of; how many of you.*
you will have to; it is important to.		C2) Tangible framing attributes
B3) Intention/prediction		*at the end of; on top of the*
it's going to be; I'm not going to		C3) Intangible framing attributes
B4) Ability		*the nature of the; in the case of*
to be able to; can be used to		D. Time/place/text reference D1. Place *in the department of*
		D2. Time *at the same time*
		D3. Text *in the next section; as shown in figure*
		D4. Multifunctional
		at the end/beginning of;

This concludes the section on the methodology of a lexical bundle analysis. Both this and the keyword analysis already described are corpus-driven, whereas the third and final methodology used in this investigation of patient information for radiography, and which we will now turn to, is a corpus-based investigation. In a corpus-based analysis, the researcher searches the corpus for one or more linguistic items that they have decided upon a priori. For my part, I had decided to investigate the modal and semi-modal verbs used to express instructions and obligations in patient information as I was interested in how one of the primary functions of patient information, that of instructing, was linguistically realised. In short, I wanted to know more about the ways patients are told what to do in written patient information.

3.5.3 Modal verbs for instructions

3.5.3.1 Introduction

One of the functions of procedural patient information is to tell patients what to do, or how or when to do it. With radiography, some examinations necessitate that the patient does not eat or drink beforehand, while other exams require the patient to remove metal objects in the body. Patients are required to tell hospital staff if they have allergies, as these can make the use of contrast dye inadvisable. Female patients are expected to inform the staff if there is any possibility that they are pregnant. We also know that modal verbs are common to certain types of medical writing (Vihla, 1999) though neither procedural or pharmaceutical patient information was discussed by Vihla (1999) (presented in Chapter 2) and, to my knowledge, there have been no studies that have investigated modal and semi-modal verbs for obligations and instructions in patient information.

We currently have little idea, then, of how these words are used in these kinds of healthcare materials.

The core modals are generally held to be *can, could, may, might, shall, should, will, would, ought (to)* and *need* (Downing and Locke, 1992; Quirk et al., 1985), while the category of semi-modals can include a range of items including *dare to, need to, have (got) to, be able to and be going to*. Semi-modals generally express a meaning that can be paraphrased with a core modal verb, and while some semi-modals are fixed expressions that cannot be marked for tense and person, e.g. *had better*, other semi-modals verbs can be marked for tense and person, e.g. *she has to be at work early* and can also combine with certain modal verbs e.g. *he should be able to* and *I might have to tell him.*

3.5.3.2 Types of modal meanings

The main function of modal and semi-modal verbs (henceforth modals) is to express stance. (Biber, Conrad & Leech, 2002). Modals can possess two different types of meaning, usually referred to personal (intrinsic) and logical (extrinsic) and are generally placed into one of three categories, depending on their meaning. Each category contains personal (intrinsic) and logical (meanings). The meanings are usually referred to as epistemic, deontic and dynamic, The meaning attached to modals and the names given to the types of modals will be discussed in more detail in Chapter 6, though I will now present a very brief description of these terms before I continue to describe the methodology.

Epistemic modality is concerned with the speaker’s attitude towards the proposition or the situation described in the proposition. This can range from an expression of doubt through to certainty. Modals commonly used to express epistemic modality include *can*, *may* and *could*. In table 4 in this chapter, we saw that lexical bundles can have an epistemic discourse function. Biber et al. (1992) refer to this category as permission/ability.

Deontic modality is concerned with obligation, requirement and necessity. *Must*, *have (got) to*, (particularly in British English) *should* and *need to* are commonly used to express deontic modality. Lexical bundles can also have a deontic function in discourse, as we saw in Table 4 earlier in this chapter. Biber et al. (1992) refer to this category as obligation/necessity.

Dynamic modality is less straightforward to characterise. Broadly speaking it refers to ability or volition - though it, unlike deontic and epistemic modality, is not subjective (Palmer, 1990, p36) which suggests that it is not inherently modal. *Will*, *would*, *shall* and *be going to* appear in this category. Biber et al. (1992) refer to this category as volition/prediction.

Neither epistemic nor dynamic modality is the focus of this study, however. Deontic modality (obligation/necessity) is the category of modal that is the subject of this analysis and thus modals and semi-modals that are used exclusively as deontic modals (rather than epistemic or dynamic) in patient information are the focus.

3.5.3.3 Methodology

The procedure for this analysis was comparatively straightforward, although there were three searches performed on three different corpora with 12 modals (listed below) the corpus of patient information; the corpus of consumer information and the corpus of general radiography. This initial frequency analysis included a range of the most common core modals and semi-modals including, though not restricted to those used to give instructions or to express obligations. The modals searched for were *will, would, can, could, may, might, must, should, have to, have got to, need to, ought to.* This was done in order to have an overview of the frequency of modal verb use in patient information which I could compare to what we know of the frequency in general discourse, spoken and written. Once this step had been carried out, the focus turned to those modals used for obligation, instruction and permission.

How the searches were carried out in Sketch Engine will now be described.

3.5.3.4 Search terms

In Sketch Engine, a search for a single item or phrase is undertaken by selecting the Search tab. The screen will offer a number of Query types: simple, lemma, phrase, word, character and CQL. A lemma will find all forms of a word, so entering *examine* will result in *examine, examined, examining*; for a search that gives you only your search term, word is the option. CQL is corpus query language, which is useful when parts of speech are being searched for (e.g. all adjectives, all conjunctions), and phrase will find examples of a sequence of tokens exactly as it is typed.

The so-called Simple search is more complex than the name implies, as Sketch Engine tries to guess what it is you are looking for based on the kind of search term you have entered. If you enter a lemma, the search is a lemma search. If you enter a term which is not a lemma, a word is searched. It was the Simple search that I used for the modal and semi-modals under investigation and the software treated my search terms as lemmas with the semi-modals and modals. This meant that the Simple search captured the changes for person and tense that took place with three of the semi-modals (*need to, have to* and *have got t*o), as well as including negative forms for all modals and semi-modals.

Absolute (raw) frequencies were normalised to 1 million. As I discussed in 3.4.1, normalising frequencies involves calculating the frequency of an item at 1 million words, or 100,000 or even 10,000. This is done so that comparison between different corpora can be made. Raw frequencies, on the other hand, while reported, cannot tell us much if the corpora are of different sizes. 1 million was the figure chosen as this a Sketch Engine default.

A second analysis was performed, using the same modals, but this time the corpus of consumer information, presented in 3.2.1 was used. The rationale for this has been presented in 3.2.1: patients are increasingly referred to as consumers or clients. The information produced for them may indeed be a type of consumer information, but I intuitively feel that the way patients are spoken to in procedural health information differs from the way consumers are spoken to consumer information. This may because the voice of authority in medicine is not the same as that voice of authority in a consumer advice agency. Authority, which will be discussed further in Chapter 6, is a significant factor governing the way in which obligations and instructions are presented - and perceived by the receiver. The relationships between obliger and obliged, instructor and instructed are quite different in medicine and in consumer advice. So, one way of investigating difference is to look at the way obligations and instructions are expressed in materials from both areas.

Once this second analysis had been carried out, a third analysis was also conducted on the corpus of radiography, which was presented in 3.2. The latter is a corpus of 719,209 words, made up of radiography research, textbooks, handbooks and patient information. I thought it would be interesting to compare the use of modal and semi-modals verbs in this corpus as it contained a radiographer's manual, and a course book, and thus was likely full of instructions and directions. The readers of this materials, however, are qualified and trainee radiographers, i.e fellow medical professionals. They are not patients.

The next step in my methodology was to understand how the different modals I had searched for were being used in the corpus. This was necessary in order to separate out those modals used to oblige or instruct from those with other meanings, as described earlier in 3.5.3.2. Some modals can have different meanings, it is not immediately obvious what meaning it carries and this necessitates an examination of the word in context. *Can* is a modal that often requires an examination of context before assigning meaning, as is *should*. This will be discussed in more detail in chapter 6. Examining the modal in context involved sampling, the procedure for which I will now describe.

3.5.3.5 Classifying the modals in the corpus

A random sample of 100 of each modal was extracted and examined in context in the corpus. Sketch Engine offers this sampling facility, and while 250 is the default, I selected 100 to make the qualitative analysis manageable, as there were eight modals (I will use modal henceforth to include modal and semi-modal verbs) to be investigated: *can, will, should, must, have to* and *need to*. My initial investigation had resulted in zero hits for *ought to, had better* and *have got to*, so these items had been discarded. *Would* and *could* had also been removed at this point as these modals are not used to instruct or oblige.

The sampling and investigation demonstrated that *can* was overwhelmingly used as an epistemic or dynamic modal and *will* as a dynamic modal. while *must, have to*, *should* and *need to* were found to be always or predominantly used with a deontic meaning. These four modal verbs were then investigated in detail in the corpus and their collocational relationships were examined. These findings are presented and discussed in Chapter 6.

In this chapter, I have presented the corpora used in my study of patient information for radiography, along with a detailed presentation of the contents of each corpus and the procedure, in Sketch Engine, of the building of the three corpora. I have presented, too, the three corpus methodologies used in my study. Each method is different, and the first two are known as corpus-driven methods as the researcher does not decide what lexical item(s) will be investigated before the computer analysis is performed. The results of the analysis provide the researcher with areas for further investigation. The first method is also one of the more commonly-used approaches in corpus-assisted healthcare language studies, that of a keyword analysis. The different stages in a keyword analysis were presented which was followed by a description of a lesser-known method in studies of healthcare language, that of a lexical bundle analysis. The three different steps of the methodology have been described. I have concluded the chapter with a description of the third investigation conducted for the study: an investigation of the use of modals verbs for obligations and instructions in patient information. After an overview of the types of modal verbs according to the literature, and following a description of the initial frequency analysis carried out for comparative purposes, I have described the steps I took to identify, and then examine, four deontic modal verbs common to patient information for radiography.

We now turn, in the following chapter, to a detailed presentation of the keyword extraction, carried out on my corpus of patient information for radiography.

4. Keywords in Patient Information for Radiography

This chapter presents the findings of a keyword analysis of patient information for radiography, the first of three analyses carried out for this doctoral study. We have already seen how useful keyword studies can be for revealing otherwise hidden attitudes and beliefs in healthcare discourse (e.g. Harvey et al., 2008) so it seems an appropriate and potentially useful analysis with which to begin my investigation of patient information. After a brief definition of a keyword, I will give a brief overview of the methodology, which has been presented in detail in the preceding chapter. I will then present an overview of the literature; key, keyword studies (e.g. Adolphs et al., 2004; Seale et al., 2006) have been discussed in detail in Chapter 2. I follow this with the results and conclude with a discussion of the findings.

4.1 Keywords

The simplest definition of a keyword, as we have heard already, is that it is a statistically significant lexical item (Scott,1997). The item is statistically significant because it appears with unusual frequency in a given text. All keyword analyses, then, involve the use of a reference corpus; a statistical analysis is carried out which produces a frequency list of lexical items in the corpus under investigation when compared to the reference corpus - usually, though not always, a large-scale, general corpus such as the British National Corpus (BNC). In this study, as I have explained in the previous chapter, I use both the BNC and a specialist corpus in order to reveal key themes that may have remained hidden by using solely a general corpus. The choice of reference corpus is important, says Scott and Tribble (2006; p65) but the greatest concern seems to be size:

> while the choice of reference corpus is important, above a certain size, the procedure throws up a robust core of KWs whichever the reference corpus used. These core KWs have largely but not exclusively to do with what the text is about; a few others are usually found which reflect some other stylistic feature.

There have been a number of statistical tests that have been used to generate keywords, though the most frequently used for keyword extraction within applied linguistics are Mutual Information (MI) (Church and Hanks, 1990), Log-likelihood (Dunning, 1993) and the t-test. An item may appear as a keyword with both positive or negative frequency, that is, appear more or less frequently than might be expected by chance. Sketch Engine uses what they refer to as Simple maths, and the calculation has been presented in 3.5.1.3. Keywords are not terms or terminology, though Sketch Engine, the software used in this analysis, offers the user the possibility of extracting terms in addition to carrying out a keyword analysis. Terms, in Sketch Engine, are two-word noun phrases that appear with greater frequency when compared to a reference corpus. ESP teachers may well refer to them as noun collocations. A keyword, however, can be any kind of word class and is not restricted to nouns.

A keyword analysis can provide an entry point into the data - Scott (2010) refers to them not as gold nuggets, but valuable indications that the text is worth mining - though in itself a keyword cannot tell us much as it does not give us any information about the use of the word in the register under investigation. The quantitative analysis is always followed by a qualitative investigation of selected keywords in context, often referred to as KWIC. The extracted keywords are also categorised semantically, which is a particularly helpful step when there is a lot of data or when semantic themes are not immediately obvious. These steps have been described in detail in the preceding chapter. Keywords, then, can ‘reveal not only a great deal about the subject matter, the ‘‘aboutness’’ of a particular genre, but they can also specify the salient features which are functionally related to the genre (Gozdz-Roszkowski, 2011, p35).

A keyword analysis has the potential to be a very useful tool to reveal more about the linguistic character of patient information and the topics that the discourse prioritises, though, to date, keyword studies of procedural patient information have not been carried out. A handful of studies of PILs (pharmaceutical patient information) have utilised the keyword method (Grabowski, 2013; 2015; 2017). Grabowski (2015) is reviewed 2.5.5. In applied linguistics more broadly and in healthcare communication studies more specifically the method is often used, and some of the studies from the literature I will overview in the following section. Other studies, particularly those relevant to medical discourse and to this doctoral thesis, have already been reviewed in detail in Chapter 2.

4.2 Keyword studies in the literature

Establishing the lexical 'aboutness' of a discipline lies behind much of the literature on keywords in the field of English for Specific Purposes (ESP). An ESP keyword analysis is of theoretical interest, of course, but it also has a great practical utility with the development of discipline-specific wordlists and teaching materials. Lecturers and teachers of discipline-specific English find that there are rarely published coursebooks to fall back on (publishers do not consider ESP to be as lucrative a market as EAP, English for Academic Purposes and thus are unwilling to invest (Bennett, 2010)), while many educational practitioners do not have expertise in the discipline or sufficient time in the programme to cover everything they feel their students require. Deciding what is essential to teach, what words their students really need in order to become members of the specific discourse community, becomes a priority.

ESP and academic word lists developed that have used the keyword approach include Gilmore and Millar (2018) who look at the language of civil engineering research papers; Watson-Todd (2017) who consider engineering more broadly; and Pacquot (2007) who applies the criterion of keyness to the more usual criteria of frequency, range and evenness of distribution for the development of an academic word list. Range and frequency are the criteria more often used in word list development (e.g. Coxhead, 2000; Gardner and Davies, 2014, Hsu, 2014; Mudraya, 2006; Wang, Liang and Ge, 2008; Ward, 2009) though Pacquot takes the view that, for productive purposes, Coxhead's (2000) Academic Word List (AWL) is less useful, as it excludes high-frequency words that have an important productive function in academic discourse. Keyword analyses, on the other hand, do not exclude on the basis of general frequency.

The concept of 'keyness' has been used in corpus-based studies which aim to reveal what Baker (2004) calls 'discourses' in the language, that is, concepts 'that may help to highlight the existence of types of (embedded) discourse or ideology' (Baker, 2004, p347). Examples of such studies include Johnson, Culpeper and Suhr (2003), who use keyword extraction to explore the discourses of political correctness (PC) in three British newspapers over the course of five years, finding an overall decline in the use of PC terms, while Baker (2004) looks at keywords in gay and lesbian erotic fictional narratives in order to identify how identity is constructed differently in each genre. Knight, Walsh and Pappagianidis (2015) investigate the discourse of e-transactional language - eBay listings - using, among other tools, a keyword analysis of the terms used by experienced and amateur eBay sellers.

Studies which aim to uncover the lexical characteristics of a particular register or domain include Goźdź-Roszkowski (2011) who investigates keywords in a range of legal registers and Grabowski (2013; 2015) looks at keywords in pharmaceutical registers. Grabowski (2015) is discussed in more detail in Chapter 2.

The introduction of the keyword technique to medical practitioners and other researchers outside the field of applied linguistics has been present in a number of studies that relate to healthcare discourse. Seale, Ziebland and Charteris-Black (2006) and Seale and Charteris-Black (2008) use a keyword analysis to investigate the impact of gender, and gender and age, respectively, on patients' experience of illness and health conditions. These studies are discussed in more detail in Chapter 2. The 2006 study, which used a corpus of forum postings and transcribed interviews, was published in the journal *Social Science and Medicine*, a particularly important journal for healthcare communication studies. The paper introduced the concept of 'keyness' to an audience of social scientists, making the case for its use alongside more traditional qualitative methods.

Gender difference as it relates to health information was the principal focus of both Seale et al. (2006) and Seale and Charteris (2008) and the two studies have findings regarding gender and information-seeking behaviour that are supported in the literature (e.g. Bidmon & Terlutter, 2015; Ek, 2013; Rice, 2006; Rutten, Squiers & Hesse, 2006). These studies confirm not only that gender has an influence on how people look for healthcare information, but that a difference in the type of information being sought is also seen. Many studies (e.g. Ek, 2013) also show that women are much more likely than men to engage in health information seeking. As Ek (2013) says ‘When it comes to health, women seem to be more engaged, more involved, more attentive and apparently better-informed decision-makers.’ (p742).

There is nothing fixed about the linguistic performance of gender (Seale and Charteris-Black, 2008), but if men and women tend to look for different information and focus on different aspects of health, illness and treatment, it seems reasonable to consider whether existing patient information, such as that in the corpus used in this thesis, is appropriate to these different needs. A keyword analysis can help reveal some of the information priorities in patient information for radiography which can help us answer the former question.

Seale and Charteris-Black (2008) also found that older men, in particular, like to reference medical experts and specialists such as radiologists, oncologists and consultants while it is woman of all ages who are much more likely to talk about, and talk with, nurses. The important role of nurses - and radiographers and technologists - in a patient’s experience of radiography does not seem to be reflected in their appearance in patient information, however. This will be discussed in more detail in 4.5.3.2.

Now let us turn briefly to the methodological steps taken in this keyword study. Full details of the methodology have been presented in Chapter 3.

4.3 Methodology

A full description of the keyword methodology has been presented in Chapter 3. There were two separate keyword extractions performed, one with the BNC (96,134,547 words) and one with the corpus of radiography (719,209 words).

For the comparison with the BNC, the minimum frequency was set at 5, the default value in Sketch Engine. For the analysis with the corpus of radiography as a reference corpus, the minimum frequency was also set at 5. With the setting at 5, the number of keywords extracted with the BNC as reference corpus was 991. When the corpus of Radiography was used, the number of keywords was 965. These numbers are pre-data cleaning which I explain in the following section.

The data cleaning (data cleaning refers to the removal of data - lexical items - that will not be considered in the final list) focused only on the first 50 items, as only these items were being evaluated in this study. Data cleaning resulted in a number of items being removed: three URLs or part domain names and two professional association acronyms. Acronyms that related to medical procedures or radiographic modalities and thus were part of the content were left (e.g. DCIS - *ductal carcinoma in situ;* CT - *computed tomography* or MRI - *magnetic resonance imaging*). These acronyms are considered part of the content as they are used in speech and writing to name procedures, modalities and examinations, and while they may be spelt out once in the text, they are then often used as acronyms without a definition. The results of the analyses are presented in the next section.

4.4 Results

4.4.1 Keyword lists

With the minimum setting at 5, the resulting list of keywords with the BNC as a reference corpus totalled 991. With the radiography corpus as a reference corpus, with all parameters untouched, the total was 965. The first 50 keywords of both analyses can be seen in Tables 7 and 8.

The Freq column refers to the raw frequency of the token in the corpus, which, as we have seen, means the number of individual occurrences of the item in the corpus. The Freq/mil is the adjusted frequency, per million words. This as we have seen, is the default setting in Sketch Engine. The score in the final column is the keyness score. The keyness score is the statistical calculation of the significance of the lexical item, though as we have seen in 3.5.1.8, the precise placings of the keywords do not mean that they are arranged in order of importance but in order of keyness. The first keyword *CT* is not necessarily any more significant than the 3rd (*radiation*) or the 20th keyword (*web*) as much depends on the frequency of these words in both corpora. A semantic categorisation and an investigation of the lexical item in the context of the corpus are necessary steps to understand the significance of a keyword.

179 *Table 7 First 50 keywords with BNC as the reference corpus*

Keyword	Freq	Freq/mill	Freq_ref	Ref/mill	Score
	Patient Information		*BNC*		
MRI	1368	2929.7	102	0.9	1536.1
radiologist	821	1758.2	26	0.2	1428.6
ct	1657	3548.6	184	1.6	1345.7
radiology	797	1706.8	46	0.4	1211.7
ultrasound	1106	2368.6	202	1.8	846.9
technologist	522	1117.9	38	0.3	836.1
tumor	370	792.4	21	0.2	668.4
imaging	1447	3098.8	441	3.9	629.4
tumors	289	618.9	4	0	598.6
interventional	276	591.1	9	0.1	548.2
x-ray	1311	2807.6	623	5.5	429.1
catheter	417	893	127	1.1	419.6
transducer	291	623.2	61	0.5	404.5
physician	986	2111.6	493	4.4	392.1
medications	252	539.7	56	0.5	360.8
radiological	416	890.9	171	1.5	353.6
scan	1090	2334.3	669	6	335.8

Keyword	Freq	Freq/mill	Freq_ref	Ref/mill	Score
	Patient Information		*BNC*		
x-rays	586	1255	327	2.9	321.2
radiologists	154	329.8	15	0.1	291.8
radiologic	138	295.5	2	0	291.3
angiography	184	394	42	0.4	287.5
tomography	283	606.1	125	1.1	287.3
anesthesia	131	280.5	5	0	269.5
radiotracer	125	267.7	0	0	268.7
copyrighted	131	280.5	6	0.1	267.3
sedation	224	479.7	93	0.8	263.0
radiographer	133	284.8	11	0.1	260.3
mammography	133	284.8	11	0.1	260.3
radiofrequency	124	265.6	3	0	259.6
embolization	116	248.4	0	0	249.4
exam	948	2030.2	868	7.7	232.8
noninvasive	105	224.9	0	0	225.9
copyright	1015	2173.7	999	8.9	219.8
download	130	278.4	33	0.3	216.0
physicians	404	865.2	363	3.2	204.7

Keyword	Freq	Freq/mill	Freq_ref	Ref/mill	Score
	Patient Information		*BNC*		
prostate	218	466.9	161	1.4	192.3
ionizing	105	224.9	22	0.2	188.9
radiation	1416	3032.5	1713	15.2	186.7
web	528	1130.7	572	5.1	185.8
anesthetic	86	184.2	0	0	185.2
brachytherapy	85	182	0	0	183
scanner	366	783.8	370	3.3	182.8
carotid	100	214.2	28	0.2	172.2
ablation	119	254.8	55	0.5	171.8
intravenous	339	726	367	3.3	170.4
scans	190	406.9	163	1.5	166.4
reviewed	1077	2306.5	1466	13	164.2
jewelry	83	177.7	11	0.1	162.8
clots	99	212	39	0.3	158.1
barium	171	366.2	153	1.4	155.5

In Table 8 below, we see the top 50 keywords when using the general Radiography corpus as a reference corpus.

Table 8 Top 50 keywords using the general radiography corpus

Keyword	***Freq***	***Freq/mill***	***Freq/ref***	***Ref/mill***	***Score***
	Patient Information		*Radiography*		
leaflet	327	700.3	1	0.9	369.3
copied	147	314.8	0	0	315.8
warranty	130	278.4	0	0	279.4
warranties	130	278.4	0	0	279.4
radiotracer	125	267.7	0	0	268.7
copyright	1015	2173.7	8	7.2	265.5
interprets	94	201.3	0	0	202.3
breastfeeding	87	186.3	0	0	187.3
illustrative	131	280.5	1	0.9	148.3
copyrighted	131	280.5	1	0.9	148.3
representations	130	278.4	1	0.9	147.1
download	130	278.4	1	0.9	147.1
fibroid	59	126.4	0	0	127.4
cryotherapy	57	122.1	0	0	123.1
dye	156	334.1	2	1.8	119.8
thrombolysis	52	111.4	0	0	112.4

Keyword	***Freq***	***Freq/mill***	***Freq/ref***	***Ref/mill***	***Score***
	Patient Information			*Radiography*	
fibroids	49	104.9	0	0	105.9
prick	47	100.7	0	0	101.7
resume	89	190.6	1	0.9	100.9
pictures	256	548.2	5	4.5	100
prohibited	130	278.4	2	1.8	99.9
disclaimer	130	278.4	2	1.8	99.9
loose-fitting	46	98.5	0	0	99.5
pals	43	92.1	0	0	93.1
web	528	1130.7	13	11.7	89.2
warned	38	81.4	0	0	82.4
sonar	38	81.4	0	0	82.4
inaudible	36	77.1	0	0	78.1
enterography	36	77.1	0	0	78.1
transvaginal	32	68.5	0	0	69.5
television-like	32	68.5	0	0	69.5
piercings	31	66.4	0	0	67.4
chemoembolization	31	66.4	0	0	67.4
vertebroplasty	30	64.2	0	0	65.2

Keyword	***Freq***	***Freq/mill***	***Freq/ref***	***Ref/mill***	***Score***
	Patient Information			*Radiography*	
kyphoplasty	29	62.1	0	0	63.1
assure	134	287	4	3.6	62.7
numb	54	115.6	1	0.9	61.4
loaf	28	60	0	0	61
up-to-date	130	278.4	4	3.6	60.8
sonohysterography	27	57.8	0	0	58.8
transrectal	26	55.7	0	0	56.7
thinners	26	55.7	0	0	56.7
magnets	26	55.7	0	0	56.7
box-like	26	55.7	0	0	56.7
outweighs	48	102.8	1	0.9	54.7
urogenital	25	53.5	0	0	54.5
breastfeed	25	53.5	0	0	54.5
please	758	1623.3	33	29.7	53
aneurysms	68	145.6	2	1.8	52.4
sting	24	51.4	0	0	52.4

4.4.2 Semantic classification

Organising the keywords into semantic categories was the next step. As I have explained in the preceding chapter, semantic classification helps reveal the primary concerns of the register, as themes can be more readily uncovered. I selected categories for each item in the top 50 in both lists, basing my decision on the broader meanings of each word and, as I have explained in detail in 3.5.1.9, using the Sketch Engine facility Word Sketch, along with examination of the item in the context of the corpus. Word Sketch gives the user collocational information, both lexical and grammatical, about a word. This collocational information references the corpus, it should be noted.

Both analyses resulted in 10 categories. Seven of these categories were shared and three categories unique. There were no items that appeared in more than one category. Examination of each word in its context revealed that two items, *up-to-date* and *reviewed,* were being used in a way that was not immediately obvious. Both items were moved from the category Information to the category Legal, after inspection in the corpus. All examples of both *reviewed* and *up-to-date*, while referring to the information contained in the leaflet, were used in formulaic phrases relating to a legal disclaimer.

As expected, the analysis with the BNC produced three categories that were more specific to radiography and radiographic procedures: *Medical instrument/equipment*; *Radiographic modality* and *Medical exam/procedure*.

In contrast, when the Radiography corpus was used as a reference corpus, there were no keywords in the top 50 from these three categories, but three new categories were created: Other: Body; Other: NHS and General. Other: Body contains items that could not be classed as body parts, but were related, nonetheless, to the human body, while Other: NHS contains just one item, an acronym that references a service offered within the NHS called PALS - the Patient and Liaison Advisory Service. *General* contains any items that were general in meaning.

The categories in common were Medical professionals; Body part or organ; Treatment or therapy; Disease or condition; Radiography or radiotherapy; Information and Legal, though as we see in Table 9, the majority of the words appearing in these categories were unique to the analysis. Words in common are bolded

Table 9 Semantic classification of keywords

BNC as ref corpus	**Category**	**Radiography ref corpus**
radiologist, radiologists, physician, radiographer, physicians, technologist,	***Medical professional***	-
clot, prostate, carotid,	***Body***	urogenital; breastfeed; breastfeeding, transvaginal, transrectal
medications, embolization, noninvasive, brachytherapy, ablation, interventional	***Treatment/therapy***	cryotherapy, kyphoplasty, chemoembolization, vertebroplasty, thinners, thrombolysis
catheter, transducer, scanner	***Medical instrument or equipment***	
CT, x-ray, x-rays, ultrasound, MRI, tomography,	***Radiographic modality***	Inaudible, sonar, loaf, magnets, box-like, television-like

BNC as ref corpus	Category	Radiography ref corpus
scan, angiography, mammography, exam, scans, barium	***Medical exam***	enterography, sonohysterography,
tumor, tumors	***Disease/condition***	fibroid, fibroids, aneurysms,
radiology, imaging, radiological, radiologic, radiotracer, radiofrequency, ionizing, radiation,	***Radiography and radiotherapy***	radiotracer, dye, pictures
download, **web**	***Information***	leaflet, **download**, **web** (site),
copyrighted, **copyright** reviewed	***Legal***	copied, warranty, warranties, **copyright**, interprets, illustrative, copyrighted, representations, prohibited, disclaimer, up-to-date, warned, assure
anesthesia, sedation, anesthetic, intravenous	***Medical: other***	numb, sting, prick
	Other: NHS/Healthcare system	PALS
jewelry	***General***	please, resume, outweighs, loose-fitting, piercings

The results of the semantic classification seen in table form gives an insight into the most important themes in patient information. We see that the technology of radiography is prominent, along with a variety of radiographic examinations. The professionals who work in radiology feature, as are treatment and therapy. An unexpected category is Legal. Without examining these words in the corpus, however, we cannot say much about how they are being used in the patient information. The results of these investigations will be presented in more detail in 4.5.3, though before this, and in the next section, I will present results relating to the class of words, i.e. parts of speech appearing in the keyword list.

4.4.3 Word type or parts of speech

We know that spoken registers are very often fundamentally different from written registers in their use of grammatical and lexical features (Biber, Conrad, Reppen, Byrd & Helt, 2002). There is more repetition in spoken discourse, with a lower lexical density than in written language (McCarthy, 1998). We have seen, too, in chapter 3, that the structure of lexical bundles differs markedly between conversation and academic discourse (Biber et al. 1999; Conrad & Biber, 2004), with academic discourse preferring noun phrases, prepositional phrases and passive structures, while conversation tends to make use of clauses, which centre around a verb. A full list of the common structure-types in academic discourse described by Biber et al. (1999) can be found in table 3 in section 3.5.2.6.

As I explained in 3.5.1.2, keywords can be lexical words, grammatical words or a combination of the two. The so-call smoothing parameter in Sketch Engine allows a user to decide whether they want more or less common words (common words are likely to include a high proportion of grammatical words) in their keyword list. Even with a high proportion of content words - which is what I wanted in this doctoral study - there will be different proportions of nouns, verbs, adjectives and adverbs. What class of words appear frequently may give us an insight into the structure of patient information, e.g. whether patient information likes naming things, and thus nouns will predominate, whether the focus is on doing and action, in which case there might be a large number of verbs or whether describing things is particularly prominent, in which case we might see a higher frequency adjectives and adverbs.

In my data, sometimes a word was used as a noun and a verb: in these few cases (e.g. *drink*; *visit*) the most predominant use in the corpus was the one selected. I could have marked these words as belonging to both categories, but in both cases, the verb was overwhelmingly preferred, so I chose to list it solely in this category in the patient information. In the section that follows, I will present the results of the word classes of the keywords.

4.4.3.1 Nouns

The keywords extracted in the first analysis, with the BNC as a reference corpus, are predominantly nouns. There are 41 nouns of various types (singular, plural, uncountable). The categories with the most nouns are Medical professional with 6; Treatment or therapy with 6. Radiographic modalities are also very present with 6 nouns that name the different radiographic technologies, while a further 5 nouns relate to the fields of radiology, radiography and radiotherapy. Examinations that use radiography account for another 6 nouns.

There is a greater variety of word type in the second analysis, with the corpus of Radiography as the reference corpus. The results included 25 nouns of various types (singular, plural and uncountable), around half of the number extracted using the BNC as a reference corpus. There are only two nouns in common: *web* and *copyright.* As *copyright* can be a noun, adjective and verb, I used Word Sketch to give me the parts of speech information I needed: 90% of the uses of *copyright* in the corpus are as a noun.

4.4.3.2 Verbs

There are 2 verbs included in the first 50 keywords with the BNC as reference corpus: *download* and *reviewed*. This is likely to be explained by what has been reported in the preceding section: the more noun-dense a text is, the correspondingly less verb-dense (and less pronoun-dense) it will be (Biber, 1988).

The second analysis with the radiography corpus does include verbs, however: there are 10 verbs (in all forms) in the list of keywords: *copied, reviewed, interprets,* *assure, resume, warned, breastfeeding, breastfeed, outweigh* and *download.* The first four verbs are categorised as legal verbs.

4.4.3.3 Adjectives

There were 7 adjectives in the first 50 keywords with the BNC as a reference corpus. In contrast, the analysis with the radiography corpus included 13 adjectives: *informational, illustrative, up-to-date, copyrighted, prohibited, loose-fitting, numb, inaudible, box-like, television-like, transrectal, transvaginal* and *urogenital.* Four of these seemed to relate to the patient information itself: *informational, illustrative, up-to-date and reviewed* though on closer inspection the first three were used in a legal disclaimer. Two further adjectives related clearly to legal issues: *copyrighted, prohibited.* This was by far the largest category in the second keyword analysis using the radiography reference corpus. It will be discussed in more detail in section 4.5.3. The distribution of the different classes of word in both keyword lists (with BNC as reference corpus and with the radiography corpus, can be seen in Table 10 below.

Table 10 Distribution by word class in keyword lists

Word Class	BNC Keyword list	Radiography Keyword list
Noun (all)	41	25
Verb (all)	2	10
Adjective (all)	7	13
Other	0	2
Total	**50**	**50**

The different range of word class seen in Table 9 above suggests that using a specialised corpus can indeed give valuable insights into the lexical characteristics of patient information that using a general corpus cannot. When the BNC was used, the technical and medical nouns that are specific to radiography predominated; they are classed as keywords because they are very rare in general English (and in any other variety of English except radiography and radiology). The second extraction produced keywords that, while sharing many of the same categories with the keywords extracted in the first analysis, were more varied in meaning and word class.

4.4.4 First 10 Keywords

We can also present a snapshot of the results of the keyword extraction by focusing on the first 10 keywords from each extraction. This snapshot gives us some important information at a glance about the prevalence of certain ideas in the corpus of patient information. I have listed these keywords again for convenience below in Table 11.

Table 11 1st 10 keywords with both BNC and radiography reference corpora

1st 10 with BNC	1st 10 with Radiography corpus
x-ray	leaflet
catheter	copied
transducer	warranty
physician	warranties
medications	radiotracer
radiological	copyright
scan	interprets
x-rays	breastfeeding
radiologists	illustrative
radiologic	copyrighted

It is worth remembering at this point that the relevance or otherwise of a statistically-significant keyword cannot be assumed by its position in a keyword list. As Scott (2010) says, ‘the greater the number of keywords extracted, the greater the possibility that the inclusion of some is based on statistical chance’. Nonetheless, focusing on just 10 words may make it easier to see themes emerging.

In Table 10, we can see that words relating to radiography and medicine predominate in the extraction with the BNC: *x-ray, x-rays* and *scan* are present (*scan* refers to a CT or MRI exam (the machines themselves are *scanners*; the exam a *scan*)). There are two medical professionals: the *radiologist*, who is a specialist doctor and not to be confused with *radiographer*, the person who performs the scan or x-ray, and *physician*, the American term for doctor. *Catheter*, *transducer* and *medications* also appear: that these terms are considered key is not overly surprising as one of the primary purposes of patient information is to explain radiologic procedures such as x-ray, ultrasound and CT. These diagnostic tests also happen to be the most commonly performed radiographic tests in the UK, in that order.[4] MRI is performed just a little less frequently than CT (in the UK) (0.26 million compared to 0.38 million in March 2016) and appears as the 11th keyword in this analysis.

Turning now to the extraction with the Radiography corpus, there are no words in common. The first word is *leaflet*. There is also a reference to *breastfeeding*, which relates to the safety of radiation-examinations for breastfeeding mothers and *radiotracer*, a reference to nuclear radiation. The remaining items in this short list all relate to the Legal category, which it transpires contains more keywords than any other category in the extraction with the Radiography corpus. The significance of this will be discussed in 4.6.1 below.

As we know, however, a keyword in a list tells us little, either about its use in a register or about its significance - its 'keyness' - in the register. It is this qualitative investigation of the words in our semantic categories, which includes collocation information, that gives us a picture of a keyword's significance in a register. It is these investigations, and their findings, that are the focus of the Discussion section that follows.

[4] https://www.england.nhs.uk/statistics/wp-content/uploads/sites/2/2015/08/Provisional-Monthly-Diagnostic-Imaging-Dataset-Statistics-2016-07-21.pdf

4.5 Discussion

The analysis with the BNC as the reference corpus results in many keywords, as we saw in Table 8, that are classified as medical and technical words, with many items that are strictly related to radiography, or to healthcare. Modalities, i.e. the type of radiographic technology; machines, examinations, along with some medical professionals and some treatment options.

When the analysis is conducted with the Radiography corpus as the reference corpus, the resulting categories are similar - there are 10 categories in common - but the keywords are usually different. In fact, there are only 3 keywords, out of 50, in common*: reviewed; commercial, copyright*. The semantic categorisation of all 50 keywords from both analyses has been presented earlier in Table 10. The finding that the categories are shared but the words they contain are different confirms, I feel, both the utility of the keyword method in highlighting the themes and areas of interest in a register, but also the value of conducting a second keyword extraction, in my case with the Radiography corpus, as it appears to have revealed a semantically richer variety of keywords.

The categories and some of the keywords that they contain will now be discussed in detail below. I have chosen the categories and keywords that seemed to me to represent particularly interesting implications for the register under investigation, that of patient information.

4.5.1 Legal

In the analysis with the BNC, there were three items in this category: *copyright*, *copyrighted* and *reviewed,* which always appear in a disclaimer. These words both reference and delimit, what can be done with the information.

(1) *Permission is granted to modify and/or reproduce this leaflet for purposes relating to the improvement of healthcare, provided that the source is acknowledged and that none of the material is used for* ***commercial*** *gain.*

There was just one example (out of 130) of *copyrighted* being used in the negative, and encouragement given to reproduce the information:

(2) *This publication is not copyrighted. The Clearinghouse encourages users of this booklet to duplicate and distribute as many copies as desired.*

Legal was by far the largest category in the analysis with the Radiography reference corpus, containing 12 words: *copied, warranty, warranties, copyright, copyrighted, prohibited, disclaimer, representations, assure, illustrative, reviewed,* and *up-to-date*. The difference in the variety of English was stark: 90% of occurrences of the words in the legal category were found in the US-sourced information. While some of these words are overtly legal, e.g. *warranty*, others, e.g. *up-to-date* or *illustrative* are general terms used with a legal meaning in this context:

(3) *Images may be shown for **illustrative** purposes. Do not attempt to draw conclusions or make diagnoses by comparing these images to other medical images, particularly your own.*

Many of these legal words appeared together with other legal keywords and thus were very formulaic in nature:

(4) *However, it is not possible to **assure** that this Web site contains complete, **up-to-date** information on any particular subject. Therefore, ACR and RSNA make no **representations** or **warranties** about the suitability of this information for use for any particular purpose.*

As we have seen, most of the legal terms came from the patient information sourced in the US. The appearance of legal disclaimers in healthcare information may seem inappropriate to those of us familiar with a different kind of health system (one that does not necessitate paying for treatment via private insurance), though given that half of all medical malpractice claims in the US relate to diagnosis[5] and treatment, and that malpractice pay-outs were valued at more than $3.7 billion in 2013 (and growing year on year) (ASC Communications, 2018), it is understandable that the producers of patient information wish to protect themselves. The legal references in my corpus

[5] https://www.beckershospitalreview.com/legal-regulatory-issues/medical-malpractice-in-america-15-latest-statistics.html

related not only to the advice and information itself but also to its commercial distribution.

In my data, encouraging communication with the patient's provider is a feature of many of the disclaimers that appeared at the end of the information, even when the word *disclaimer* is not used (as it is in much of the US information, suggesting that its inclusion is necessary by law). Patients are urged not to use the information in the leaflet to make treatment decisions and that the information is not intended to replace a visit to a doctor.

(5) *This leaflet tells you about the procedure known as nephrostogram. It explains what is involved and what the possible risks are. It is not meant to be a substitute for informed discussion between you and your doctor, but can act as a starting point for such a discussion.*

The encouragement to communicate with a healthcare provider may be falling on deaf ears, however, as studies suggest (e.g. Diaz, Griffith, Ng, Reinert, Friedmann & Moulton, 2002; Silver, 2015; Yeo, 2016) that more than 50% of patients who use healthcare information found online do not discuss the information with their provider.

Other studies find that accessing healthcare information online may even, in some cases, even lead to fewer subsequent doctor’s visits (Shim, Ailshire, Zelinski & Crimmins, 2018). While many studies concern themselves with online information, known hereafter as e-health (Eysenbach, 2001), their findings may also be relevant to the kind of information that hospitals and medical services produce, much of which ends up online on the hospital website, in addition to being made available as a printed leaflet. All of the documents contained in my corpus were available to be printed off at home, and many documents seemed to replicate what would also be available in a hospital or GP surgery.

In a relatively early study on e-health, Diaz et al., (2002) found that 59% of the respondents did not discuss information found online with their doctors. Patients who did discuss internet information with their medical provider rated online information as more reliable than those who did not, and between 50-60% of all patients surveyed felt that information recommended by a doctor or on a hospital site was more trustworthy than that from a non-profit organisation (30%) or a site sponsored by a pharmaceutical company (16%) (p183). 60% of the respondents in this early study rated online information as good as, or even better than that from their doctor.

A more recent study (Waring, McManus, Amante, Darling & Kiefe, 2018) found an even lower percentage of patients, just over 33%, discussed their online health information with their providers, suggesting that as we become more familiar with the internet, we may use information more independently and feel less inclined to share it with healthcare providers. This, and other studies, find that a higher education level, higher socioeconomic status and frequency of online-information seeking is associated with provider discussions (Graffigna, 2017; Waring et al., 2018). Expressed another way, the less educated you are, the lower your income and the less familiar you are with e-health searching, the less likely you are to discuss any information you may find online with your healthcare provider, in spite of the encouragement in patient information to do so. Clearly, in this scenario, there is a negative impact on shared-decision making which is defined as ‘an approach where clinicians and patients share the best available evidence when faced with the task of making decisions, and where patients are supported to consider options, to achieve informed preferences’ (Elwyn et al., 2010).

What are the factors that explain what appears to be an increasing number of patients choosing not to discuss information with their doctor? It seems the reasons behind this decision are complex and varied but can include feeling embarrassed, feeling that the doctor does not want to hear about it, believing there is no need to talk about it and simply forgetting (Silver, 2015). As we have already seen, the availability of printed or digital patient information is central to the policy of shared decision making; it is ironic, then, that the very people who would benefit from discussing information with their doctors are less likely to do so.

Let us return for a moment to the characteristics of ‘legalese'. Could the language of legal disclaimers be off-putting for readers? My personal response to the US legal disclaimer was negative, though this may be because I am not used to seeing overtly legal text in healthcare information. However, the differences between the insurance-based medical systems in the US and the NHS in UK, free-at-the-point-of-care, may lead some readers to the conclusion that legal disclaimers are unlikely to be a feature of UK written patient information. My research tells me that disclaimers are included if the writer considers it necessary in the UK, but there is currently no obligation to do so.

A 2018 post in the ‘Patient Information Forum', a respected association for professional patient information writers in the UK, in response to a question regarding the necessity of writing a disclaimer on patient materials, was:

> I do include disclaimers when we reference a third-party website as we have no control over the content they contain. In our own information however, we don’t include formal disclaimers but do often state that the leaflet isn’t meant to replace discussion with a healthcare professional and also include wording to encourage patients to contact us (or NHS24 or their GP if appropriate) if they have questions and concerns. (Thomson, Patient Information Forum, 2018)

The writer worked for the NHS. Another respondent, also working for an NHS hospital Trust, included a link to their disclaimer, developed after consultation with the Trust’s solicitors, and which I reproduce in part below:

> Please note, while the information contained in the leaflets has been created, reviewed and checked by our medical and surgical teams, the information is designed to complement the advice of professional healthcare staff. The leaflets should not be used on their own without appropriate medical advice. Procedures described should only be undertaken after you have received training by healthcare professionals and the Trust will not be liable for injury or loss incurred as a result of actions taken by individuals after reading the materials.
> While every effort is made to ensure that the information contained in these leaflets and web site remains up to date, from time to time clinical advice or the management of clinical situations may change with new medical practice/knowledge. While we try to update our information promptly, we must emphasise the need for you to seek advice from a healthcare professional about your particular situation.
> Care has been taken to describe the treatments, conditions and risks associated with treatment in a sensitive manner, however, due to their nature you may find some of the content distressing. (Glaister, Patient Information Forum, 2018)

The passage above (which I must make clear is not included in my corpus of patient information) contains a number of the keywords that were also extracted in my analysis of my corpus: *treatment*, *healthcare*, *web*, *up-to-date*, *leaflet* and *reviewed.* The legal disclaimer drives home the point regarding advice and involvement from a professional, as we see in the following four sentences that express the idea:

(6) *…the information is designed to complement the advice of professional healthcare staff.*

(7) *The leaflets should not be used on their own without appropriate medical advice.*

(8) *Procedures described should only be undertaken after you have received training by healthcare professionals…*

(9) *…we must emphasise the need for you to seek advice from a healthcare professional about your particular situation.*

There are also two sentences that are more legal in tone and thus more formulaic:

(10) *…the Trust will not be liable for injury or loss incurred as a result of actions taken by individuals after reading the materials*

(11) *All information is copyright and must not be adapted or reproduced without permission*

That printed information is not to intended to replace information from a medical provider (spoken or written) is clearly of great relevance in the NHS: nearly half of the sentences in the disclaimer above of 11 sentences reinforce this idea.

But is legal language appropriate in healthcare communication? From the point of view of comprehensibility, it seems in conflict with the advice to write in plain English, to avoid jargon and complex language. I am not aware of any studies that have surveyed a patient’s responses to the legal information in many documents, though I have presented studies above that show that an increasing majority of patients who have sourced information online choose not to speak with their provider about their findings, which suggests that the encouragement to do so - often expressed in these legal disclaimers - is not successful. We do not know if this is because patients are actively ignoring the legal disclaimers and the encouragement to seek further information from their doctor or, in fact, because they have not read or understood this part of the document.

On the other hand, legal action against hospitals and doctors is big business: we have already seen that in the US, medical insurance claims reached $3.7 billion in 2013, but even free-at-the-point-of care medical systems, like the NHS, face increasing claims against it from patients. In just six years, the cost of negligence claims paid out by the NHS Litigation Authority has risen from £0.6 billion in 2006/7 to £18.9 billion of outstanding liabilities in 2012. (NHS Litigation Authority Factsheet, Factsheet 2, 2012). Many of these litigation cases concern clinical negligence, i.e. errors made during an operation or treatment, though the Citizen’s Advice Bureau, a consumer advice agency in the United Kingdom, estimates that communication failure lies behind 20% of these claims. An example of a recent legal case brought against an NHS hospital by a patient that relates specifically to the provision of information is reported below:

> Following an assault to his head, the claimant attended the A&E department of Mayday Hospital, accompanied by a friend. He was booked in by a receptionist at 20.26 and was told that it would be up to four or five hours before he would be seen. He was not informed that a triage nurse would examine him within 30 minutes and determine how soon he needed to see a doctor. As he was in pain he decided to go home after just 19 minutes and take paracetamol. Unfortunately, his condition rapidly worsened. He returned to hospital by ambulance and it was discovered that he had an extradural haematoma, but too late to prevent serious brain injury. The experts for the parties agreed that if the claimant had remained in A&E he would have been treated sufficiently quickly to have avoided the brain damage. The essence of this claim was that the A&E receptionist owed the claimant a duty of care to give him accurate information about waiting times. The trial judge had accepted Mr D’s assertion that had he been told that he would be seen by a nurse within 30 minutes, he would have stayed in hospital and therefore avoided his permanent injury. Lord Justice Jackson gave the main judgment. He said that this case was significant because roughly 100,000 people visit A&E departments across England every week. He was satisfied that there was no duty upon receptionists to keep patients informed about waiting times. It would not be fair, just or reasonable in his view to impose liability in such circumstances. Were this type of claim to be permitted, litigation about who said what and to whom in A&E could become prevalent. Trusts might then instruct receptionists to say nothing to patients other than ask for their details, which would be unhelpful. The claimant had been told to wait but chose not to do so. People had to accept responsibility for their actions in his opinion. (NHS Resolution, 2017)

While this extract relates to spoken information, it also has relevance for the role of written information. NHS hospitals, authorities and trusts have to strike the right balance between providing information to patients as a fundamental aspect of patient-centred care, while, at the same time, covering themselves against misinterpretation, misunderstanding or misuse of that information.

In the US, legal disclaimers are very visible and appeared in every US information leaflet in the corpus. The disclaimer always appears at the end of the document, as they do in the UK-sourced information, and this fact makes it worth reflecting for a moment on the question of sampling, discussed in 3.1.4. If my patient information documents had been subject to sampling, that is, only a section of each document used - and it is generally the middle - then I would have remained unaware of the legal information that is so evident in much healthcare material. The keywords that led me to this hugely significant area would have been lost.

Let us now move from the category Legal to another category of interest, the Medical professional.

4.5.2 Medical Professional

There are six words for medical professionals in our keyword list, and all 6 revealed by the analysis with the BNC: *radiologist, radiologists, physician, physicians, technologist* and *radiographer*. *Technologist* and *radiographer* are synonymous, the former is the American term for the latter. *Radiologist* is the medical specialist who diagnoses and interprets radiographic images. When we look closely at the frequency of these keywords in the corpus, along with the frequency of two other key health professionals working in a radiography department, a nurse and doctor, an interesting picture emerges, as illustrated in Table 12 below.

Table 12 Frequency data for Medical professionals in patient information

Professional	**Raw freq.**	**Adjusted freq. per million**
radiologist	893	1,912.4
physician	1388	2,927.5
doctor	914	1,957.3
radiographer	181	387.7
nurse	234	501.1

We can see that the two jobs that occur with the lowest frequency are *nurse* and *radiographer*. *Doctor* and *radiologist* have similar adjusted frequencies while *physician* appears nearly 3 thousand times per million words. This is a word used with great frequency in patient information and is nearly six times as frequent as *nurse*. *Radiologist* occurs with nearly 5 times the frequency of *radiographer*.

These results strongly suggest that there is an underlying focus on the professional who diagnoses and treats: the *doctor*, the *physician*, the *radiologist*. This imbalance is ironic when one considers that a patient attending hospital for radiography may not have any contact at all with the radiologist. The person they will likely have most contact with is the radiographer (aka technologist) and quite possibly a nurse. This privileging of roles in patient information - if that is what it is - seems to mirror the hierarchy seen in medicine, where doctors have more prestige and power than nurses and other allied health professionals.

The over-emphasis on ‘the doctor’ is also seen in healthcare research: Candlin and Candlin (2003), discussed in 2.4.1, refer to the fact that numerous studies of communication by nurses have been carried out by nursing professionals and published in nursing journals but these studies, unlike studies of doctors, are not referenced in applied linguistics or discourse analytic studies (p144). While recently there has been a little more visibility of nursing communication studies in mainstream journals such as the *English for Specific Purposes* journal (e.g. Lu, 2018; Staples, 2015; Bosher and Stocker, 2015), the number remains small. This finding, along with the large number of keywords from both extractions that relate to treatment and therapy - there are 13 of them - is particularly significant when we consider gender differences in healthcare information-seeking behaviour and linguistic expression of disease and health discussed earlier in this chapter and in chapter 2.

Seale and Charteris-Black (2008), as discussed in 2.5.4, found that men in general, and older men in particular, were significantly more likely to refer to specialists, general practitioners and consultants than were women. The professionals named significantly more frequently in my corpus are specialists and general practitioners. *Doctor* was used by men more or less equally in Seale and Charteris-Black (2008), suggesting that men, but particularly older men, give particular importance to the specialist knowledge of the medical professionals they interact with. *Nurse*, in contrast, was referred to significantly more often by women; in my data, as we have seen, both *nurse* and *radiographer* are referred to far less frequently than the *radiologist*, *physician* and *doctors*.

This is not the only evidence that the patient information may contain information that is more relevant to men. Seale et al.’s (2006) finding, discussed in 2.5.3, that men tend to focus on information relating to treatment, to medical staff and to medical procedures is very relevant to my findings in this doctoral study. The semantic categories in my keyword analysis were predominantly those that Seale et al. (2006) highlight as likely to be of more interest to men: Medical staff, Treatment & therapy, Medical equipment, Radiographic procedures and Examinations and radiographic technology.

If women (and men) are (also) looking for information or confirmation in the patient information leaflets of their emotional or mental responses to their looming radiographic examination they will be disappointed: all modifiers of the verb *feel* in my corpus relate to physical or physiological sensations. e.g. *pain, pressure, discomfort, prick, tired, warm, unwell.* There were six uses of the adjective *anxious,* and three of *depressed* (and two of these from the same document), but no uses of typical expressions of fear or distress that we expect to find in a healthcare information leaflet relating to a procedure that has such a close relationship with cancer, e.g. worried, scared, frightened, concerned, sad or tearful.

In spite of these well-reported differences, gender is very rarely considered to be a determinant of healthcare, and, as we have seen, rarely appears as such in health policy documents (Gelb et al. 2011). The question is raised, then, of whether gender is considered by healthcare information writers.

Healthcare communication writers employed by large charities and those working for the NHS that were contacted during this research for their views on gender and healthcare information either assumed I was referring to transgender (sometimes referred to in their communication as 'third gender') and/or said they did not consider gender at all in the production of their materials:

> We (I) certainly don't consider gender when writing materials, in the sense of writing "for" one particular gender over another. I also had a look through our brand guidelines but there is nothing specific about gender when it comes to our [*name of charity*] tone of voice. We aim to be "inspiring, authentic, confident, frank and human" in all our comms and these values apply across gender boundaries. (S.Newton, personal communication, August 24, 2018)

The charity the respondent worked for focused on diabetes, a very common health condition and one which research consistently shows disproportionately affects women. An editorial in the Lancet Diabetes and Endocrinology (Editorial, 2017) states that in Western countries, not only do fewer women than men receive the level of treatment outlined in healthcare guidelines but woman suffering the Type 1 variety have a 40% higher risk of premature death.

There are similar findings of disparity in healthcare and prognosis for a large number of common conditions (Legato, Johnson & Manson, 2016) including heart disease (Westerman & Wenger, 2016), certain cancers (Williams et al., 2017; Yuan et al., 2016), kidney disease (Jindal, Ryan, Sajjid, Murthy, Baines, 2005). Regitz-Zagrosek (2012, p. 596) writes that the scientific literature contains 'more than 10,000 articles [that] deal with sex and gender differences in clinical medicine, epidemiology, pathophysiology, clinical manifestations, outcomes and management'. Table 13 below illustrates the sheer volume of papers - which are likely to have to increased since 2012 - focussing on sex and gender differences.

Table 13 Publications with sex and gender differences in the most frequent clinical entities (Regitz-Zagrosek, 2012)

Cardiology	Rheumatology/ Immunology	Pneumology	Nephrology	Gastro-enterology/ Hepatology	Neurology	Endo-crinology	Oncology	Haematology
Hypertension (414)	Lupus erythematosus (68)	Asthma (140)	Renal failure (27)	Hepatitis B (22)	Multiple sclerosis (65)	Diabetes mellitus (447)	Skin carcinoma (45)	Anaemia (44)
Myocardial infarction (275)	Rheumatoid arthritis (41)	Lung cancer (116)	Diabetic nephropathy (11)	Hepatitis C (26)	Stroke (129)	Obesity (349)	Gastric cancer (25)	Leukaemia (49)
Heart failure (153)	Systemic sclerosis (3)	Chronic obstructive pulmonary disease (36)	Glomerulone--nephritis (9)	Hepato-cellular carcinoma (37)	Alzheimer's disease (104)	Osteo-porosis (123)	Renal cell carcinoma (17)	Lymphoma (34)
Atrial fibrillation (38)	Fibromyalgia (15)	Pulmonary hypertension (12)	Polycystic kidney disease (12)	Inflammatory bowel disease (13)	Epilepsy (56)	Hypo-thyreoidsm (33)	Bladder cancer (22)	Thrombo-cytopoenia (6)
Coronary heart disease (207)	Sjögren's syndrome	Pulmonary embolism (110)	Renal artery stenosis (0)	Colorectoral cancer (24)	Parkinson's disease (69)	Hyper-thyreoidsm (16)	Thyroid carcinoma (16)	Purpura (2)
Cardiomyopathy (41)	Ankylosing sponylitis (11)	Sarcoidosis (6)	IgA Nephropathy (2)	Autoimmune Hepatitis (2)	Muscular dystrophy (11)	Morbus Addison/ Cushing disease (5)	Pancreatic carcinoma (10)	Agranulo-cytosis (0)

Numbers in brackets refer to the number of publications.

It is misguided, then, to believe that being human trumps sex and gender differences in health when the evidence from the scientific literature is clear. This evidence suggests strongly that biological sex and gender have a relationship with the manifestations of, experiences of and outcomes of disease. This, in turn, suggests that a one-size-all approach to healthcare information is not only inappropriate but may also result in information that is unwittingly gender- and age-biased: my examination of the keywords in patient information for radiography strongly suggests that the content prioritises information that is likely to be of particular interest to older men. Without acknowledgement of these differences, and without data that can inform public health campaigns and patient materials, there is a real danger of health messages not being transmitted.

The increasing numbers of people who use internet forums for healthcare advice may, in part, be explained by the need to find information other than that which is published or presented officially. In a study of internet forums relating to chronic cough (Sinha, Porter & Wilson, 2018), traditional medical advice was sought and given (and judged to be of good quality by raters), along with emotional support for the psychological stress associated with the condition. Of note is the attitude towards medical consultations:

> Chronic cough is a condition with which patients often visit their doctor multiple times. Our data show forum users avoiding doctors' appointments after bad experiences or lack of effective treatments, citing them as a waste of time. These patients are lost to follow-up in the medical system, but may frequent online health forums, seeking advice from other sources. (Sinha et al., 2018, n.p)

We have already seen in Chapter 2 that the 'biomedical discourse' that Dixon (2002) refers to is a feature of much pharmaceutical patient information. My keyword extraction reported in this chapter strongly suggests that it is also the primary discourse in procedural patient information. 'Bio-technical' is also an appropriate term. The appropriacy of this discourse for all patients, and whether it provides the information all patients would like to have is an area worth further investigation.

Let us now turn to another category that Seale et al. (2006) found to be particularly interesting to male healthcare-information seekers, and one that falls under the heading of bio-medical discourse: Treatment and therapy.

4.5.3 Treatment/therapy

This was a large category of keywords with 6 words, predominantly nouns, in each list. There were no words in common, however. The generic term *medications* appears in the extraction with the BNC and *thinners*, a reference to blood thinners in the extraction using the Radiography corpus, while two descriptive adjectives for the type of medical procedure also appear: *interventional* and *non-invasive*. The rest of the words were specific therapies: *chemoembolization, brachytherapy, embolization, cryotherapy, kyphoplasty, vertebroplasty, thrombolysis* and *ablation.*

In many cases, these words appeared only in US materials and were described in fairly complex terms. There is one example of *cryotherapy* in the UK materials and it is interesting to compare the comparative clarity of the description with that found in an US leaflet, which is more technical and lexically complex:

(12) ***Cryotherapy** is an alternative technique that freezes tissues instead of burning them. It involves insertion of small needles (cryoprobes) through the skin, which circulate very cold gas and freeze the tumour by producing ice. (UK)*

(13) *During **cryotherapy**, liquid nitrogen or argon gas flows into a needle-like applicator (a cryoprobe) creating intense cold that is placed in contact to diseased tissue. (US)*

As I have said, the vast majority of the named therapies in the category were found in US materials. While it seems possible that UK patient information is avoiding obvious mention of something considered frightening and unpleasant, there is another, more likely, explanation: the majority of the 94 documents sourced from the NHS and the Royal College of Radiologists was information pertaining to diagnostic radiography or procedures involving radiography, such as angiography. Diagnostic radiography is that which is used to diagnose or rule out a disease. On the other hand, the 136 documents sourced from RadiologyInfo.org, also contained some information about therapeutic radiography, which is radiography used for the treatment of cancer. This partly explains the quantity of precise therapies listed.

Closer inspection of the corpus also revealed that a very small number of long, therapeutic documents were responsible for the appearance in the keyword list of a number of the keywords: *chemoembolization* (3 documents; 138 per million words); *embolization* (5 documents; 414 per million words); kyphoplasty (1 document; 62 per million words). The name of the specific therapy was repeated frequently in each document, though it is clear that these words are used quite idiosyncratically and only one word, *embolization*, is present in 5 documents. *Kyphoplasty* appears 29 times but only in one document. This finding underlines Scott's (2010) reminder that statistical chance (and not necessarily linguistic value) is a factor in keyword extraction (p50), and that keywords are useful for suggesting areas worth investigation, but they are not always, in themselves, of much value.

4.5.4 Information

Leaflet, *web* (site) and *download* were categorised under Communication, a heading that covers information from print and non-print sources, such as websites. *Download* and *web*(site) appeared together on RadiologyInfo.org's website and were used to tell patients to return to the site to check for, and download, updated or further information. *Leaflet* was also sometimes used to encourage patients to read more information:

(14) *If you would like more information about this, please ask a member of staff for* ***a leaflet*** *called What to do if the contrast injection leaks out (extravasation)*

With *this* as determiner, *leaflet* was also used to focus attention on the purpose of the information being read:

(15) ***This leaflet*** *will give you some general information about the clinic. If you have any further questions, please speak to a doctor or nurse caring for you*

(16) ***This leaflet*** *contains information on gadolinium (also known by its brand name Dotarem®), which is a contrast (dye) used during MRI scans*

There were also some examples of a hospital leaflet referencing their no-smoking policy; in these cases, the patient is instructed to read the *leaflet* with a polite imperative. In this directive, *our* is always used in place of *the* or *this*, perhaps to emphasise that this is a policy 'owned' by the entire institution or hospital (and thus to be taken seriously).

(17) *Please read our* ***leaflet*** *'Policy on Smoke Free NHS Premises' to find out more*

As we have seen in 4.5.1, *this leaflet* is also used as part of a disclaimer where the idea of its scope being limited is expressed:

(18) *Legal notice. Please remember that* ***this leaflet*** *is intended as general information only. It is not definitive, and the RCR and the BSIR cannot accept any legal liability arising from its use*

(19) *Some of your questions should have been answered by* ***this leaflet****, but remember that this is only a starting point for discussion about your treatment with the doctors looking after you.*

The message to patients seems to be: read our leaflets; they should answer most of your questions. If they do not answer your questions, read some more leaflets or look at another website. However, any information you read must be considered general information, not always specific to you and perhaps not even relevant to you. Communication with medical professionals, particularly the patient's doctor, is frequently encouraged and it is likely that this encouragement is motivated by legal concerns. Legal is the biggest category in the analysis, as we saw in 4.5.1, while disclaimers and legal statements in general are very common in the patient information collected for the corpus, even when presented in an indirect way. Subtle disclaimers such as the above are far more common in the UK-patient information, while more formulaic, legalese is used in the US-information.

4.5.5 General

There were a number of the words classed as General: *please*, *resume*, *outweighs, loose-fitting, piercings and jewelry.* The latter 3 are found in the instructions that patients are routinely given in preparation for an exam, while *resume* refers to the post-exam recovery period:

(20)   *You should wear comfortable, **loose-fitting** clothing for your ultrasound exam*

(21)   *We will ask you to remove all jewellery and body **piercings** before the scan as the scanner uses a very strong magnet*

(22)   ***Jewelry** and other accessories should be left at home if possible*

(22)   *You will be able to **resume** all other normal activities 8 to 12 hours after the exam*

Please also appears in this category and is used very frequently in the corpus: 759 occurrences which corresponds to a 1,625.45 per million words.   What is particularly interesting is that *please* is used exclusively with a verb in the imperative, e.g. *please ask your doctor for more information.* Some of these imperative structures functioned as invitations and offers, particularly of further information or advice,

(23) *If you have any questions or concerns,* ***please*** *do not hesitate to speak to a doctor or nurse caring for you*

though many of the structures are imperatives and are obliging the patient to do something:

(24) *Arriving at the clinic:* ***Please*** *report to the receptionist on arrival.,*

(25) ***Please*** *make sure that you understand the risks and benefits of the procedure and that it has been explained to you in the detail you need.*

(26) *If you are not able to attend,* ***please*** *let the department know in good time*

(27) *You will be asked to undress in a cubicle and you will be given a cotton gown to wear;* ***please*** *bring your own dressing gown.*

Many of the uses of *please* referenced further communication between the patient and the hospital or care provider, in common with *leaflet* and *web* as we saw earlier in section 4.5.4,

(28) *If you have a query,* ***please*** *ring your breast surgeon's secretary or a breast care nurse*

(29) *If you have any questions about the procedure* ***please*** *ask the doctor who has referred you for the test or the department which is going to perform it.*

Other uses were as part of a legal disclaimer, a category discussed earlier in this chapter in section 4.5.1,

(30) *All information is provided "as is" without express or implied warranty.* ***Please*** *visit the RadiologyInfo Web site at http://www.radiologyinfo.org to view or download the latest information.*

(31) ***Please*** *remember that this leaflet is intended as general information only. It is not definitive, and the RCR and the BSIR cannot accept any legal liability arising from its use.*

There were just seven uses (50 per million words) of *please* + not + verb; four of these examples invited patients to seek more information, while three of these also functioned as negative obligations,

(32). *If you have any questions or concerns,* ***please do not*** *hesitate to speak to a doctor or nurse caring for you.*

(33) ***Please do not*** *bring children with you to the department. This is to avoid exposing them to unnecessary radiation.*

In the corpus of patient information in this doctoral study, *please* was used with imperatives to give instructions but also to invite and encourage an exchange of communication. Table 14 below shows the 10 most frequent verbs collocated with *please* in order of descending collocational strength.

Table 14 Common collocates of please in patient information for Radiography.

Verb	Example of use in the corpus
1. visit	*for more information…please visit the KIC on the Ground floor*
2. contact	*if you have any questions or concerns about ablation, please contact…*
3. ask	*please ask your doctor for more information*
4. bring	*please bring an overnight bag with you to hospital*
5. remember	*please remember that this information is intended as general information only*
6. tell	*please tell us before the injection if you think you might be pregnant*
7. let	*please let us know if you are taking any antiplatelet medicines*
8. telephone	*please telephone xxxx to cancel or make changes to your appointment*
9. consult	*please consult with your physician as to whether or not you will be admitted*
10. read	*please read out leaflet: Policy on smoke-free premises*

Please collocates most strongly with *visit*, *contact* and *ask*. All three of these verbs are inviting and encouraging communication between the patient and the professional, which underlines how important an aspect this is in procedural patient information. The remaining collocates - the majority - are all instructing the patient to do something, suggesting that this is also an important function of patient procedural materials. I will pick this topic up again in Chapter 6, where I report the results of my analysis of modal verbs of obligation.

Before I conclude this chapter, I would like to refer not to a category, but a significant defining feature of patient information that has been clearly illuminated by the two keyword extractions, that of the use of patient-friendly vocabulary alongside complex medical vocabulary.

4.5.6. Patient-friendly vocabulary

The complexity of vocabulary in radiography, and in medicine in general, has been the focus of many studies, as we saw in Chapter 1. Studies repeatedly show that patients struggle to understand medical consultations (Chapman et al., 2014; O'Connell et al., 2014) and readability, as we have seen, is a constant concern. (e.g. Morony et al., 2015).

In my two keyword extractions, a total of 100 keywords were examined. Three categories are shown below in Table 15 which illustrate the kinds of words that are being used as patient-friendly terms, and some of their medical equivalents.

Table 15 Patient friendly keywords with their medical equivalents

BNC as ref. corpus	Category	Radiography as ref. corpus
CT, x-ray, x-rays, ultrasound, MRI, tomography,	***Radiographic modality***	Inaudible, sonar, loaf, magnets, box-like, television-like
anesthesia, sedation, anesthetic, intravenous	***Medical: other***	numb, sting, prick

In the category 'Radiographic modality', we have *magnet, box-like* and *television-like* and *loaf* which are all used to describe or explain aspects of a scanning machine to the patient. While *magnet* may be comprehensible if you know that MRI stands for magnetic resonance imaging, *box-like* and *television-like* are less so; without seeing the data, *loaf* seems entirely out of place in the context of radiography. Examining the words in context reveals how these words are used:

(34) *Some of the **magnets** used for MRIs are like narrow tunnels and others are more open.*

(35) *The CT scanner is typically a large, **box-like** machine with a hole, or short tunnel, in the center.*

(36) *The equipment typically used for this examination consists of a radiographic table, one or two x-ray tubes and **a television-like** monitor that is located in the examining room.*

(37) *CT imaging is sometimes compared to looking into a **loaf** of bread by cutting the **loaf** into thin slices*

We also have *inaudible* and *sonar* which are both used to refer to the ultrasound examination.

(38) *The transducer sends out **inaudible** high-frequency sound waves into the body and then listens for the returning echoes from the tissues in the body. The principles are similar to **sonar** used by boats and submarines*

Examples (36) and (37) are good examples of simplified medical language being anything but (37) is simply obscure in its imagery. When does cutting a loaf involve looking into it? (38) assumes vocabulary knowledge with *inaudible* and cultural knowledge with *sonar* that just cannot be assumed.

I would like to conclude this section by considering three further words that showed up our keyword list, *numb, sting, prick.* They appear in the category Medical: other and were listed when using the Radiography corpus as a reference corpus. They also very neatly appear alongside the medical procedures they are referencing (as do the keywords 32-36) which were listed as keywords with the BNC as reference corpus: *anesthesia, sedation, anesthetic, intravenous.* How these items are used in the corpus is illustrated below in 37-41.

(39) *… a local anaesthetic will be injected into your groin area. This will **sting** at first but will then **numb** the area so that you do not feel any pain.*

(40) *Your physician will **numb** the area with a local **anesthetic***

(41) *It may **sting** a little when the local anaesthetic is injected.*

(42) *You will feel a slight pin **prick** when the needle is inserted into your vein for the **intravenous** line (IV)*

(43). *Infants and young children usually require **sedation** or **anesthesia** to complete an MRI exam without moving*

Sedation was used only in the US material; (41) above is almost threatening and would, I feel, frighten many parents.

There is also a vocabulary comprehension issue in this sentence that Clerehan et al. (2005) raise, discussed in 2.6.1, which is that of presenting two terms as synonyms in the same sentences, separated by the word ‘or’, e.g. ‘Methotrexate may cause a reduction in the number of white cells **or** platelets in the blood’ can be doubly-confusing for patients if they are unfamiliar with either or both terms; it is also the case that sometimes the second word is a synonym but sometimes a new, additional word. How is the patient expected to know this? There are 39 examples of *sedation* used with *anesthesia* in this way in my corpus. Is sedation the same as anesthesia?

Out of a random sample of 10 sentences of content word + or + content word from the patient information corpus, I found 4 that could be taken to be synonymous. E.g. *sometimes a small plug or stitch is placed in the artery; other risks or complications include*… It is important to remember, too, that health literacy is not literacy. Zarcadoolas (2011), discussed in 2.6.2, refers to the variety of knowledge that we bring to the interpretation of a text: the social, cultural and environmental aspects of health literacy. We have to know already something about *sedation* and *anaesthesia* to be able to know whether these terms are synonyms; we need to know something about medical vocabulary to feel confident that *risks* means more or less the same thing as *complications*.

There are 6,685 (14,316.34 per million) *word* + *or* + *word* combinations in my corpus. Not only is this structure very frequent in healthcare materials, but we have seen that the two content words (often nouns) are only sometimes synonymous. Generally, no information is given in the text to help readers with this interpretation. The complexity of health literacy means that a lot of knowledge -and some of it is specialist, medical knowledge - is required by readers to be confident that they fully understand these sentences. The factors reported here strongly suggest that this is an aspect of healthcare materials that warrants further research.

4.6 Conclusions

My keyword analysis carried out with two different reference corpora uncovered a number of areas of linguistic interest.

An overarching theme in the corpus is the role of patient information in healthcare education, and, more particularly, the limits to that role. These limits are expressed in clear, legal disclaimers in the US-sourced information, while the UK materials prefer a linguistically softer approach that encourages the patient to not rely on anything they have read (as it may not be accurate or appropriate) and to speak about the information with their healthcare provider. As we have seen, however, considerably more than half of people who engage in ehealth information-seeking do not subsequently refer to their doctor to discuss it. We do not know the reasons for this though they seem to be various and complex. It is also the case that some legal disclaimers are long and linguistically complex, which may be very off-putting for readers (who may not read them at all).

On the other hand, disclaimers that try to be non-threatening and non-disclaimer-like may not be understood to even be disclaimers. Many people report trusting hospital-produced information and to rate it as highly as the information their doctor provides. That patients trust the printed information seems at odds with the message that information often contains (in the form of non-threatening disclaimer): that it is irrelevant or not up-to-date or appropriate, which is the message that much UK materials seem to transmit. It would be interesting to find out what the impact is of messages such as this in healthcare materials.

Another important theme that emerged was the reflection of the power hierarchy in the naming of professionals. Medical doctors and specialists (radiologists) were referred to far more often than the radiographers and technologists - the very personnel who are responsible for performing the radiographic examinations and therapeutic sessions - and considerably more often than nurses, who are often present for radiographic procedures. One explanation is the focus on the significance of diagnosis for the patient reading the material. As we have seen, most, if not all, radiographic examinations can be (and very often are) used for finding or excluding cancer and a diagnosis is not given by a radiographer or a nurse. While this may be the case, (and many appointments for radiographic exams do not relate to cancer), not referring by name to the healthcare professional performing the examination is a strange omission. Many patients will have no contact with a radiologist: the results of a radiographic examination will be sent to the referring doctor or physician.

The keyword analysis reported in this study also raised some very interesting questions about the kind of information being presented in procedural patient information. It is, in common with much pharmaceutical patient information, a discourse that is overwhelmingly biomedical and biotechnical. This may be the result of the technological nature of this branch of medicine, though it may be, also, that healthcare information materials habitually present the kind of information that only a small section of the population finds satisfying: older men. All age groups and both sexes undergo radiographic examinations, and it is possible that other kinds of information is wanted, information that is currently not present in the types of published materials that make up my corpus.

Another area of interest revealed by the keyword extraction relates to how information is presented in what we can refer to as ‘or' structures. There is an assumption on the part of many materials writers that readers have the capacity to judge whether the content words in these structures is additional or synonymous. Given that many of the words relate to medicine and radiography parsing the text demands a high level of health literacy. Understanding a word when it is used alone, in a clear context, may cause less problems than trying to decide whether two words, both of which you think you know, are synonymous. We return to the notion of understanding, raised by John Skelton in Chapter 1. The issue is not really whether we know what words mean, but how we understand them. Presenting two items in a sentence that are often very similar, but leaving it up to the reader to decide just how similar, seems an unnecessary complication in healthcare materials, which, after all, are striving for clarity.

We stay with questions of structure in the next chapter, in which I present the results of my analysis of 4-word lexical bundles in patient information for radiography.

5. Lexical Bundles in Patient Information

This chapter complements the keyword analysis described in the preceding chapter by reporting on an analysis of the frequency, distribution and discourse function of four-word lexical bundles in patient information for radiography. Both keywords and an analysis of lexical bundles can reveal aspects of the lexical characteristics and lexical patterns found in patient information.

Lexical bundles are multi-word lexical sequences that frequently reoccur in a register, e.g. *in the light of* and *at the end of*. They have been described as ‘characteristic features of language use in particular settings’ (Hyland, 2008, p8) and as ’text building blocks’ (Biber et al., 2004, p443). Usually transparent in meaning, they tend to be structurally incomplete and often bridge two structural units, i.e. a clause or phrase, very often functioning as the pragmatic head of an utterance and acting as an interpretative frame for the discourse that follows (Biber and Barbieri, 2007, p. 8). Lexical bundles are generally made up of grammatical words while the keywords discussed in the previous chapter tend to be content words, belonging to the noun, verb and adjective class predominantly

Variously referred to in the literature as formulaic sequences (Wray 2002; Schmitt and Carter 2004), lexical bundles (Biber & Conrad, 1999), n-grams (Stubbs and Barth 2003) or lexical phrases (Nattinger & DeCarrico, 1988), lexical bundles have received a fair amount of attention in the literature, though not, as we shall see, in the healthcare discourse literature. While different terms are used for these multiword sequences, for McEnery and Hardie (2012, p110), lexical bundles are, ‘methodologically and technically’, simply recurring sequences of *n* words, i.e. n-grams. They add that the term ‘lexical bundle’ has become associated with the work of Biber and colleagues on register description, and on their focus on the structural and functional interpretation of lexical bundles. As it is the structural and functional interpretation that interests me, it is Biber and colleagues’ terminology and approach that I have chosen to use. While earlier studies on multiword units relied on intuitive lists of prefabricated expressions (e.g Pawley and Syder, 1983; Nattinger and DeCarrico, 1992), corpus software has permitted an evidence-based approach to studies, with Altenberg’s study (1998) of the phraseology of spoken English, being one of the earliest.

In this chapter, after an overview of the methodological steps taken to extract lexical bundles, which has been presented in full in chapter 3, I will present the results of my analysis of the bundles, describing their frequency and their distribution in patient information for radiography. As I have already reported in chapter 3, for this doctoral study the analysis was restricted to 4-word lexical bundles. These bundles are less common than 3-word bundles, which occur very frequently in both spoken and written discourse (Conrad and Biber, 2004) but are not as rare as 5-and 6-word bundles, meaning an analysis of 4-word bundles results in a sufficient, but not an overwhelming quantity, of data. This will be followed by an analysis of the discourse functions and the communicative purpose of the identified bundles in patient information.

I begin, however, with an overview of the literature of the lexical bundle literature. As we have seen in chapter 2, unlike a keyword analysis, a lexical bundle analysis has rarely been used in healthcare discourse studies, though in studies of academic registers, however, lexical bundles have been the focus of many studies.

5.1 Lexical bundles in the literature

Lexical bundles are a powerful tool for the understanding of the unique characteristics of registers (Biber 1988) and have been described as the ‘building blocks of discourse’ (Biber, Conrad & Cortes 2004, p. 401). Bundles are found in both spoken and written discourse though their frequency and distribution differ. Conrad and Biber (2005) showed that the 3-word bundle *I don’t know* appears repeatedly in conversation at over 1,000 times per million words, and, while individual bundles also appear often in academic prose, the most-used items appear far less frequently, at between 200 and 400 times per million words (Conrad and Biber, 2005). Conversation, then, might be said to possess a repetitive quality that is generally not seen in written discourse.

The range and frequency of lexical bundles, however, are not solely defined by the mode of discourse. A seminal study by Biber et al. (2004) found that university classroom talk uses a wider range of types and higher frequency of lexical bundles than informal conversation and academic prose, evidence, says Barbieri (2018) of the communicative purposes of classroom teaching, ‘which combines the informational focus typical of academic prose with the expression of personal stance and interpersonal meanings typical of casual conversation.’ (p. 253)

The university is the focus of many studies in the literature, and while some studies have focussed on or included spoken university registers in their studies (e.g. Biber et al., 2004; Biber & Barbieri, 2007, Csomay, 2013), the literature on academic writing predominates, almost certainly because of the increasing importance of English in global academia and because of the rising number of foreign students studying in English. The research article has been a particular focus, with studies that identify bundles specific to different sections of the research article, and explore the functions of those bundles (Cortes, 2013; Jalali et al., 2015).

The frequency and type of bundle have been found to vary considerably across different disciplines (Cortes, 2002; 2004; Durrant, 2017; Hyland 2012; 2008a; 2008b). Durrant (2017), used Hyland’s (2008a) taxonomy, findingbb evidence for a clear distinction between the hard and soft sciences, with two further groupings of life sciences and commerce sitting between the two. In addition, evidence is found of disciplines which are essentially heterogeneous in nature such as engineering and cross-disciplines which draw on a variety of influences, such as the health sciences.

Difference in frequency and type of bundle are found between and within spoken and written academic modes as we have already seen (Biber et al., 2004; Biber and Barbieri, 2007), while variation has also been seen between expert writers and novice writers (Cortes, 2004), both in the range, type and function of the lexical bundles that they use. Cortes (2004) found that university students of history and biology rarely used lexical bundles in their writing, and when they did, their use did not correspond to the uses of bundles employed by professional authors.

More recently, studies have increasingly focussed on language background, finding that L1 and L2 speakers of English use different kinds and quantities of lexical bundles (Ädel and Erman, 2012; Bychkovska and Lee, 2017; Chen and Baker, 2010; Pan et al, 2016). It is perhaps less surprising that there are differences between L1 and L2 users, but it turns out that the differences are not related to frequency alone, but also to the function, the structure, and type of the bundles. Pan et al. (2016) compared the use of bundles in telecommunications journal articles written by English L1 academics with those of Chinese peers writing in English. The study found that while both groups used lexical bundles, the L2 professionals preferred bundles made up of verbs plus clause fragments, in particular, passive structures, while L1 speakers used more bundles made up of noun plus prepositional phrases. Similar results were found by Efandiari and Barbary (2017) in their comparative study of English and Persian writers of psychology research articles.

Staples et al. (2013) looked in more detail at the development of bundles in L2 writers, focusing on their frequency, function, and degree of fixedness. For a corpus, the study used essays written by candidates in the TOEFL iBT exam. With important implications for language teaching, there were few differences in the fixed versus variable slot bundles used by different proficiency levels of learners, and while lower levels actually used more bundles, closer inspection revealed that many of these bundles were copied from the essay prompts. This last finding echoes that of Wray and Perkins (2000) who found that L2 learners are much more likely to rely on the imitation and repetition of formulaic sequences (p10). Referential bundles, e.g *that kind of thing; the end of the; as shown in fig*, were very rarely used by any candidate, irrespective of proficiency level. Pan et al.'s study also found that noun plus prepositional phrase bundles (which many referential bundles tend to be) were not used by their learners, irrespective of level.

Lexical bundles, then, need to be learned. The evidence presented above (e.g. Cortes, 2004) that expert and novice L1 writers use different quantities and types of bundles and the L2 users, in addition to using different types of bundles, also use different structural types of bundle (e.g. Pan et al. 2006) suggests that bundles are not something that are easily acquired and may need to be overtly presented and taught by ESP teachers. Nesselhauf (2005, p. 69) describes L2 learners as using bundles like 'lexical teddy bears', a reference to students' tendency to overuse a small range of (favourite) bundles. Additionally, referential bundles, which are a feature of expert and L1 writing and very prominent in informational text (I use text to refer to both written and spoken language), are generally avoided by L2 learners of English (Staples et al., 2013). It is certainly the importance of bundles, combined with their proliferation and the need to learn them in order to become a member of the particular discourse community - a need that is the same irrespective of language background - that has resulted in the majority of lexical bundles studies focussing on the academic.

What of studies of non-academic registers, or studies that have compared bundles across registers from the same domain? A small number of domain-specific studies have looked at lexical bundle frequency and function in a range of registers within one domain, e.g. law (Breeze, 2013; Gozdz-Roszkowski, 2011) and pharmacy (Grabowski 2013; 2015) - Grabowski (2015) has been presented in chapter 2 - while studies of non-academic registers include Barbieri (2018), who looks at lexical bundles in blogs. Barbieri (2018) finds that blogs are characterised by a combination of stance expressions and make heavy use of verb-phrase structures. These types of bundles, as we have seen earlier in this chapter, are more commonly found in conversation (Conrad & Biber, 2005). What is particularly interesting, however, is that blogs also rely on referential bundles and narrative expressions, a bundle combination reflects both the communicative purpose of a blog and the mode.

Lexical bundle studies in medical registers are generally limited to medical research papers (Jalali & Moini, 2004; Jalali et al., 2015; Mazzi, 2016), with occasional exceptions: Kopaczyk (2013) looked at 3-word lexical bundles in Early Modern English medical writing, and there have been no studies of lexical bundles, to my knowledge, in other medical registers. One reason for the absence of non-academic focussed studies may well be the emphasis in the literature on the academic, as I have previously discussed. This emphasis might explain why the language of medicine, in these studies, is treated as homogenous, instead of a genre that is made up of a wide variety of clinical specialities. Medical students, after all, do not specialise while they are still students. Studies of research papers or abstracts tend to use a corpus made up of articles from a range of specialities. Mazzi (2016) uses articles from journals from 14 different specialities, selected after asking for advice from his university medical and scientific colleagues. The journals are the *British Journal of Dermatology, Journal of the American Academy of Dermatology, Blood Cells, Molecules and Diseases, Proceedings of the National Academy of Sciences of the United States, Cancer Research, British Journal of Haematology, Artificial Organs, Proteome Science, Clinical Chemistry, Journal of Pharmaceutical and*

Biomedical Analysis, Science and Current Opinions in Genetics and Development. My understanding of these specialities is that they might be expected to demonstrate not only different vocabulary but, as Durrant (2017) shows, a different use and range of lexical bundles. Mazzi (2016) cites Hunston (2008) in suggesting that the well-established meanings and functions of 'phraseologies' (p. 14) override any subject differences, though if the objective of such lexical bundle studies is to help train students to write 'as a medical researcher [which] implies being capable of talking to the expert members of the relevant discourse community in ways they find most effective', I believe subject-specific studies would be more useful.

Aside from medical research and historical investigations, other medical registers have yet to be studied. Given the success of a lexical bundles analysis in revealing the true communicative purpose of a text, as reported by the studies I have referred to in this section, and the importance and ubiquity of written patient information, a bundle analysis of these healthcare materials seems overdue.

5.2 The discourse function of lexical bundles

Lexical bundles serve important discourse functions and can be broadly categorised as referential (e.g. *at the same time; the rest of the*), discourse organising (*let's have a look; if you have any*) and stance conveying (*it's not possible to; if you want to*) (Biber et al., 2004a; Conrad & Biber, 2005; Cortes, 2004; 2006; 2013). These categories contain further, more defined sub-categories, e.g. the category Discourse-organising bundles contains two sub-categories of bundle: Topic introduction and Topic elaboration and clarification. The taxonomy is presented in full 3.5.2.7 and is summarised below in Figure 8

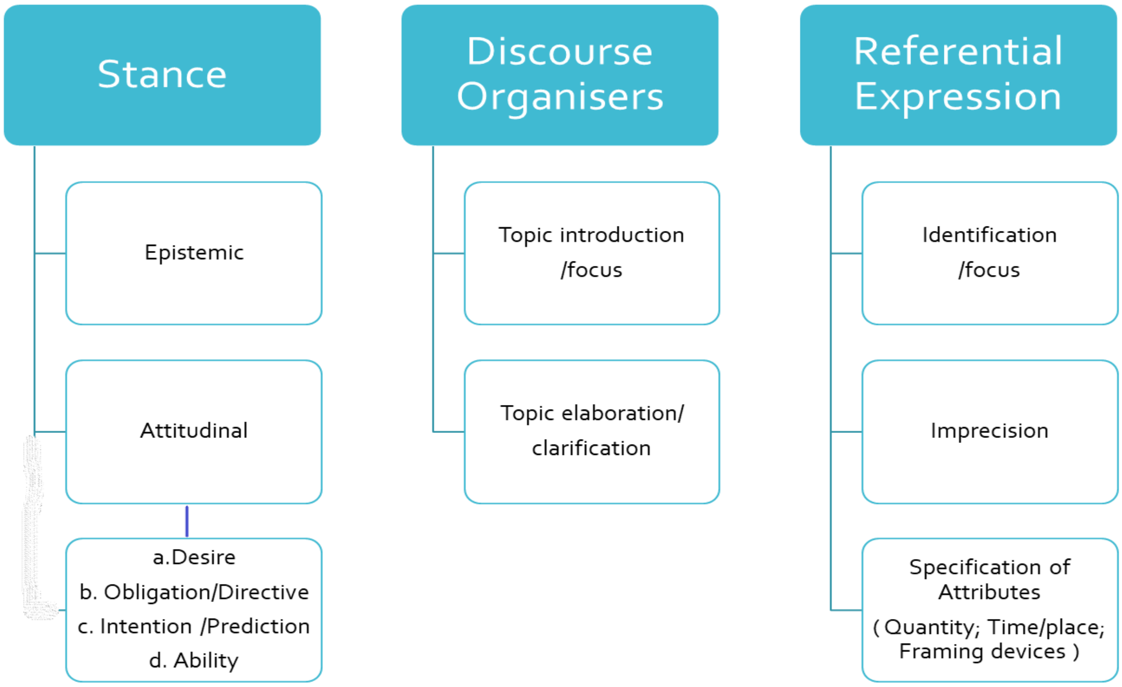


Figure 7 Taxonomy of discourse categories, after Biber et al., 2004a; Conrad and Biber, 2005.

Assigning a discourse function to the bundle is a significant step in the methodology, which is re-visited below and presented in detail in chapter 3.

5.3 Methodology

As this is one study in a larger investigation of patient information, the focus of the analysis was solely on 4-word bundles in order to avoid unmanageable quantities of data. Sketch Engine refers to ‘n-grams’ and thus a search was carried out limiting the span of the n-gram to 4 words.

The cut-off point was set at 20 per million words. This means that all 4-word bundles appearing at least 20 times in the corpus were included. While 40 is also a common cut-off in similar studies (e.g Biber and Barbieri, 2007; Goźdź-Roszkowski, 2011), the cut-off point it is always fairly arbitrary and much depends on the size of the corpus, the researcher’s preferences and the length of the bundle. A lower cut-off point, for example, is generally selected for the rarer, 5-and 6-word bundles (e.g Cortes, 2013). The small size of the corpus used in this thesis, just over 400,000 words, and my impressions of the quality of the data between 20 and 40 per million words (pmw), were factors in the decision to use the lower cut-off. There were a number of bundles that I felt were interesting to investigate further between the 20 and 40 per million cut-off point that would have been excluded by choosing the upper cut-off point.

Dispersion, as I have explained in chapter 3, is also an important aspect to control for, in order to minimise idiosyncratic uses of a bundle (one or two writers favouring a bundle that is not used by anyone else, for example) and thus each bundle needed to appear in at least five documents in the corpus to be included in the final list.

Once the settings had been decided upon, a search was carried out and a list of 4-word bundles produced. The second important stage, once the list was extracted, was the identification of suitable bundles for further analysis. Sketch Engine automates the search but cannot distinguish easily between 4-word sequences that are random, or part bundles. To help me identify suitable bundles a list of exclusion criteria had been drawn up and appears below in Table 16. The exclusion criteria were arrived at based on my readings of the literature.

Table 16 Exclusion criteria applied to extracted list of lexical bundles

Exclusion Criteria
Fragments of other bundles i.e. *eat or drink any; tip of the part*
Topic/Name specific e.*g in X Plain-T; University College Central Clinic*
Bundles with random or meaningless numbers e.g. *know page 40 if*
Web noise e.g. *at www.radio.com*
Clear Legal disclaimers *e.g. do not copy this*

Once the exclusion criteria had been applied, the bundles were classified according to their grammatical type using the taxonomy first described by Biber et al. (1999).

The final step was to classify the bundles according to their discourse functions, using the categories described by Biber et al (1999) and expanded in Biber et al. (2004) and Conrad and Biber (2005).

Assigning discourse function is not obvious simply from looking at an isolated bundle. Bundles, in fact, do not necessarily possess a function irrespective of context, and some bundles can appear in different categories as I explained in chapter 3: they possess multiple functions that are context dependent, making it imperative to investigate how the bundles are used in the data before assigning them to a category. By way of example, *at the end of* can be used to refer to both time and to place, e.g. *at the end of the corridor* or *at the end of the day*. It can also be used as an expression of identification or focus in a sentence such as *at the end of the process.* Checking the use of the bundles in the corpus, then, though a long process, is an essential one.

5.4 Results

5.4.1 Overall distribution of lexical bundles in patient information

In this section, I will first report on the overall frequency of lexical bundles and the distribution and frequency of their structural type. Then I will present the results of the categorisation of discourse function.

109 unique types bundles were extracted from the corpus of 408,997 running words. There is a total of 3725 bundles, representing 3.6 % of the total number of running words in the corpus. To put this figure into some comparative perspective, Goźdź-Roszkowski (2011) found that 4-word bundles represented 4.2% of the running words in a corpus of legal textbooks, 2.4% of the running words in a corpus of professional articles and 9.4% in a corpus of legislation. Conrad and Biber (2005) found that 4-word bundles made up 3% of a corpus of conversation (compared to 25% for 3-word bundles), and 2% of a corpus of academic prose. 3.2%, then, is a finding that seems appropriate for a discourse type that seems, at first glance, to lie midway between conversation and more formal prose. The final list of bundles, ordered by frequency, can be seen in Appendix B.

5.4.2 Structural type

A closer look at my data reveals that as many as two-thirds of the bundles are of a structural type that is more commonly found in academic prose (Biber et al., 1999). The structural types found in my corpus of patient information can be seen below, in Table 17

Table 17 Distribution of structural types of lexical bundle in patient information after Biber et al. (1999)

Distribution (%) of bundles by grammatical type in patient information corpus*			
More common in conversation		*More common in academic prose*	
pronoun + lexical verb phrase	8.0	NP + post-modifying fragment	9.0
pronoun/NP + (AUX) + be	5.0	Prep + NP fragment	27.0
(pronoun) (AUX) + active verb	16.0	'it' + VP/adjP (+ complement clause)	5.0
yes-no + QU-word fragment	5.0	Passive verb + PP fragment	18.0
(verb) + WH-clause fragment	0.0	Verb (+ that) clause fragment	1.0
		Other expressions	6.0
Total	34.0	**Total**	66.0

*rounded to the nearest 0.5%

Two structures more common to academic prose are particularly frequent in the corpus: Passive verb + PP fragment, e.g. *that may be used; can be treated with; may be needed to*; and Prep + NP fragment, e.g. *during the course of; at high risk for*.

This last type, Prep + NP fragment, are the most frequent type in the discourse and well over a quarter of all the bundles fall into this category. These two categories alone account for 45% of all of the bundles in the final list of 109.

A third of the bundles were of a kind that predominate in conversation, the most common being Pronoun + AUX + Active verb, e.g. *you may feel a,* representing 16% of the number of bundles. The high frequency of modal verbs used in patient information may well explain the predominance of this kind of bundle. We shall report on the use of modal verbs in patient information in the following chapter.

These results show that the frequency of structural types varies quite markedly, with some bundle types being used repeatedly, while others appear very infrequently. Five bundle types, two more common in conversation and three more usual in academic prose, represent around 78% of the final list of 109.

5.4.3 Discourse function of bundles

With regards to the discourse function of the bundles, the results appear below in Table 18. As we can see, the most frequent bundle types are split more or less evenly between referential and stance: 52 bundles are categorised as stance bundles while 54 are categorised as referential bundles. Discourse organising bundles, on the other hand, are used far less frequently and make up just 12% of the total number of bundles.

Table 18 Discourse function of bundles in patient information for radiography

Discourse Function	Number of individual bundles
STANCE	**55**
Epistemic	4
Attitudinal	
Desire	2
Obligation/Directive	9
Intention/Prediction	27
Ability	13
DISCOURSE ORGANISING	**13**
Topic introduction/focus	6
Topic elaboration/clarification	7
REFERENTIAL BUNDLES	**41**
Identification/focus	6
Imprecision	0
Specification of Attributes	
Quantity specification	4
Tangible framing attributes	2
Intangible framing attributes	7
Time reference	11
Place reference	10
Text reference	1
TOTAL	**109**

5.4.4 Frequency of structural types

We have seen that some bundles are used with great frequency in certain types of discourse. Conrad and Biber (2005) reported that one bundle, *I don't know* was used over 1000 times per million words in conversation. Such high rates of use are less common in academic prose, however.

In patient information, there are very few bundles that are used with great frequency, and none at all used with the frequency of certain bundles in spoken discourse. There are five bundles used more than 200 times per million words: *you may be asked*, *if you have any*, *you will be asked*, *if there is any* and *how do I get.* The most frequent bundle in my data, *you may be asked*, appeared 221 times (471 pmw). This was followed by *if you have any,* which occurred 207 times (443 pmw) and *you will be asked*, at 186 occurrences (396 pmw). *If there is any* occurred 140 times (299 pmw) and *how do I get* appeared 97 times (207 pmw).

In their study, Conrad and Biber (2005) found that the most frequent bundles in academic prose appeared between 200-400 times per million words, a similar finding to that reported here for patient information. Compared to conversation, then, written patient information shows evidence of being formulaic, but it does not have the repetitive characteristics of spoken discourse. A wide variety of bundles are used, but only five bundles appear with any notable frequency - i.e. more frequently than 200 per million words.

While patient information makes use of a number of bundle types that are more commonly found in conversation, bundle types that are more common in academic prose and informational discourse predominate. I now turn to a detailed discussion of these findings.

5.5 Discussion

In this section, I will discuss some of the more significant findings reported above. I will begin by focussing on the structural types of bundle found in the data, proposing some explanations for the reliance on these bundle types, and, with my research questions in mind, considering what the occurrence of these bundle types tells us about the characteristics of patient information. I will then move on to investigate the discourse functions of the bundles extracted in more detail, an investigation that will further my understanding of the lexical characteristics of patient information but also, perhaps, reveal some more of its underlying discourses, some of which have been reported on in chapter 4.

5.5.1 Structural types

5.5.1.1 Passive verb + PP fragment

We have seen that two-thirds of the bundle types in patient information are of the kind more often found in academic writing. This is surprising when one considers the need to produce healthcare information materials that are easy to read and accessible to the greatest number of people. Academic prose and easy-to-read text do not seem compatible.

We saw in chapter 1 that simplifying text to meet a certain reading age is the usual approach taken to making patient information readable. This generally involves a focus on shorter sentences, simpler vocabulary, with definitions provided for any medical words that need to be used and active sentences. An NHS guide for their information writers expresses it thus:

> Various studies have shown that the average reading age of a British ADULT is between 9 and 12 years. So if you are writing a leaflet it might be an idea to get an average nine or ten-year-old to try to read and understand it! Readability is simply a measure of how easy a piece of text is to read. Readability can be calculated in lots of different ways, but basically the following applies: Short words + short sentences = information that is easy to read. (NHS Scotland, 2007)

The guide also suggests that writers use the active voice and avoid passive structures, which is advice common to all communication guides. It is interesting, then, to note that the 2nd most common bundle type in my data is Passive verb + PP fragment. Examples of passive structures in the data include:

(3) ***It will be performed*** *in the interventional radiology suite* ***. You will be asked*** *to lie on your back on an x-ray table. Monitoring equipment* ***will be attached to*** *you to measure your blood pressure and heart rate.*

(4) *Baby soap* ***may be used to*** *wash the treatment area.*

It is difficult, at first, to understand why the passive has been used in some sentences. (4) appeared in a section where the active voice was predominantly used:

(5) *If possible, shower instead of bathing, use lukewarm water, not hot and do not stay in the shower for long periods of time. Do not use shower gel, bath oils, and bubble bath as this may cause a skin reaction. Baby soap may be used to wash the treatment area. Pat the skin dry with a soft towel, do not rub as this may make the skin sore.*

In other cases, an entire section was written using passive structures, including both 3- and 4-word bundles:

(6) *To stop your bowel moving on the x-rays* ***you may be given*** *a small injection in your arm. The tube* ***will be removed*** *and* ***you will be taken*** *to the toilet.* ***You may be asked*** *to go into a different room for a further x-ray after you have been to the toilet.*

The data in (6) came from information regarding a barium enema, a rather unpleasant procedure which the writers had previously referred to as *a little undignified.* Is the passive used here precisely because the procedure is considered unpleasant or embarrassing? I did not gather this information in the course of my thesis as it lay outside the scope of my inquiry, but it would certainly be an area worth further investigation.

In our first example, (3), the passive is used to describe what will happen during the examination. It also seems unnecessary to use the passive here, rather than an active verb plus 'we', e.g. *we will ask you to lie down.* The effect of the distance created by the passive is not at all reassuring but, on the contrary, cold and unfeeling. Who will do the actions referred to in the example? Almost certainly the radiographer or radiography nurse, both of whom, as we saw in the keyword analysis, are rarely named in the patient information. Would it not be more appropriate to use an active sentence with either one of these two professionals as subject, or 'we'? e.g. *The radiographer will attach monitoring equipment to you…*

The passive structure is often portrayed as a structure that is less clear and direct than its active counterpart and too complex for readers to process, irrespective of how and where it occurs in a text (Minton, 2013). This is an idea particularly common to communication guides. It is overly simplified, however. Zarcadoolas (2011) reminds us that context is key and references Coleman (1964) who showed that children comprehend passive structures in context, even when they could not comprehend the same structures in isolation. And while passive structures do require different processing skills on the part of the reader (Mack, Meltzer-Asscher, Barbieri & Thompson, 2013), the passive has an important function in spotlighting the focus of the sentence. As Minton (2013, p. 4) says,

> When active-voice and passive-voice sentences are properly composed and appropriate to the context in which they are used, there are no grounds whatsoever for claiming that one voice is clearer or more direct than the other.

Minton (201, p. 5) illustrates his position by pointing out that each of the following sentences is appropriate depending on who, or what is the focus of interest. ‘Columbus discovered America in 1482’ (Active) and ‘America was discovered by Columbus in 1482’ (Passive). Leaving aside the fact that America was already populated and thus did not need discovering, the importance of clarifying the key message may well explain the use of passive bundles in patient information.

The examples we have seen in (3) and (6) foreground the experience for the patient. To the extent that patient information is written for patients to better understand what will happen to them in the radiography suite, the passive seems a more appropriate structure than the active (e.g. *we will ask you to…* or *we will remove the tube)* where the focus is on the medical professional. The sudden appearance of the passive bundle in (5) can, I feel, also be explained by the need to focus the attention on the most important piece of information in the sentence, which is the noun, *the baby soap*, as opposed to the *baby oil*, *bubble bath* and *shower gel*, none of which should be used by the patient, and as opposed to the verb *use*, which is the key verb in the paragraph and is presented early on. It is not new information.

With this in mind, avoiding the passive entirely may be impractical advice for patient information writers. It may also be the case that presenting information in the active voice only, when it would be more logical sometimes to present it in the passive, may have an impact on how the information is read and comprehended. In the following two sentences (not taken from my data), the active sentence in (8) seems less clear than the passive in (7):

(7) A CT scanner is used to take multiple x-ray images of your body

(8) We use a CT scanner to take multiple x-ray images of your body

The active sentence also leaves open the question of whether other machines could have been used, or whether CT scanners have other functions, as well as take multiple x-ray images. They do not. Based on this one example, it would seem that active sentences are not always clearer than passive sentences.

To conclude this section, further studies of how and why passive bundles are being used in patient information, and how patients feel about them in terms of ease of comprehension and clarity, would be very welcome.

5.5.1.2 Prep + NP fragment

What of the other bundle-type very commonly used in patient information: the Prep + NP fragment? Almost a third of the total number of bundles were accounted for by this bundle-type. Examples of this structure include *at the end of*, *in the area of* and *during the course of*. Investigating the bundle structure in the corpus I discovered that, while some of these structures had a framing function (e.g. *as a result of*), the majority of the structures referenced time and place:

(9) *This procedure combines special x-ray equipment with sophisticated computers to produce multiple images or pictures* ***of the inside of*** *the body.*

(10) *The technologist will attach electrodes to your chest, wrists, and ankles. These will be used to record an EKG* ***at the same time*** *the echo is taken.*

The reliance on these types of structural bundles underlines one of the primary functions of patient information, that of providing information. Referential bundles are, in fact, a strong feature of informational discourse (Biber et al., 2004; Biber & Barbieri, 2007; Barbieri, 2018) In fact, information-giving appears to be the primary function of the majority of bundles in the corpus, as I will demonstrate in this discussion section. The information included the benefits of the medical procedure, patient preparation for the procedure, the steps of the procedure itself, the time required for the procedure, any equipment used, the meaning of certain significant terms and post-procedural recovery.

Referential bundles will be discussed in more detail in 5.5.2.

5.5.1.3 (pronoun) (AUX) + active verb

While structure-types were predominantly those more commonly found in academic prose, a third of the bundles are those that are more commonly found in conversation. These structures are clausal, often involving a verb or auxiliary verb. Of these, the most commonly used, representing 16% of the 109 bundles, was (pronoun) (AUX) + active verb.

Examples of this type include *you will have a; you will need to* and *may need to be.* As will see in the following chapter, auxiliary verbs *may* and *will* are very frequent in patient information, and, as might be expected, the future is often referenced. *Need to* is also more frequently used than expected when its use in general English is compared (e.g. Johansson, 2010). Modal verbs for giving instructions will be discussed in detail in the following chapter.

Let us now turn to a more detailed look at the discourse function of the bundles. I will begin by looking at Referential bundles, which represented more than a third of the bundles extracted.

5.5.2 Referential bundles in patient information

5.5.2.1 Specification of attributes: Time

Referential bundles that identify some specific attribute of the following head noun often relate to time, place or text. In the patient information corpus, these make up more than 50% of all referential bundles. On closer inspection, these bundles are very evenly distributed between those referencing time and those that reference place. There was just one occurrence of a bundle referencing the text.

The time referential bundles are nearly always imprecise. When they reference a ‘window’ of time they generally refer to the medical procedure itself and preparation for it:

(11) *You should not eat or drink after midnight* ***on the day of*** *the procedure*

(12) *The skin becomes darker* ***during the course of*** *radiotherapy, similar to tanning from the sun*

(13) *…a doctor will examine you* ***before you leave the*** *department.*

(14) *…patients should avoid blood-thinning medication for the recommended* ***period of time before*** *the treatment.*

(15) *Detailed instructions will be given* ***at the time of*** *booking your appointment.*

(16) *The therapy is usually given* ***over a period of*** *several weeks*

When the bundles refer to recovery time or possible side-effects, they name the unit of time (weeks, hours or seconds) but, generally, they too are imprecise and approximate:

(17) You may feel a warm sensation ***for a few seconds*** *when the dye is injected*

(18) You may feel sore at the end of the biopsy ***for a few days***

(19) Skin reactions usually heal completely ***within a few weeks*** *of completing radiotherapy*

The difficulty of predicting with any great certainty the duration of anything medical is a likely explanation of the imprecision. It is also the case, as we have seen in 4.5.1, that the issue of time is a legally sensitive one. A patient may feel that they have grounds for complaint or legal action if their experience does not match official information. Being vague is legally advantageous.

5.5.2.2 Specification of Attributes: Place

Unlike imprecise time bundles, those that reference place are relatively precise, and nearly always reference the body area being examined, or a part of the scanning machine:

(20) *Tissue samples are removed* ***from the area of*** *concern using a hollow needle*

(21) *It also is possible that the catheter tip will separate material from* ***the inner lining of*** *the artery, causing a block downstream in the blood vessel.*

(22) *[…] the CT table moves you very slowly towards the hole* ***in the centre of*** *the "polo" shaped scanner.*

There is one bundle, *at the end of*, that is used to reference both place,

(23) *A balloon* ***at the end of*** *the catheter is inflated with contrast*

(24) *Small balloon-like sacs called alveoli are* ***at the end of*** *the bronchial tubes*

and time,

(25) ***At the end of*** *the procedure, the applicators are removed*

(26) *[…]**at the end of*** *the operation, the anaesthetist will stop giving anaesthetic drugs and you will start to wake up.*

The same bundle was also used once to reference the text:

(27) *[…] details can be found* ***at the end of*** *this leaflet.*

The focus in patient information on explaining what instrument will be acting on what body part explains the reliance on these bundles. There is more precision with these bundles as place information comes without the legal pressures that accompany time referential bundles.

5.5.2.3 Specification of Attributes: Tangible and Intangible framing

Framing bundles also identify attributes of the noun that follows. These attributes can be more concrete in nature (tangible) or abstract (intangible). Framing bundles represent around 10% of all bundles. Some of these bundles are used to identify and name something:

(28) *This radioactive material accumulates in the organ or area of your body being examined, where it gives off a small amount of energy* ***in the form of*** *gamma rays*

(29) *Follow-up imaging may be necessary to ensure that no foreign bodies remain in the body and to check* ***for the presence of*** *any side effects such as infection.*

The bundle *as a result of* seems to have a specialised use in the corpus, and is overwhelmingly used to refer to the side effects of a treatment or the (negative) result of a disease:

(30) In P.A.D., the arteries that carry oxygenated blood throughout the body become narrowed or even blocked, usually ***as a result of*** *atherosclerosis, or plaque*

(31) Side effects of radiation treatment include problems that occur ***as a result of*** *the treatment itself*

(32) *[…] there are rare reports of people having died* ***as a result of*** *infection thought to be due to the biopsy.*

Likewise, *the best way to*, which also seems to have a special use in the corpus, and is predominantly used to justify the proposed medical procedure or to give post-surgery/treatment advice:

(33) *An angiogram is* ***the best way to*** *find out if arteries are blocked or restricted by plaque*

(34) ***The best way to*** *fight fatigue is to get on a daily exercise regimen that is tolerable and sustainable, eat a healthy diet and rely on friends and family for support*

(35) *Follow-up examinations are sometimes* ***the best way to*** *see if treatment is working or if an abnormality is stable over time.*

5.5.2.4 Specification of Attributes: Quantity

There were a handful of bundles that expressed a quantity, none of which were precise. This lack of precision we have seen previously with referential bundles related to time.

(36) *Sometimes, **one or more of** these warning signs may happen and then disappear.*

(37) *In a conventional x-ray exam**, a small amount of** radiation is aimed at and passes through the part of the body being examined*

(38) *Some patients may cough up **a small amount of** blood after the procedure.*

The warning signs, radiation, and blood appear almost inconsequential by the use of these imprecise quantifiers, which may indeed be the objective.

5.5.2.5 Specification of Attributes: Identification / Focus

Finally, identification/focus bundles, are often used to name or define something medical, with the intention of clarifying things for the reader:

(39) *The procedure is also **sometimes referred to as** Uterine Artery Embolization (UAE)*

*(40) X-rays **are a form of** radiation like light or radio waves*

The bundle *any of the following* is used to present a list of options, sometimes relating to existing complaints or current medication, but also to potential side effects of treatment:

(41) *Please inform a member of staff if you answer 'yes' to **any of the following** questions…*

(42) *Please indicate if you have **any of the following**…*

(43) *You should report to your physician immediately if you experience **any of the following** after your procedure…*

5.5.3 Stance bundles in patient information

Stance bundles are the most frequent bundle in the corpus, even more so than Referential bundles, accounting for around half of all the bundles used. Stance relates to the expressions of attitudes or expressions of certainty that frame some other proposition (Biber et al., 2004, p. 384); there are two main categories of stance in Biber et al.'s (2004) taxonomy, Epistemic, which relates to expressions of certainty, and Attitudinal, which includes a number of sub-categories: Intention/prediction; Obligation/Directives; Ability and Desire.

Intention/Prediction are by far the most frequent, accounting for around 50% of the total number of Stance bundles and more than 20% of the total number of bundles. Intention/Prediction bundles are followed by Ability with 13 bundles and 15% of the total number of bundle types, and Obligation/Directives with 9 bundles, around 10% of types. The Stance bundles seen in patient information are overwhelmingly impersonal, that is, they are not overtly attributed to the writer but to the organisation (the hospital in many cases) or the medical system itself.

5.5.3.1 Intention/Prediction Bundles

These bundles have a clear use in patient information and that is to say what is certain or likely to happen during and after the patient's visit to the hospital, and what the patient is certain or likely to (be expected to) do. The ratio of bundles expressing a possibility (very often with the modal verb *may*) to those expressing a certainty (with *will*) is around 2:1. As we have said earlier in this chapter, predicting with certainty in the field of medicine is not straightforward and it is not surprising that *may* is used twice as often as *will* in these bundles.

There are two Intention/Prediction bundles in the five most-used bundles in the corpus: *you may be asked* (appearing 220 times (471 pmw) and *you will be asked*, occurring 186 times (398pmw). Both of these bundles are in the passive form, possibly to focus on the 'you' of the patient. I have discussed this in more detail earlier in this chapter in 5.5.1.1.

The fact that these bundles appear in the category Stance is not the end of the story, however. Examining these bundles in the corpus, I understood that both *you will be asked* and *you may be asked* generally function as instructions in patient information. A variation on these two is the active form *we will ask you*, which appears far less frequently in the corpus at just 13 raw occurrences. Though it seems to merely state what will happen, closer inspection reveals that it too is used to reference an instruction:

(44) The scan is taken very quickly and ***you will be asked*** *to hold your breath whilst it is taken.*

(45) On arrival ***you will be asked*** *to undress in a cubicle.*

(46) ***You may be asked*** *to change into a gown before your scan*

(47) ***You may be asked*** *to remove any piercings, if possible.*

(48) ***We will ask you*** *to remove all jewellery and body piercings before the scan as the scanner uses a very strong magnet.*

Irrespective of whether *will* or *may* is used, the information content is generally the same, suggesting that the choice of *may* or *will* in these bundles is down to the writer's preference.

In other cases, the choice of *will* over *may* seems to relate more closely to the idea of something that can, with certainty, be predicted. These include events that are invariable, such as the steps a patient is expected to follow when they arrive at the hospital, or the stages of a procedure that are the same for any patient.

(49) ***You will have a*** *blood test at the start of treatment*

(50) ***You will be asked*** *to lie down on an x-ray table*

(51) *The Radiologist performing the Nephrostogram* ***will be able to*** *let you know the results of the test before you leave the Department.*

How long something will take, however, i.e. the procedure, waiting time and recovery time, is variable, and not something that can be predicted with any great accuracy. As we have seen earlier in this chapter, nor is it something that medical bodies want to state with too much precision in case it lays them open to complaints and legal action when the reality fails to match what has been stated in writing. *Will* is not used here, but *may* or, less frequently, *should*. Likewise, aspects of a procedure that are only sometimes necessary or not appropriate for every patient appear with *may* or *should*:

(52) *For ultrasound of the aorta,* ***you may need to*** *avoid eating for eight to 12 hours before the test.*

(53) ***You should be able*** *to resume your normal activities within a week.*

(54) *This IV infusion* ***may take up to*** *two hours.*

How a person experiences a health condition, pain or the procedure itself, varies from individual to individual of course and, as a result, most references to sensation or pain or possible side effects are modified with *may*:

(55) *Occasionally,* ***there may be some*** *bleeding inside the breast and a bruise or swelling (haematoma) will form.*

(56) ***You may also be*** *aware of pressure from the biopsy needle as it takes the sample.*

(57) *When the radioactive material is injected into your arm,* ***you may feel a*** *cold sensation moving up your arm.*

In (55), the use of *may* along with the adverb *occasionally* and the vague quantifier *some*, contribute to reducing the likelihood - and threat - of the event described

Interestingly, when the topic is the opposite, i.e. not feeling pain or discomfort, patient information writers are more certain, and *will* is overwhelmingly used. In fact, *you may not feel* appeared just once in the corpus and with reference to a symptom of diabetes. It was never used to reference a treatment or examination.

(58*) **You will not feel** the catheter in your artery, but when the contrast material is injected, you may have a feeling of warmth or a slight burning sensation.*

(59) ***You will not feel** the treatment and the machine is very quiet.*

Pain is subjective, as we know (e.g. Coghill, 2010), so it is interesting that patient information writers should be so certain in these contexts.

5.5.3.2 Ability

There are 13 individual Ability bundles in the corpus, representing around 25% of the Stance bundles. Nine of these bundles involve the verb *use*, in all cases as a passive form, and in most cases modified with auxiliary verbs *may* and *can*:

(60) *A biopsy needle* **may be used to** *obtain a sample of lung tissue.*

(61) *Several imaging tests* **can be used to** *diagnose P.A.D.*

(62) *Occasionally, a device or plug* **will be used to** *seal over or close the hole in the artery.*

(63) *A nephrostogram is an x-ray procedure* **that is used to** *check your nephrostomy catheter and flow of urine through your ureter (water pipe between the kidney and bladder).*

In patient information, the bundles that include *used to* always refer to physical objects - medical equipment - and to medical procedures or tests. This is in line with Durrant's (2017) finding that most bundles for the description of procedures and processes in science and technology are centred around the bigram *used to*.

5.5.3.3 Obligation/Directive Bundles

There are nine individual Obligation/Directive bundles in the corpus, though very few are overt directives. The most frequently used 4-word bundle in this category is *you will need to*, the eighth most-frequent bundle overall. It appeared 49 times in the corpus (104 pmw). This bundle is used to issue instructions to the patient regarding the examination/procedure itself,

(64) ***You will need to*** *stand for the treatment, holding a bar within the treatment frame.*

(65) ***You will need to*** *have an empty bowel and a full bladder for your treatment;*

(66) *While the camera is taking pictures,* **you will need to** *remain still for brief periods of time.*

or to give the patient instructions regarding the recovery period:

(67) *If you go home the same day,* ***you will need to*** *arrange for someone to take you home by car or taxi and to stay with you overnight*

(68) ***You will need to*** *come back to the hospital for regular CT scans to check that the treatment has worked and that there is no recurrence*

(69) ***You will need to*** *stay in hospital after the biopsy for about four hours.*

As I will discuss in my examination of modal verbs for instructions in the following chapter, *need to* structures in patient information are invariably used to refer specifically to medical procedures. While 2nd person *need to* structures in other contexts are often perceived to be strong directives (e.g. ‘you need to be home by midnight’, uttered by a parent to a child), in patient information these structures are invariably presented as either a necessary part of the procedure (and thus closer to dynamic necessity), or as requirements that are for the good of the patient, such as getting someone to stay over with you after returning home, or returning for regular check-ups.

The most common obligation/directive 4-word bundle is *it is important that* which appears 90 times in the corpus, followed by the pronoun *you* in over 90% of the cases.

(70) ***It is important that*** *you follow the instruction below*

(71) *You have been given a laxative to take before your barium enema;* ***it is important that*** *this is taken following the enclosed instructions*

(72) *These devices will be used for the treatment to achieve the same position daily, so* ***it is important that*** *the patient can maintain that position*

(73) *In order to see the bowel* ***it is important that*** *it is empty. This is why we ask you to take the laxative prior to your scan.*

In each of (70)-(73), the patient is being told to do, or not to do, something, though the instruction is presented in terms of the importance of the action, leaving the patient to fully understand that this, in fact, is an instruction. In (72) and (73), why it is important is also stated, though not so in (70) and (71).

The most overt directive in this category is an imperative structure, an appeal to the patient to inform the medical staff if certain conditions apply. It was unusual however, and appeared just 14 times in the corpus:

(74) *If you are known to have an allergy,* ***please let us know*** *on the day*

(75) *If you are pregnant or think that you may be pregnant,* ***please let us know*** *before you have your scan.*

5.5.3.4 Epistemic and Desire bundles

While stance was the category of bundle most frequently used, neither epistemic or desire bundles, a subset of atttitude bundles, are included in this. In fact, there are only two desire bundle types in the entire corpus and just four epistemic types.

With regards to the two desire bundles extracted in the analysis, they are both used to offer further help or information to the patient:

(76) ***If you would like*** *information about any medication you may be given during the scan please contact us or speak to the radiographer when you attend for your appointment.*

(77) ***If you need any*** *assistance with transport to the hospital please contact your GP.*

Two of the epistemic bundles contain overt references to risk and chance and all came from one (American) website (www.radiologyinfo.org). It is possible that one individual was responsible for writing all or much of the patient information on this site and thus the two examples below might well be examples of idiosyncratic use:

(78) *Individuals* ***at high risk for*** *developing colorectal cancer should be screened more often and begin screening before age 50.*

(79) *There is always* ***a slight chance of*** *cancer from excessive exposure to radiation.*

That there were so few examples of bundles relating to likelihood or risk does not mean that information related to these aspects are not presented in patient information. *Risk* is frequently used, appearing nearly 1,000 times in the corpus (2,118 times per million words). Its most common collocates are *possible* and *potential*, followed by *high*, *slight* and *small*. *Benefit*, on the other hand, appears with a third of the frequency at 342 times (732 pmw) and with only one collocate that appears more than twice: *potential*.

This finding suggests that risk is presented in radiography patient information as something that is gradeable and quantifiable, unlike benefit. A lexical analysis of the use and connotations of risk and benefit in medical information for radiography, and how patients understand these messages, would be very useful, particularly in the light of the studies I presented in chapter 2, showing that radiography patients consistently under-estimate the concomitant risk associated with certain radiography procedures and seem equally uninformed about which procedures involve radiation, and thus risk. (Singh et al., 2017; Ukkola et al., 2017)

The final category of bundle to be discussed in the chapter are Discourse organising bundles, which, as we shall see in the next section, were suprsisingly infrequent in the corpus.

5.5.4 Discourse Organising Bundles

5.5.4.1 Topic Introduction/Focus Bundles

Discourse organising bundles are the least used bundle-type in patient information, representing just 12% of the total proportion of bundles. They are evenly split between Topic Introduction/Focus and Topic elaboration/clarification. In spite of their infrequency, however, three discourse organising bundles are among the five most frequent bundles in the corpus: *if you have any*; *if there is any* and *how do I get*, though as we shall see, these bundles are often used to frame instructions.

If you have any is the second most frequent bundle in the entire corpus, occurring 207 times (443 pmw). This bundle is a topic introduction or focus bundle which is overwhelmingly used with the object *allergies* or *questions/queries*. In both cases, the information that follows is generally presented as an instruction. The surrounding text usually contains an imperative or a modal verb of obligation:

(80) ***If you have any*** *of these warning signs, call 911 right away.*

(81) *You should tell the radiographers if you have had an allergic reaction to iodine or contrast dye in the past or* ***if you have any*** *other allergies*

(82) ***If you have any*** *queries please telephone 020 7351 8220*

The next most-frequent Topic Introduction/Focus was *if there is any,* appearing 34 times in the corpus. On closer inspection, it transpired that all but one of these uses appeared in the American materials. While the bundle appeared in 25 different documents, the fact that one author may have been responsible for writing much of what appears on the site, or that a ‘house-style’ may have been in use, cannot be ruled out.

5.5.4.2 Topic Elaboration/Clarification

These bundles precede more detailed information about an already-introduced topic. In patient information, this can refer to the steps involved in a medical procedure,

(83) ***How do I get*** *the results of my scan?*

the reasons for something procedural,

(84) *Many imaging tests are not performed during pregnancy* ***so as not to*** *expose the fetus to radiation.*

(85) *A chest x-ray will be taken* ***to make sure that*** *the lung has not collapsed from an air pocket created during the procedure*

(86) *In a biopsy, a small amount of tissue is removed under local anesthesia* ***so that it can*** *be examined in a laboratory*

or an explanation of an imaging modality:

(87) *Positron emission tomography (PET)* ***is a type of*** *nuclear medicine scan that uses a small amount of radioactive material to image body functions*

How do I get is a particularly frequent bundle in patient information and is almost always used in the context of scan results. This is one question that radiographers field on a daily basis, underlining the importance to patients of knowing what the scan or x-ray has seen. The job of a radiographer, however, is usually restricted to carrying out radiography. Aside from specialist radiography roles which permit some diagnosing, it is the radiologist who diagnosis. Radiographers report being asked continually, sometimes pressurised, to give results by worried patients, but they are not permitted to do so. Nor, in many cases, do they possess the skills. The *How do I get my results*? section in patient information serves to inform patients prior to coming to the department that their radiographer will not be diagnosing them. Whether patients read this or fully understand this, we do not know.

The discourse organising bundle *if you do not* is often used to present what will or could happen if the patient acts in a manner that is contrary to that advised or desired by the hospital. Sometimes this presented in a manner that is quite alarming, as in (88).

(88) ***If you do not*** *follow your diet, exercise, and perform sugar level tests, serious complications can arise.*

(89) ***If you do not*** *get treatment, chest pain may happen more often.*

Discourse organising bundles are used far less frequently than might be expected, given the nature of the text. On the other hand, we have seen that three of these bundles are, in fact, used very often in the text, with three in the first-5 most frequent. This demonstrates that while there is little variety in the bundles, a small number of them are relied upon and have an important function in patient information. Discourse organising bundles ‘reflect relationships between prior and coming discourse’ (Conrad et al; 2005, p67) and serve to introduce a change in topic or to add more detail to the topic being discussed. It may be that the style and format of much patient information render discourse organising bundles less necessary. Patient information is often arranged as a series of questions and answers - and some of the 4-word, discourse organising bundles in patient information are part-questions - e.g. *when will I get; why do you need.*

All advice to patient information writers to dispense with long sentences and complex structure should mean a minimal number of complex paragraphs that require a range of connecting, cohesive devices. The following paragraph is typical of the patient information in the corpus. There are ten sentences but no relative pronouns. The average sentence length is 15, though the shortest sentence is just six words and the longest is 24. The paragraph contains no 4-word, discourse organising bundles (other 4- and 5- words bundles underlined):

Bronchoscopy *The doctor uses a bronchoscope during bronchoscopy. A bronchoscope is a long, thin, and flexible fiber optic tube that transmits pictures from the tip to an eyepiece or to a video set. During a bronchoscopy, the bronchoscope is used to look at the larynx, trachea, and bronchial airways of the lungs. This procedure shows more details from the inside of the airways than pictures taken with X-rays. The bronchoscope has an open channel. This allows instruments to go through the scope and be used to take tissue samples, cauterize bleeding, or remove thick mucus blocking the airways. The doctor that performs the bronchoscopy procedure is a pulmonologist, a specialist in the respiratory system. A bronchoscopy can be used to examine many different respiratory tract symptoms. These include pain in the trachea, difficulty breathing, bleeding, tumors, and chest pain. Clear and detailed images and video projected on a monitor helps the doctor diagnose problems.*

To conclude, I believe it would be an interesting exercise to see how this paragraph would look with the addition of a discourse organising bundle or two. It may be that the simplification of the text, the shorter sentences and the repetition of the nouns, in place of pronouns, perform the functions ascribed to discourse organising bundles, that of topic focus, elaboration and clarification. Some targeted experiments would help answer that question. The text would look differently with some relative pronouns too (*i.e. that, which, who*), of which there are none. The sentence *the doctor that performs the bronchoscopy procedure is a pulmonologist, a specialist in the respiratory system* could also be written as *the doctor that performs the bronchoscopy procedure is a pulmonologist, who is a specialist in the respiratory system.* Which would patients find easier to read?

Students asked to simplify healthcare text as part of their language studies should, in my view, also be looking at cohesion and coherence, which discourse organising bundles contribute to, and not only at vocabulary clarity and complexity. Zarcadoolas (2011), discussed in chapter 2, cites Redish and Seizer (1985) and Ancker (2004) in saying:

> Often the mandate to write or revise text to meet formal readability criteria leaves writers and materials developers in a Catch-22, and can result in actually trying to game the system by artificially dividing sentences and using sentence fragments. Adding the very words or sentence types that would make the text more com-prehensible unhappily increases the readability score of the material, and thus is judged inappropriate. (p343)

5.6 Conclusions

My analysis has shown that four-word lexical bundles appear frequently in patient information, though none are used with anything like the frequency seen in conversation, where the most frequent individual bundles appear over 1,000 times per million words (Biber and Conrad,2005). In patient information, the most frequent bundles appear between 200 and 400 times per million words, a similar rate to that found in academic prose (Biber and Conrad, 2005). Similarities with the bundles found in academic discourse are also seen in the structural types of bundles in patient information. Two-thirds of these bundles are of the kind found more often in academic writing, and almost a third of the total number of bundles were accounted for by just one structure: Prep + NP fragment (e.g *at the end of*). With regards to the discourse function of the lexical bundles, they are split between referential and impersonal stance bundles, the latter predominating.

Referential bundles are frequent in academic discourse but also in informational discourse. It is the informational content in patient information that explains the preponderance of bundles that are common to academic prose. Indeed, the primary function of the majority of bundles in the corpus appears to be that of information-giving. In patient information, there is a premium put on transmitting practical information relating to the procedure, and on the patient's experience at the hospital. And where there is a precision seen in many of the referential bundles, particularly those that relate to place, many of these bundles are often uncertain: modified by *may*, or with imprecise temporal terms which may well be explained by the fact that a patient's experience of a medical intervention is highly individual, and the daily workflow of a hospital environment unpredictable.

It is also true that avoiding certainty in patient information may confer some legal protection for the hospital or healthcare system. This imprecision is seen in the Stance bundles too, many of which relate to Intention and Prediction. Interestingly, some of these Stance bundles are, in fact, functioning as instructions. *You will be asked* and *you may be asked* are two examples, and as we have seen, they are also the second and third most-used bundles in the entire corpus, highlighting the importance of instruction, alongside information, in patient materials.

And while instruction is one of the two primary functions of patient information, many of the bundles that function as obligations are not direct and do not come from the Obligation/Directive category. We have seen, too, that while Obligation/Directive bundles are relatively common in patient information, the vast majority of these bundles are not direct either. In fact, directness in patient information seems to be something to be avoided, particularly when telling the patient what to do. In my corpus, the majority of bundles in the Obligation/Directive category involved the use of *need to* and, to a lesser extent, *should*. These two modal verbs, along with others used to instruct, are the subject of the third analysis, reported on in the next chapter.

Indirectness is also achieved in patient information for radiography by using an impersonal structure such as *it is necessary to* or *it is very important.* The question is raised, however, of how a phrase like *it is important to* are perceived and understood by patients, for whom the reading ‘important for others but not me’ is always a possibility.

In terms of lexical characteristics, the lexical bundles analysis has provided clear evidence of the primary communicative concerns of patient information: information and instruction. We have seen, too, that the bundle types that are more often found in conversation, (pronoun) (AUX) + active verb, often contain modal auxiliary verbs such as *may*, and many references to time or experiences in the patient information are vague or imprecise. The pronoun used in the majority of the cases of this bundle type is ‘you’.

As discussed in Chapter 1, The NHS in the UK produces a guide to writing patient information and the use of *you* and *we* are encouraged. Bundles in patient information, though, very rarely make direct reference to the hospital or the medical system, and very rarely do they include *we*. The pronoun *you* appears 6,544 times (14,014.38 pmw), while *we* appears just 326 times (698 pmw). Some documents never use *we* at all. There are just two 4-word bundles in my analysis, *we may have to* and *we will ask you*, that use *we*, against 31, 28% of the total number of 4-word bundles, that use *you*. The focus, then, seems very much to be on the patient. This focus on the patient was also seen in the number of bundles that use a passive structure, putting the patient in the initial position, e.g. *you will be asked* instead of *we will ask you*. *You may be asked* and *you will be asked* are the first, and the third, most frequent 4-word bundle in my corpus. This finding seems as odd with the advice to patient information writers to avoid the passive, and I wonder whether the desire to focus on the patient at all costs may, unwittingly, be forcing the use of passive structures when an active structure, with *we*, would be a more natural choice.

This concludes my chapter on lexical bundles in patient information. I now focus my attention on one of the two primary functions of patient information, that of obliging and instructing the patient. How this is achieved through the use of modal verbs is the subject of the third analysis, reported in the following chapter.

6. Modal verbs as instructions in patient information

The purpose of patient information produced for radiography is generally twofold: to instruct and to inform. Patients, ideally, are given useful information about the medical procedure, told what might or will take place during the examination and, additionally, are told what is required or desired by the hospital before, during and after the procedure. (Patient Information Forum, 2013; Tutty & O'Connor, 1999).

In the analysis of lexical bundles in the preceding chapter, a number of bundles that function as instructions were revealed, e.g. *you will be asked to* and *it important that.* We saw, too, in chapter 4, that *please* is a keyword in patient information, and is used to preface an imperative, e.g. *please go to the main hospital reception desk*. Inviting contact is one of the uses of this structure, while its other principal use is to instruct the patient. Another finding from the lexical bundle analysis of relevance here is the frequent use, in patient information, of various structures with *need to* and *should*. *Need to* is classed as a semi-modal, and *should* a modal verb.

How these words are used in patient information, with what frequency and with what effect is the focus of this chapter, as is the use of the other modal verbs and semi-modals that are used in English to give instructions.

I will begin by explaining the role and importance of instructions in patient information in general, and my reasons for selecting modal verbs as the subject of my analysis. This will be followed by a section that presents an overview of modal verb meaning and some findings from the literature regarding frequency and use in different varieties of English. The methodology, which has been presented in full in chapter 3, will be summarised before I present my results and a discussion of those results. Let us begin by considering the role of instructions, and the importance of following instructions, in patient information.

6.1 The importance of instructions in patient information

When instructions are not followed, the patient is often said by medical professionals to be exhibiting 'non-compliance' or 'non-adherence'. The terms refer to two different kinds of behaviour, the latter suggestive of unintended consequences, the former more complex and intentional behaviour (Jones, 2013).

This is an important and much-researched healthcare topic which, while it is outside the scope of my research, is of great relevance when we consider the role language may have in explaining why patients intentionally, or unintentionally, fail to follow instructions.

Jin, Sklar, Min Sen Oh, & Chuen Li (2008) in their meta-analysis of studies of patient compliance (I will use the term compliance to include adherence in this chapter), demonstrated just how complex and varied the reasons are. Jin et al. (2008) identified as many as 25 factors that could affect compliance, though conflicting results in different studies they looked at suggest that the factors governing an individual's ability or willingness to comply are very complex.

Where study results were unequivocal, however, was in the area of the patient-provider relationship and communication. Studies consistently show that patients are far more likely to exhibit compliance when they feel that they are being treated as an equal partner, when there is empathy from a provider, and where patients exhibit higher levels of trust towards their provider (Jin et al., 2008, p277). Feeling informed contributes to this feeling of trust, and how medical professionals communicate with their patients also contributes to this trust and to the experience of being treated as an equal partner. In ideal patient-centred care, a patient needs to feel confident that they have a voice and that decision-making power is being held by both parties. (Patient Information Forum, 2013). Language is pivotal to the development and maintenance of this relationship, and this relates to both spoken language, as in a consultation or written language, as in a patient information leaflet.

People are told what to do by other people very frequently in certain settings: in the workplace, school and in healthcare interactions particularly, and this is a well-researched area in sociolinguistics (e.g. Holmes and Stubbe, 2003; Vine, 2004). Most of this research has focused on spoken interaction, while the language used for instructions and obligations in written registers has received scant attention in the literature. This absence in the literature is a motivating factor for me to focus on the instructions and obligations in patient information.

In the next section I explain why I elected to focus particularly on modal verbs for instructions, rather than any other means of instructing, such as imperatives.

6.2 Why modal verbs?

Overt instructions seem to be largely absent in patient information for radiography, based on my two analyses thus far: keywords and lexical bundles. My analysis of the latter, presented in the previous chapter, suggests that aside from an imperative with *please*, direct obligations and bald directives are not at a feature of the register. Appeals to the importance or necessity of something, e.g. *it is important to*, and *it is necessary to* are preferred, and I also found a number of bundles making use of *need to*. This semi-modal verb, along with the modal verb *should*, appeared in a number of very frequently used bundles, although their stronger counterparts, i.e. *must* and *have to*, did not. While the findings from my lexical bundle analysis has suggested this is an area worth further investigation, a targeted analysis will give us more detailed information.

Another reason for investigating modal verbs in patient information is that modal verbs are very common in medical writing in general. This is particularly true of epistemic modals, which are often used in hedges or boosters in research papers (e.g. Salager-Meyer, 1994), and of modal verbs of obligation, particularly *must* and *should* which have been found to be frequent in a number of medical registers, included case notes and editorials (Vihla, 1999). Vihla (1999) did not include patient information in her study of a range of modals in medical registers, however, and as I have stated already, I am not aware of any studies that have looked at instructions in patient information, or at modal verb use generally in patient information. The investigation described in this chapter is a response to some of this gap in the literature

In the next section I will present a summary of modal and semi-modal verbs in English, describing their range of meaning, their use in different modes and varieties of English, and the changes in use that have been documented by applied linguists.

6.3 Modals and semi-modals in English

Modality is the expression of possibility or necessity, and in English can be expressed by many means including modal verbs, semi-modal verbs, adjectives, nouns, adverbs, and particles. For the purposes of this study, the categories of auxiliary modal verb (also called *central or core* modals and one of the most common means to express modality) and semi-modal verbs (Palmer, 1983, p208) are considered.

The core modals are generally held to be *can, could, may, might, shall, should, will, would, ought (to)* and *need* (Downing and Locke, 1992; Quirk et al., 1985) while the category of semi-modals can include a range of items including *dare to, need to, have (got) to, be able to and be going to*. Semi-modals express meanings that can usually also be paraphrased with a core modal, e.g.*I have to lose weight*, and *I must lose weight.* (Biber et al., 1999). Some semi-modals, unlike modal verbs, can be marked for tense and person, e.g. *have (got) to.*

The main functions of modal and semi-modal verbs (henceforth modals) is to express stance. (Biber, Conrad & Leech, 2002). Modal meaning is usually categorised as epistemic, deontic and dynamic, though Biber et al. (1992) propose three other names for the categories: permission/ability; obligation/necessity and volition/prediction. Epistemic, deontic and dynamic categories of meaning I will present in 6.4.1.

The literature on modal verbs, their meaning and use, is huge and beyond the scope of this study, however. I am interested in how a small selection of modal verbs from one category of meaning (deontic) are used to give instructions in patient information, though some background information regarding meaning and modals is necessary for the sake of clarity. To this end, I will present an overview of modal meanings in the following section, with particular emphasis on the category of meaning under investigation.

6.3.1 Modal meanings

As we have seen in the preceding section, modal meanings fall into three categories: epistemic, deontic and dynamic. I will summarise epistemic and dynamic modal meaning first, before moving on to a more detailed consideration of deontic modality.

6.3.1.1 Epistemic modality

Epistemic modality is concerned with the speaker's attitude towards the proposition or the situation described in the proposition. This can range from an expression of doubt through to certainty. It is concerned with '…the speaker's assumptions or assessment of possibilities and, in most cases, it indicates the speaker's confidence (or lack of confidence) in the truth of the proposition expressed' (Coates, 1983, p18). Modals commonly used to express epistemic modality include *may* and *might*. E.g. *He may be the right man; it might be the right decision*. Studies of hedges and boosters in medical academic writing (e.g. Salager-Meyer, 1994; Skelton, 1997) are concerned with epistemic modality. Lexical bundles can also have an epistemic discourse function, as we saw in chapter 5.

6.3.1.2 Dynamic modality

Dynamic modality is less straightforward to characterise. *Will*, *would, can, shall* and *be going to* appear in this category and, broadly speaking, dynamic modality refers to ability or volition - though it, unlike deontic and epistemic modality, is not subjective (Palmer, 1990, p36) which suggests to some that it is not inherently modal. Gisborne (2007) says that *can*, when used dynamically, 'is not a modal meaning, but rather is simply the retention of an earlier sense which persists after CAN has joined the modal verb system of English (with similar arguments applying to WILL)' (2007, p45).

6.3.1.3 Deontic modality

Deontic modality is concerned with obligation, requirement and necessity. It is this category of modal that is the focus of this chapter. As we have seen, no studies have looked at the linguistic mechanisms of instruction and obligation in patient information, in spite of instruction being one of the two primary functions of patient information. English has a particularly wide range of deontic modal and semi-modals from which to choose, all of them evidencing different collocational behaviour and different connotations, including *must, should, need, need to, have to, have got to, ought to, have to* and *allowed to.* We have already seen a suggestion, in the finding of the lexical bundle analysis, that patient information may prefer *need to* and *should* over other deontic modal verbs. A targeted modal verb analysis will explore this finding more fully.

Deontic modality receives far less attention in the literature than epistemic modality, and when it does get any attention is, say Nuyts et al., (2005) it is ‘nearly exclusively as a ‘byproduct’ in the context of analyses of the formal category of the modal auxiliaries’ (p. 7). This imbalance of attention is another reason for investigating deontic modal verbs in my corpus.

Vihla (1999) uses the term ‘performative’ to refers to the function of deontic expressions (including modals) saying that ‘when using them, the speaker permits, demands, or forbids something, and they can be used prescriptively to create norms of action’ (p18). In the context of patient information, these norms of action might relate to behaviour around diet, lifestyle, drug or alcohol use, and equally to the behaviour expected in the context of a radiographic examination, i.e. to wear or not wear certain types of clothing, to eat or drink appropriately prior to an exam, and to inform the medical staff if pregnancy is suspected or allergies known about. The relationship of deontic expressions to norms of action had previously been stated by von Wright (1983), who says that deontic expressions ‘imply the existence of an authority having the power to say what is right or wrong, i.e. ‘norm authority’ (p68). The authority of the speaker over the addressee is a ‘felicity condition’ for deontic expressions, says Vihla (1999) if the authority does not exist the utterance is not regarded as a valid command, request or permission (Vihla 1999, p18).

The notion of authority is pertinent to this study. Patient information produced by hospitals and healthcare trusts exists to inform and instruct. Giving or denying permission to the patient to act in a certain way, telling the patient what to do and what is acceptable or otherwise are its primary functions. The authority in patient information can be the hospital named in the patient information leaflet, or a more generic authority, that of the medical system, of which the named hospital and its staff are a part.

Patient-centred medicine, however, has, as an objective, a rebalance of the power relations between provider and patient. Shared-decision making means, in theory, both patient and provider possess the authority to command, request or grant permission. According to Lindstrom and Weatherall (2015), both professionals *and* patients have what they refer to as 'deontic authority': medical professionals have the right to propose courses of treatment and behaviour while patients have the right to refuse to comply. Both epistemic and deontic authority plays a fundamental role in medical interactions and are 'complex and powerful structural forces scaffolding doctor-patient interactions and the ways treatments are recommended and responded to.' (Lindstrom & Weatherall, 2015, p51). While Lindstrom and Weatherall's work focuses on face-to-face consultations, written patient information is also concerned with recommending, proposing and outlining medical treatments and procedures and thus is very likely to demonstrate deontic authority. Quite how much deontic authority it demonstrates, and how this authority is realised linguistically is the objective of my analysis.

The results of the lexical bundle analysis presented in the previous chapter suggests that the authority of the hospital, the professional and medical system is not visible: only three lexical bundles that reference the authority using the pronoun 'we' were found, out of a total of 109 bundles. Bundles containing 'you' were 10 times as frequent. In the corpus itself, the pronoun 'you' predominates; it is used around 20 times more often than 'we' (6,544 occurrences of 'you' against 326 for 'we'). The question of how deontic authority can be expressed when the identity of the authority is unclear is a pertinent one.

We will return to the discussion of authority later in this chapter but let us now return to the subject of modal meaning. We have seen that modals can possess epistemic, dynamic or deontic meaning. Additionally, modals can also be used with two different types of meaning, which can mean modals can appear in different categories depending on the type of meaning being expressed. These types of meanings are usually referred to as personal (intrinsic) and logical (extrinsic).

6.3.2 Personal vs logical modal meaning

Personal (intrinsic) and logical (extrinsic) are two types of meaning that most modals possess. Personal (intrinsic) refers to the control of events and acts by human agents, with intention, volition, obligation and permission meanings. Logical (extrinsic) refers to the logical status of states or events. Logical modal meanings are necessity, certainty or likelihood (Biber et al., 2002, p. 176). The structure of the clause can usually indicate what meaning is being expressed by the modal. Personal or intrinsic meanings have two characteristics: the subject of the verb phrase is usually human, while the main verb is dynamic and references an event or activity that can be controlled. (Biber et al., 2002). *You can't sit there* and *John should ask for a raise* are examples of personal/intrinsic meanings. Logical meaning, on the other hand, usually has a non-human subject and/or a main verb that express states: *The photocopier can be found on the ground floor* and *That chicken should be done now* are examples of modal verbs used with dynamic meaning.

The deontic modals that I will be discussing in this chapter may be used with a personal meaning (obligation) or a logical meaning (necessity). As we have seen, English has a number of modals that can be used to tell people what to do: *must, should, have to, need to,* etc. Sometimes, of course, these modals are presented in the negative, when people are told what they cannot do.

When we talk about instructions and getting people to do things, the term 'directive' is sometimes used, particularly in studies from the fields of discourse analysis and pragmatics. I have chosen not to use the term, and in the following section I present an explanation of why.

6.4 Instruction, obligation or directive?

In speech act theory, a directive refers to an utterance that is used to get the addressee to do something. A directive can take many different forms, including that of a request, an invitation, a challenge, a threat and a direct obligation. Sometimes, a combination is possible. *You must eat with us* can be an invitation to dinner or be a direct obligation, most likely from a parent to child, perhaps: *No! You can't eat in front of the TV; you must eat with us at the table.*

We have already seen lexical bundles used in patient information that are classed as intention/prediction stance bundles but, on closer examination, are revealed to be functioning as instructions. *You may be asked* and *you will be asked*, the most frequent and the third-most frequent bundle in the corpus, are not referring to questions, or requests for information, e.g. *you will be asked about your hobbies/what you want for dinner,* but are used to refer to instructions that will be given once the patient is in the radiography department: *you will be asked to remove your clothing* or *you may be asked to drink a liquid.*

Notwithstanding the fact that many of the instructions I refer to in my data are also directives, to avoid confusion - for I am not directly referring to speech act theory or pragmatics in my study - I will refer to instruction or obligation when referring to an utterance that functions as an obligation, requirement or instruction, while deontic will be the term used for the modal or semi-modal that is used to express the obligation. I accept that instruction and obligation are not always the same thing, but both function to tell someone what to do. It is the telling-someone-what-to-do that interests me, though, for the sake of brevity, I refer to obligation or instruction in the text.

6.5 Frequency of modals in English

Dispersion and the frequency of a lexical or grammatical feature has long been held to be an important predictor of register variation (Biber 2012). Comparing modal verb frequency information from the literature with their frequency in patient information is a first step to describing the characteristics of the register.

Studies show that *will*, *would*, *can* and *could* are the most frequent modals in written English (Biber et al., 2002; Kennedy, 2002; Leech et al., 2009). Kennedy (2002) suggests these four modals account for as much as 72.6% of the modals in written English. Leech et al. (2009) found that in spite of an overall reduction in the number of modals being used in written British and American English between 1961 and 1991, *would*, *will*, *can* and *could* (in that order) were still the most commonly used.

Biber (2002) found that both central modals and semi-modals are more common in spoken language than in written academic prose. That semi-modals feature so much in spoken language is possibly less surprising than the fact that central modals also do as researchers had long thought the latter were more common in writing (Biber, 2002, p177). Not all modals are more common in spoken language, however. *May* is considerably more frequent in academic prose than in conversation (Biber, 2002, p. 177) while *must* and *should* are found slightly more frequently in academic prose.

It is worth remembering, however, that deontic modals are less frequent overall in general language than common epistemic and dynamic modals (e.g. *may*, *can*, *will*). Collins (2009) points out that only in the deontic category are semi-modals increasingly more frequent than core modals, evidence he says that semi-modals (referred to as quasi-modals) are 'regularly replacing their auxiliary counterparts'. (p. 33).

I will now present frequency information from the literature for deontic modals. I do not present detailed frequency information for epistemic or dynamic modals as they are not the focus of my study.

6.5.1. Frequency of deontic modals

The reported frequency of deontic modals varies across different corpora, depending on the mode of discourse (e.g. written or spoken) and the language variety (e.g. British English, Australian English, etc.). Generally speaking, *must* and *have to* appear with far greater frequency than *need to* and *have got to*, with the use of *have to* equalling or surpassing that of *must* in spoken corpora. The status of *must* as a direct obligation and the need to avoid a face-threatening act renders it unsuitable for most situations in spoken language.

Collins (2009) investigated modal verbs from all categories of meaning in three varieties of English: British, Australian and American. The corpora of 1 million words for British and Australian English were made up of spoken and written material, with a range of registers. The material was collected in the first half of the 1990s. The US corpus was smaller in size though contained a close match of document types. The figures in brackets relating to the US corpus are adjusted frequencies (to per million). His results of a frequency analysis of deontic modals can be seen below in Table 19.

Table 19 Deontic modal frequency in GB, Aus and US English (Collins, 2009)

		ICE-AUS	ICE-GB	C-US	TOTAL
Modals	*must*	613	675	402 (79)	1,690 (1,367)
	should	1,141	1,124	850 (167)	3,115 (2,432)
	ought to	36	80	51 (10)	167 (126)
	need	19	34	15 (3)	68 (56)
	Total	1,809	1,913	1,318 (259)	5,040 (3,981)
Quasi-modals	*have to*	1,311	1,244	1,385 (272)	3,940 (2,827)
	have got to	332	339	173 (34)	844 (705)
	need to	343	280	473 (93)	1,096 (716)
	had better	48	33	41 (8)	122 (89)
	be supposed to	47	99	127 (25)	273 (171)
	be to	135	221	76 (15)	432 (371)
	be bound to	9	17	5 (1)	31 (27)
	Total	2,225	2,233	2,280 (448)	6,738 (4,906)
Total		4,034	4,146	3,598 (707)	11,778 (8,887)

I will discuss some of the changes in modal use over time in the next section, and we must bear in mind that the materials contained in the corpora investigated by Collins (2009) date to the early 90s, but, nonetheless, Collins's study produced some very interesting and useful findings. The table shows a strong preference for certain modals over others, in all varieties of English. The most frequent in all varieties is a semi-modal: *have to.* This is closely followed by *should*, which appears to be more frequent in both British and Australian English (though the corpus of US English was considerably smaller in size and frequencies are adjusted.) The frequency data relating to *must* and *need to* are particularly interesting, and as we shall see, of relevance to my analysis of patient information.

Must is used with more or less equal frequency in British and Australian English. In US English, however, *must* is used marginally less often than *need to*. In the GB and AUS data, *need to* is used, but at a similar rate of frequency to *have got to*, and half as frequently as *must*. Language change is often seen in US English before appearing - if it appears at all - in other varieties of English. Collins's (2009) findings seen in Table 18 suggests that *must* has been eclipsed by *have to* in all three varieties of English, and in the early 1990s was in the process of being replaced by *need to* in US English. More about the changes in modal use in the US are reported in the next section.

6.5.2 Diachronic change in deontic modal use

Our use of modal verbs has changed over time and continues to change. Johansson (2010) used the COCA (Corpus of Contemporary American) corpus (Davies, 2009) to investigate changes in *must*, *have to, need to* and *have got to* from 1990 through to 2008, providing, perhaps, some of the answers to the questions raised by Collins's (2009) study. Johansson (2010) confirmed that *must* is gradually falling in frequency, while *have to*, the most frequent deontic modal by far also seems to be dropping off. *Need to*, on the other hand, is rising steadily in use.

I used the COCA corpus to look at the use of *must*, *need to* and *have to* in 2017 and found that the frequency rates remain very similar: *have to* is still the most frequent deontic modal, and *must* and *need to* are used at a similar rate. I found that *should*, which was not considered by Johansson (2010), is used marginally less than *have to* in the COCA but more frequently than *must* and *need to*. Without knowing how Johansson's (2010) search was conducted, or how the frequency figures were treated (rounded up or down) a true comparison cannot be made, however.

Some of the reasons for the increase in use of *need to* and the fall off of *must* may well relate to the use of the *need to* as a democratic, non-threatening term (Nokkonen, 2006, p46). Smith (2003) found that *need to* was used around 130% more in written American English over the course of 3 decades (between approximately 1960-1990) and 249% more in British writing, while in spoken British English its use increased by more than 600% in the same period. '*Need to* grows in use in all syntactic environments, and in some of these it is likely to be a competitor with *must* and *have to*' said Smith (2003, p255) who adds that *need to* 'can acquire the force of an imposed obligation -something that does not happen with other markers - the writer or speaker can claim that the recommended action is merely being recommended for the doer's own sake (2003, p260). Medical advice may well be the kind of recommended action that Smith has in mind.

Should seems to be consistently frequent in studies of modal verbs (e.g. Collins, 2009). Leech (2004) suggests that *should* is less categorical than *must* in both obligation and logical necessity and this obligation can be reduced to 'something like desirability'. (p158). Nokkonen (2006) says that *should* 'gives the impression that the speaker is appealing to the assumed needs of the addressee' (p64) which suggests that any advice or perceived obligation is principally for the good of the recipient. This has also been said about *need to* as we have seen and suggests that we may see a high rate of occurrence of *should* and *need to* in patient information. Health advice is generally offered, after all, for the good of the patient recipient. I expect, too, to see *have to* used frequently, based on the findings of studies of modal use reported in this section (Collins, 2009; Johansson, 2010; Smith, 2003).

At the outset of my doctoral investigations, I was unsure what to expect with regards to the frequency of *must* in patient information. Collins (2009) found *must* to be used slightly more frequently than *need to* in GB and Australian English, though *need to* was preferred in the American data. The change in frequency of *must*, however, is reported as gradual by Johnsson (2010) and the drop off in use is greater in spoken language, where it retains a strength that many users would find inappropriate. Patient information is written material produced by a medical authority, however, and as a result, might be considered fairly formal in style. In spite of the simplified language and question-answer format that is intended to simulate a conversation, my analysis of lexical bundles revealed that two-thirds of the 4-word bundles extracted were of a structural type more often seen in academic prose, with one-third conversational bundle types. Perhaps *must* would be retained as the deontic modal of choice in patient information.

This concludes my presentation and discussion of the frequency of deontic modals in British, Australian and American English and the changes in use over time, as reported in the literature. I now turn to the methodology employed in my analysis, presenting a summary of the key steps. The methodology has been presented in full in chapter 3. This is followed by a presentation of the results of the different analyses undertaken and a discussion of the key findings.

6.6 Methodology

The methodology of this corpus-based analysis has been presented in detail in chapter 3, though in this section I re-present the key steps. I will first present the corpora used in the analysis.

6.6.1 Patient information corpus

The 408, 997-word corpus of patient information was made up 221 downloadable patient information leaflets, sourced from three principal organisations: the NHS, the Royal College of Radiographers, and RadiologyInfo.com, a website associated with the Radiologic Society of North America (RSNA). The corpus was first compiled in 2011, with later additions in 2014 and 2016. Both diagnostic and therapeutic procedures were included, though the majority of documents relate to diagnostic exams, and medical procedures involving the use of radiographic technologies.

Two further corpora were also used to provide a comparison of the frequencies of deontic modals. These are described below.

6.6.2 Comparative corpus 1: consumer advice

This was a small, 104,670-word corpus of consumer information, with material from both the UK and the US. The inclusion criteria for the consumer information corpus was very close to that for the patient information corpus. All texts were available as Word or pdf documents on the Citizens Advice website, a recognised authority in the UK for consumer information, and its US equivalent, the Federal Trade Commission (https://www.ftc.gov/). This was a much smaller corpus than the patient information, however, with a word count of just 104,670 and the majority of texts came from the UK Citizens Advice site (https://www.citizensadvice.org.uk). The topics covered included housing, health, children, consumer topics and the law. The length of the documents included in this corpus also varied, from the longest at over 7,000 words to the shortest at under 300 words. The longest documents in this corpus were from the UK, unlike those in the Patient Information corpus, where we saw that the longest documents were US-sourced.

By comparing the frequencies of deontic modal verbs in consumer advice with those found in patient information, I wanted to see if patient information resembled consumer information, particularly in light of the fact that patients are increasingly referred to as consumer or clients. The latter is very much focused on consumer rights in the law. Patient information is, in itself, a right; the right to be informed of healthcare-related events. The information confers on the patient the ability, in theory, to also be involved in decisions relating to healthcare and treatment by virtue of being informed. And while we are led to believe the customer is always right, a position that the law can uphold, we never hear the same said about patients. On the basis of this, my hypothesis was that deontic modal verb use would not be the same, as there are fundamental differences in the functions of healthcare information and consumer information. This comparison I hoped would add to my growing understanding of the lexical characteristics of patient information.

6.6.3 Comparative corpus 2: General radiography

The 719,209-word corpus of General radiography is made up of a radiographer handbook, *Clark's Positioning in Radiography*, a training textbook, *Patient Care for Radiography* and research from *Radiography*, a peer-reviewed journal of the Society and College of Radiographers and the European Federation of Radiographer Societies. The research was included because it is written for and by radiographers, and not radiologists, and thus deals with the issues that are relevant for radiographers: patient safety, radiation dose, patient position and workflow, for example. Radiologists, on the other hand, are doctors who specialise in radiology. Their job is to diagnose and propose treatment. Radiographers, on the other hand, are the healthcare professionals who carry out radiographic examinations.

By comparing the frequency rates of deontic modals in patient information for radiography with those found in other radiographic registers - textbooks, manuals, and research papers - I thought I would be better able to characterise some of the uses of deontic modal verbs specific to patient information, rather than specific to the field radiography.

6.6.4 Search criteria

Sketch Engine was the software used in this analysis. The steps taken, when compared to those of the keyword extraction and the lexical bundle analysis, were straightforward. In Sketch Engine, I used the so-called Simple search, which is cleverer than the name implies. The software works out what it is you are looking for based on the kind of search term you have entered. If you enter a lemma, the search is a lemma search, meaning *go* will also find *goes*, *going*. If you enter a term which is not a lemma, the software will search only for that word. It was the Simple search that I used for the modal and semi-modals under investigation, and the software treated my search terms as lemmas. This meant that the Simple search captured the changes for person and tense that took place with three of the semi-modals (*need to, have to* and *have got t*o), as well as including negative forms for all modals and semi-modals. It did so quickly and effectively.

There were three distinct steps in my analysis which I summarise in the follow sections.

6.6.5 Methodological steps 1-3

6.6.5.1 Step 1: General frequency rates of modals in patient information

In order to find out how the frequency of modal verbs in general compared with what we know of their frequency in general English, a summary of which I presented in 6.4, I carried out a Simple search in the patient information corpus of the following modal and semi-modal verbs: *can, could, will, would, may, might, must, have to, should, have got to, need to, need, ought to, be allowed* and *be supposed to.* Raw frequency counts and their adjusted frequencies in per million words were noted.

6.6.5.2 Step 2: Deontic modals frequency rates in all three corpora: patient information compared with consumer advice and general radiography

The deontic verbs from the list in 6.6.5.1 were searched, using the same Simple search described above, in all three corpora. Results were compared using raw frequency counts and adjusted frequency.

6.6.5.3 Step 3: Deontic modals in patient information investigation

Four deontic modals, found to be the most significant in patient information, were investigated qualitatively in the corpus: *have to, must, need to* and *should*. This was done by investigating a sample of 100 examples of the modal verb in context. This facility is available in Sketch Engine, with a default setting of 250. I chose 100 to reduce the amount of data.

6.7 Results

The first analysis was to investigate the frequency and the use of a range of common modal and semi-modal verbs in patient information, not only deontic modals but epistemic and dynamic modals. The results are seen in Figure 9 below. The figures are adjusted to per million words. Adjustment was necessary as the three corpora were of entirely different sizes.

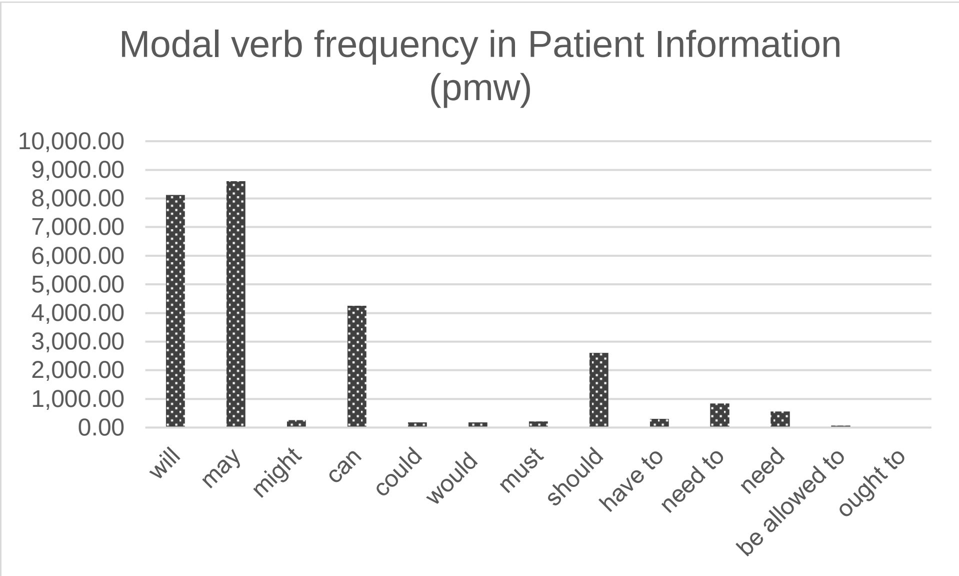


Figure 8 Frequency of common modal and semi-modal verbs in patient information

May, *will*, *can* and *should* are the most frequent modal verbs in patient information, in descending order of frequency. This is followed by *need to* and *need*. *Would* and *could* are very infrequent, as is *might* and *must*.

With regard to the second analysis, the frequency of deontic modals in the three corpora, the corpus of patient information, the corpus of consumer information and the corpus of general radiography, the results are shown in Figure 10 below. *Had better*, *ought to* and *have got to* resulted in zero or a single count (*ought to* in the corpus of patient information) and thus were not included in the graph.

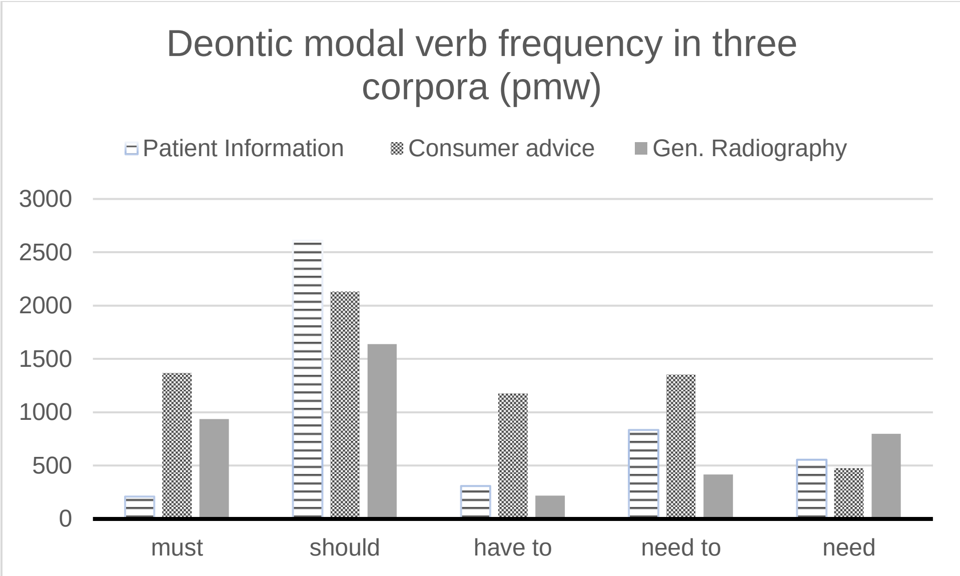


Figure 9 Frequency of deontic modal verbs in the three corpora

The same data is also shown Table 20, to allow for easier comparison, giving the raw occurrences in brackets following the adjusted frequencies to per million words. The corpus with the highest frequency of the modal is shaded.

Table 20 Deontic modal verb frequency in the three corpora

Modal verb	Patient Information	Consumer advice	Gen. Radiography
must	209.87 (98)	1368.78 (172)	936.99 (674)
should	2610.56 (1219)	2132.76 (268)	1639.04 (1179)
have to	308.38 (144)	1177.79 (148)	218.87 (156)
need to	835.21 (390)	1352.87 (170)	417.06 (300)
need	556.40 (260)	477.22 (223)	797.57 (574)

6.8 Discussion

6.8.1 General frequency of modal verbs in patient information

I will begin this section by discussing the findings of the first analysis, the use and frequency of common modal verbs, in patient information. This analysis was carried out to give me an idea of how patient information reflected modal verb use in general English.

It is immediately obvious from the initial frequency analysis, shown in Figure 9, that modal use in patient information differs from the frequencies reported in studies of general and academic English, where *will*, *would*, *can* and *could* are the most frequent modals, in both spoken and written discourse. (Biber et al., 2002; Kennedy 2002; Leech et al., 2009). In my study of patient information, *may* is the most frequent modal, with an adjusted frequency of 8,609 per million. In the Longman Spoken and Written English (LSWE) corpus, as reported by Biber et al. (2002), *may* is used around 1000 times per million, putting the very high frequency of this modal verb in patient information into perspective. What might be the explanation?

We have seen in our discussion of lexical bundles in chapter 5 that there are a lot of vague and imprecise references in patient information, particularly related to time. Medicine itself is sometimes very vague, as so little can be predicted with any great certainty and *may* also appears to be very common in some medical writing: Vihla (1999) found *may* to be the most common modal in a number of the registers she examined. We have also seen that the threat of legal action is never too far away in modern medicine: *may* confers a legal advantage over the surety offered by *will*. Without examining the use of *may* in detail in patient information, however - and *may* is not the focus of my study - we cannot be sure how the modal is being used. It would undoubtedly make an interesting future study.

The next most frequent modal in patient information is *will*, close behind at 8,122 per million. This is also used at a very high rate of frequency in patient information. *Will* is the most frequent in general academic English and conversation, though Biber et al. (2002) report a frequency rate in the LSWE of just over 3,500 per million, less than half the rate in patient information. *May* and *will* together occur more frequently than all of the nine modal verbs in the LSWE corpus combined, as reported by Biber et al. (2002). Looking at my data again, it seems that a few modal verbs are being used with great frequency (*may, will, can, should*) and an equal number used barely at all (*might, could, would, must* and *have to*). Between the two extremes we find *need to* and *need,* which are used with reasonable frequency, and to which I will return later in this discussion section.

Could and *might* are often used in academic prose and conversation to mark logical possibility, along with *may*, but are barely used in patient information. *Could* appears at a rate of 182 per million, while *might* occurs at an adjusted rate of 263 per million words. *Could* and *might* usually express doubt, as does *may*, though *could* and *might* seem more tentative. (Biber et al., 2002). Perhaps patient information writers, though happy to use epistemic *may* with great repetitive frequency, do not wish to sound overly tentative and thus avoid *could* and *might*. An alternative explanation might be that *could* and *might* are victims of the message simplification that patient information is subject to: after all, why use *could* and *might* when you can repeat *may*?

Message simplification does not seem a likely explanation for the absence of *must* and *have to* in patient information, however. The very low rate of frequency, particularly of *have to*, is especially interesting when we remember the reported rates of *have to* and *must* in Collins (2009) and Johansson (2010) reported in 6.5.1 and 6.5.2: *have to* was by far the most frequent deontic modal in both studies. Why are these modal verbs not being used in patient information when they are so common in many other registers, including consumer advice, radiography textbooks and radiography research? I will return to this question later in this section, as the use - or not - of deontic verbs in patient information is the central focus of this chapter and will be explored in more depth.

Returning to the general frequency of modal verbs in patient information, *can* is the next most frequent in patient information, but at 4,253 per million, it occurs at around half the rate of *may* and *will*. The fourth most frequent is neither *would* nor *could*, which barely feature in patient information, but *should*, the first of our deontic modal verbs. *Should* is used in patient information at a rate of 2,610 per million. Biber (2002) finds *should* as common as *must* in academic writing, and more common than *must* in conversation: in patient information, however, *should* occurs more than ten times as frequently as *must*, which appears at a rate of 209.8 per million. In Collins's (2009), the results of which were shown in Figure 8, *should* appears with about twice the frequency of *must* in all three varieties of English studied.

In patient information, modal verb use and frequency patterns do not resemble those seen in general English. A small number of modal verbs are used with very high rates of frequency in the register, while *could* and *might* along with hypothetical *would*, all very common in general academic English and conversation, are used with very low frequency rates in patient information.

Let us turn now to a discussion of the findings of the second analysis, a comparison of the frequency of deontic modals in the three corpora built for this doctoral thesis: patient information, consumer advice and general radiography.

6.8.2 Deontic modals in consumer advice and general radiography

Consumer information differs from patient information in that its primary purpose seems to be informing consumers of their legal rights. Obligations are most frequently referred to in consumer advice when they are legal obligations, though even here, the emphasis is very much on the legal rights - the possibilities within the law - that the consumer has. The emphasis given over to rights can be seen in the repetition of the word *rights* and the categories that greet a visitor to the Citizens Advice UK website[6], probably the best-known consumer advice association in the United Kingdom. The Health section homepage refers to a consumer's rights to healthcare on the NHS; their rights to dental care when abroad; the rights to healthcare for people resident abroad; how to report discrimination and how to complain.

We saw that patient information does not use *could*, *might, have to* or *must* at all frequently. In consumer advice, however, they are all used with more or less with equal frequency, at around 1300 times per million. *Should* is used most frequently at 2132 per million. *Need to* and *must* are used with very similar frequencies in

---

[6] https://www.citizensadvice.org.uk/

consumer information, at around 1350 per million, while *have to* is just a little less frequent at around 1117 per million. In short, a wide variety of deontic modal verbs is used in consumer information and used with high rates of frequency.

In consumer advice corpus, more than 90% of the uses of *have to* involve the 2nd person pronoun. The modal is generally used to refer to obligations, legal or procedural, or to give what amounts to instructions.

(1) *You'll probably **have to** pay a fee to cancel a contract if you've decided you don't want it anymore.*

(2) *You can't get your service provider to chase someone who's not named on the bill - you'll **have to** get the money from them yourself.*

The frequency of *have to* in consumer information is more in line with reported studies of modal verb use in UK and US English, and there is variety shown that patient information seems to be lacking. I have already referred to the emphasis on legal rights and obligations in consumer information; *must* is generally used in the context of laws, requirements and other enforceable aspects.

(3) *To qualify they **must** meet all of the following criteria:*

(4) *Landlords **must** tell you if they will not rent to you because of information in your credit report or background report.*

(5) *The law which says you **mustn't** be discriminated against is called the Equality Act 2010.*

Unlike patient information, one of the primary functions of consumer advice is to tell consumers about their legal rights. In nearly every case in my corpus, when the topic is the law, *must* is the modal verb used. While patient information does concern itself with the law, it does so in the form of a disclaimer, as a form of protection for the hospital or medical system against the patient; consumer advice, on the other hand, transmits information about the legal rights of its readers. This is a significant difference.

Before I move to an in-depth discussion of the deontic modals in patient information, let us see how the frequency of deontic modal verbs in the general radiography corpus compares. The data is reported above in Table 19.

The first thing we see is that there is no one modal that is used with greater frequency in this corpus. *Should* is the most frequent deontic modal, but it occurs less frequently in this corpus than in the other two. *Have to*, in complete contrast to its frequency in consumer advice, is used even less often here than it is in patient information, at little more than 218 times per million words. This is perhaps less surprising when we remember that a large part of the general radiography corpus is made of research papers from the *Radiography* journal, where the informality of *have to* would be inappropriate. *Must* is four times as frequent here than it is in patient information, though less frequent than in consumer advice. Inspecting the uses in the corpus, I found that *must* in the general radiography corpus is not concerned with legal rights, as it is consumer advice, and is rarely used to express an obligation; rather, it is overwhelmingly related to logical necessity, as the following examples show:

(6) *The central ray* ***must*** *pass through the joint space at 90 degrees to the humerus, i.e. the epicondyles should be superimposed*

(7) *The examination* ***must not*** *proceed unless the radiographer is sure of the identity of the patient*

(8) *The patient* ***must*** *be monitored with a pulse oximeter during the procedure, because it is not possible to monitor the patient directly.*

Aside from radiography research, the general radiography corpus is made up of a radiographer’s manual related to patient positioning, and a training textbook, with a large number of instructions, many of which relate to successful and safe imaging. They are, as a result, instructions presented as necessities.

Let us now move to a more detailed look at the deontic modal verbs as they appear in patient information for radiography. I will take each modal verb, being with *need to*, which is followed by *need*. I will then discuss the use of *must* and *have to* before turning to a discussion of *should*, the most frequent deontic modal in my corpus of patient information. Before the chapter concludes, I present the findings from a survey of modal verb use that I carried out in the course of this doctoral research.

6.9 Deontic modal verbs in patient information

6.9.1 *Need to*

Need to is the second-most frequent modal with a deontic meaning in the patient information corpus, and is used three times as frequently as *must* and two and a half times more frequently than *have to*.

The appearance of *need to* in my corpus may be explained by the fact that, as was suggested by Johansson (2010) earlier in this chapter, *need to* has seen an enormous rise in use in spoken English over the last few decades. Smith (2003) found that *need to* was used around 130% more in written American English over the course of 3 decades (between approximately 1960-1990) and 249% more in British writing, while in spoken British English its use increased by more than 600% in the same period. '*Need to* grows in use in all syntactic environments, and in some of these it is likely to be a competitor with *must* and *have to*'' said Smith (2003, p255) who adds that need to 'can acquire the force of an imposed obligation -something that does not happen with other markers - the writer or speaker can claim that the recommended action is merely being recommended for the doer's own sake' (2003, p260). Medical action may well be the kind of recommended action that Smith has in mind.

The most common forms of *need to* in the literature do not correspond to that used in patient information, however. Smith (2003, p. 261) reports that 1st person plural (we) and passivized 3rd person are far and away the most common grammatical subject. Examining my data, it transpires that in patient information, the 2nd person is far and away the most common grammatical subject: 75% of structures involving subject pronouns used *you*, which was used exclusively in the singular:

(9) *Lose weight if you **need to**.*

(10) *You will **need to** stay in bed for two to four hours.*

(11) *You **need to** drink the contents of this bottle.*

The 1st person plural, *we*, was barely used with this structure and appeared just five times in the entire corpus. These structures, then, are generally addressing the patient and, as Smith (2003) suggests, are often imposing an obligation on the patient, though that obligation is within the context of medicine, and thus for the patient's own good.

The use of *need to* structures in the passive were also a feature - around one quarter were of this kind. either using the semi-modal in the passive or where the following verb was passivised. Most of these structures involve medical procedures or treatment,

(12) *These injections may need to be given several times a day.*

(13) *Certain foods and medications may need to be avoided prior to taking the test.*

We have seen already that a number of lexical bundles were in the passive form and I have suggested that the focus on the patient in printed information sometimes forces a passive when an active structure would be more appropriate. In (12) and (13) above, there seems no explanation for choosing a passive over an active sentence so the surprisingly high use of passive structures may be explained by patient information writers failing to follow guidelines that encourage not to use the passive.

In (12) and (13), and most of the examples of passivised *need to,* the modal is not being used as an obligation but to express what Collins refers to as dynamic necessity (2009) or what Biber et al. (2002) call logical necessity. There is a sense that the situations referred to are, by being a requirement of an objective medical procedure, outside of the control of either the patient or the health professional. This fits with the notion expressed by Collins (2009, p74) that 3rd person deontic uses of *need to* are very often expressions of institutional requirement.

The idea of institutional requirement is also expressed by the core modal *need*, which appears at a surprising rate in patient information, almost three times as often as *must*, and twice as often as *have to,* which is, as we have seen, the most frequent modal in general spoken English after *should*. *Need* was also very frequent in the corpus of general Radiography, which is made up of manuals, textbooks and research paper. The very high rate of use of *need* in patient information compares to its very infrequent use in Collins' (2010) analysis, seen in Figure 8 in this chapter. Leech (2003) and Smith (2003) refer to the huge decline in *need*, which seems to be matched by an equally large increase in the use of *need to.*

Although not strictly a deontic modal as the others in this chapter, *need* is relevant to my study as it is used in patient information to present requirements. These requirements are either personal to the patient, as in (14), a general medical requirement (15), or a requirement specific to the radiography appointment.

(14) *If you **need** an interpreter or information about your care in a different language or format, please get in touch*

(15) *In extremely rare cases, surgery may **be needed***

(16) *The images from the scan will **need** careful analysis by our staff*

Returning to *need to* for a moment, in terms of its behaviour, *need to* is not immune to subjectivity, and it is with the 2nd person singular that we see *need to* at its strongest. Collins (2009) says that 'in the contexts where there is an obvious authority structure, the utterance will have the force of a directive' (p73). The uses of *need to* in the patient information corpus are often ambiguous, however, and the vast majority are related very closely to aspects of the medical procedure to be undertaken, even when used with a 2nd person pronoun. There are very few that might be interpreted to be a strong directive.

And while there can be very strong exhortations made with the 'I / we need you to…' structure, there were only two examples in the entire corpus, and both found in the same document (and thus almost certainly written by the same person). Johansson (2010) found that the 'I/we need you to…' structure, though small in absolute frequencies, was consistently growing in use and in the COCA corpus was most common in the Spoken and Fiction genres. There is a strength and non-compromising authority to this structure (that Yagoda (2006) refers to as the 'kindergarten imperative') that seems very out-of-place in patient information, which, as we have seen avoid overt references to obligation and authority. However, as linguistic innovations are seen first in spoken discourse, perhaps it is only a question of time before this structure becomes more acceptable (and less face-threatening) in written discourse.

A number of studies have found that *need to* tends to be used by those in positions of authority (Glass, 2015; Nokkenen 2006; 2012), such as teachers talking to students and bosses to workers. It is, at the same time, very frequent among teenagers (Nokkonen, 2006, p46) suggesting that it is also considered to be a democratic, non-threatening term. *Need to* is also used to appeal to the needs of the addressee and to issue an instruction or obligation in an objective, polite way (Smith, 2003).

While its multiple uses might explain its rise in popularity, the question is raised of how *need to* can be at the same time authoritative, democratic, directing, *and* polite? Glass (2015) believes the relationship between the speaker and addressee is fundamental to understanding the use and meaning of *need to* and this relationship (or perceived relationship) explains why the use of *need to* is neither monolithic or consistent across contexts. In some contexts, *need to* can come across as bossy; in others, it can appeal to external, objective needs. In (17), these needs relate to information regarding risks. This sentence has a very bossy, we-know-best-tone about it, which may be because the subject is risk information. In my data, reference to risk and complications were often presented in this way, as if the patient was being told to take their fingers out of their ears. The right to know is at the centre of patient centred medicine, but the right not to know also exists, and is often an ethical dilemma. Perhaps it is precisely because of the ethical considerations that I react negatively to the use of *need to* in (17).

(17) There are, however, several possible risks and complications. These are very unlikely, but possible. You **need to** know about them just in case they happen.

In the following two examples, *need to* is used with another instruction, an imperative in (19) and reference to future instruction, in (18). There is no doubt that *need to* is used to tell the patient what to do in these sentences, but the tone is quite different from that in (17). These are obligations presented as being for the good of the patient (Smith, 2003)

(18) *You will be taken to the recovery area where you will **need to** stay in bed for two to three hours, or as instructed by your nurse.*

(19) *Please bring an overnight bag with you to hospital as you may **need to** stay overnight*

By referencing the medical procedure in the majority of cases, and not using overtly direct structures (e.g. *we need you to*), patient information avoids a subjective or an overly authoritative tone when instructing.

Let us now turn to *must* and *have to,* which, as we have seen, were very infrequent in patient information.

6.9.2 *Must* and *have to*

Have to, a semi-modal and one that has been reported to have increased in use in both written and spoken English as *must* has declined (Collins, 2009, p67), represented less than 10% of the deontic modal tokens used. Surprisingly, *must* was used even less frequently. These results could well be evidence that in some registers Smith's (2003) prediction of the increasing status of *need to* viz *have to/must* is correct. *Need to* appears three times as often as both *must* and *have to* in my corpus of patient information.

Many of the uses of *must* in the patient information corpora refer to the medical procedure (as seen with *need to*), but in addition to patient preparation and communication, before and after the procedure:

(20) *If you have any allergies you **must** let your doctor know.*

(21) *If you have any allergies or have previously had a reaction to the dye (contrast agent), you **must** tell the radiology staff before you have the test.*

In contrast to *must*, however, *have to* is not used on any occasion in the patient information corpus to directly tell the patient what to do. It is not used to issue a directive or obligation but seems to refer more to procedural necessities - and often to the *possibility* of procedural necessities. It seems to be used to mitigate necessity - at times almost apologetically as we can see in the examples below:

(22) *If you **have to** undress for the procedure, you will be shown to a private cubicle where you will be asked to put on the gown provided.*

(23) *If you are given fluid to drink on arrival, you might **have to** wait an hour before entering the scanning room.*

(24) *As treatment progresses, you may find you **have to** pass urine more frequently*

(25) *It may be possible to perform the scan without you **having to** change your clothes*

(26) *We recommend that you do not wear jewellery as we may **have to** ask you to remove this during your examination*

The grammatical flexibility of *have to* can be seen in the use of other modal modifiers, usually *may* and *will* which are used in around one-third of the occurrences of *have to,* though there were very few examples of negatives with *must* or *have to*. *Must not* was used as an obligation just 6 times in the entire corpus, e.g. *you must not drive*, and all examples came from the American material. *Do not have to* was used just twice in the British material and 13 times, to refer to the same thing, in the US materials. In the UK, guidelines (NHS Toolkit, 2003) for the writers of patient information frequently state that negatives are to be avoided; the data in this thesis strongly suggests that these guidelines are being followed.

Given the very infrequent use of *you* + *must*, (at around 115 pmw compared to *you* + *should* at 798 pmw), it is evident that this is not a preferred modal for giving instructions in patient information. *Must is* used with great frequency in radiography textbooks and handbooks, however, evidence that it still has a place in certain types of discourse - the discourse where the relationships between the participants (addresser and addressee) are not considered equal or where politeness is not something to be concerned about. In the case of textbooks and handbooks, the relationships are teacher/learner or expert/trainee and here *must* is seven times as frequent than it is in patient information.

If *must* is rarely used to issues obligations in patient information, *have to* never used for obligations, and *need to* refers primarily to procedural necessities, it appears that *should* is the modal verb relied upon to issues instructions and to refer to behavioural obligations in patient information.

6.9.3 *Should*

In patient information *should* appeared 1,219 times, (2,610.56 per million). This is a frequency rate that is more than six times that of as *have to*, more than twice that of *need to* and a whopping nine times that of *must*. Its proportional frequency in this study is similar to the studies that have considered general and academic English (which find *should* used as often as *must* in academic writing and more often than *must* in conversations) (Biber 2002).

Leech (2004) suggests that *should* is less categorical than *must* in both obligation and logical necessity and this obligation can be reduced to '…something like desirability' (p. 158). Nokkonen says that *should* 'gives the impression that the speaker is appealing to the assumed needs of the addressee' (2006, p64) which suggests that any advice or perceived obligation is principally for the good of the recipient. This has also been said about *need to* as we have seen earlier in this chapter.

Myhill (1995) considers *should* to be an individually-oriented modal in contrast to *ought to*, which he calls group-oriented, along with deontic *must* and intentional *will* (1996, p339). *Have to*, in contrast, he refers to as an objective modal and one which denies any personal involvement; the increasing tendency to avoid overt claims to authority by the speaker or writer favours *should* (weak obligation) over *must* (strong obligation) and explains the rise in use of *should* and *have to* (p 339) - though in patient information, as we have seen, it is *need to*, not *have to* that predominates.

There seem to be few differences in the use of *should* in British and Australian varieties of English and it is the deontic use that predominates, though *should* is less used in US English (Collins, 2009) in both spoken and written discourse. It is also the case that the deontic use of *should* is more frequent in written discourse (Collins, 2009). I will discuss the meanings of *should* as they are used in patient information later in this section.

In the case of written patient information, the authority is not personal but is the authority of the medical system, the hospital and the team of professionals. The use of ‘we’ as a subject pronoun always references this authority in the corpus and is not a ‘we’ that includes the patient, with a handful of exceptions when ‘we' means ‘we humans,' e.g. *the air we breathe*. However, as discussed in chapter 5, ‘we’ is noticeable by its absence in patient information. There are just 326 examples (698.15 per million) of ‘we’ in the entire corpus. This compares to 6,544 (14,014.38 per million) for ‘you’. Nor are references to the hospital evident: There are 100 (214.16 per million).

My keyword analysis did suggest, however, that the authority being referenced in patient information are certain medical professionals: the radiologist, the doctor and the physician. As I discussed in chapter 4, these professionals predominate in patient information, while radiographer and nurse - the professionals with less status - are far less frequently referred to. The authority obliging the patient with *should* in patient information, then, seems to be the age-old, elevated authority of the doctor, in contrast to the image of a democratic, modern and equitable health system suggested by the term patient-centred care.

We have seen that modal verbs possess different meanings and *should* is no exception. In fact, *should* possesses a range of meanings, some of which are very evident in patient information.

6.9.3.1 Meanings of should in patient information

Should generally possesses two main meanings: one is epistemic in nature and refers to the likelihood of something happening. The other meaning, the most common, is deontic, and refers to the ‘desirability’ of something, though the strength of this desirability ranges from an obligation, ‘do (not do) this please’ through to ‘it would be a good idea if’.

To understand how the 1219 examples of *should* were being used in my data, I sampled 150. Though not a fine-grained analysis, sampling in this way gives an idea of proportions, which can be very helpful. The distribution of the meanings of *should* in patient information can be seen below in Table 21.

Table 21 Distribution of epistemic and deontic should in patient information

Uses of *should* in patient information (1,219 /2,610.56 per million)		
Epistemic (Likelihood)	Deontic (Obligation)	Deontic (Desirable)
31 (66.55)	108 (231.87)	11 (23.54)

Examples from the data of the different uses can be seen below. Epistemic uses can be seen in (27) and (28).

(27) *This should not last more than a few hours.*

(28) *It should not be painful and will heal*

Negative obligations and obligations can be seen in (29) - (31).

(29) *Patients with epidural electrodes should NOT have an MRI.*

(30) *You should not have a bone scan if you are pregnant or think you might be pregnant.*

(31) *You should tell the radiographers if you have ever had an allergic reaction to iodine.*

In (32) and (33), *should* is used with the sense of desirability.

(32) *A prevention plan should be discussed with your doctor*

(33) *Ideally all diabetic patients should be given an early morning appointment*

The desirability of the utterance in (33) is made clearer by the adverb *ideally*, though (32) contains no such adverb and requires reader interpretation. Reader interpretation, however, is not always reliable as it requires a certain level of health literacy, as we have previously seen. What is a prevention plan? Will my doctor contact me or is it my job? What happens if I don’t call my doctor? Will there be a problem?

There is an assumption underlying some of these sentences that all languages use an equivalent modal verb with the same meaning. It is assumed, perhaps, that speakers of English as a foreign or second language can easily decide which meaning of *should* is being expressed. This is a very significant, but under-appreciated problem with *should*, which I shall discuss further in 6.9.3.2 below.

In my corpus of patient information, *should* is used in a variety of grammatical structures to refer to a wide range of subjects, procedural, pre- and post-procedural, but also to general medical concerns. This is in contrast to *need to*, which, as we have seen, restricts itself to procedural concerns, and is primarily used with a 2nd person pronoun. *Have to* is not used to issue any instructions or directives in the corpus, but seems particularly related to possible occurrences and experiences of the medical procedure. *Must*, while used in a variety of contexts, is the least frequent of the four deontic modals. The utility of *should* seems to lie in the fact that it has a range of meanings, and when used to issue a directive it is generally non-threatening and can reduce the obligation to 'desirability' in some cases. These very characteristics may also make its interpretation problematic, as I will now discuss.

6.9.4.1 Interpreting should

We have seen in the preceding section that *should* is used with both epistemic and deontic uses patient information, and that its deontic meanings lie on a continuum from suggestion to advice through to an obligation. While my sampling of 150 examples (out of the 1219 that are found in the corpus) of *should* is not fine-grained, it suggests that around 70% of the uses of *should* are deontic and at the obligations end of the meaning-continuum. About 20% are epistemic with the remaining 10% used to mean that the action is desirable, but not obligatory. Interpreting which is which requires, at the very least, familiarity with the subtle pragmatics at play.

A fundamental understanding is that which relates to the medical experience itself. If a patient has never been to hospital, never had radiography or any other medical intervention, it becomes less easy to navigate the text: health literacy, as we have seen, multi-faceted and complex, and a powerful factor in the comprehensibility of patient information (Zarcadoolas et al., 2005).

Patients for whom English is a second or foreign language may also find interpreting *should* more challenging than health literate L1 speakers. Studies show considerable variation in cross-cultural use and expressions of modality, with non-native speakers being prone to pragma-linguistic transfer (Hinkel, 1995; Hussin, 2013). As Hinkel (1995, p. 329) says:

> Pragmatic and sociocultural assumptions represent fundamental points of reference that may not be frequently questioned by members of a language community in which values and common background beliefs appear to be mutually shared.

While most of the studies that have investigated the pragmatic problems encountered in intercultural medical encounters focus on spoken interaction (Yates et al., 2016; Staples, 2015; Dahm & Yates, 2013; Dahm, 2011), the findings that relate to a mismatch between meaning and intention may also be relevant to written patient-provider communication. As the use of deontic modals appears to be culturally dependent (Hinkel, 1995), it is a reasonable assumption that their interpretation may also be, and while this seems particularly true for *should*, it may also be evident with *need to*. As I have shown, *should* and *need to* are used as important expressions of obligation and necessity in patient information. This tendency may be potentially problematic for patients, however. Low health literacy and language and cultural differences may result in the non-threatening, indirect modals *should* and *need to* being interpreted as suggestions, rather than instructions.

My experience working as an English language teacher to speakers of European languages suggests that *should* is usually interpreted to be advice and it is the weak deontic meaning of desirability that is attached to *should*, not the strong deontic meaning of obligation. In Italian, for example, desirability is usually expressed with the conditional form of the verb *dovere* (must/have to): *dovresti arrivare per le 19:00* (You should arrive for 7pm) This desirability can sometimes be strong, depending on the context, but in Italian this is not the usual meaning. When students learn modal verbs in English, the tendency in many classrooms and coursebooks is to present *should* only or primarily with its desirability meaning. For students who have not had the opportunity to study in an English-speaking country (or to have a teacher who uses corpus linguistics to inform her materials), obligations are expressed with *must* or *have to*, and advice is given with *should*. In patient information materials in English, however, *should* is used for obligations, and neither *must* nor *have to* are used frequently.

Investigating reader interpretation of modals in healthcare discourse is, I feel, important research that needs to be carried out. While investigations of that size were outside the scope of my doctoral study, I decided, nonetheless, to carry out a preliminary investigation to find out how L1 patient information writers use deontic modals compared to L2 speakers of English. To that end, I developed a survey which I describe in the following section.

6.9.4 L1 and L2 uses of deontic modals in healthcare contexts

In order to investigate whether there are any differences in the way non-native speakers and native speakers use deontic modal verbs in the context of medical information, I conducted an online survey with 20 questions, the results of which can be seen in Appendix C.

Table 22 Survey of deontic modal use in radiography materials: participant info

Survey: use of deontic modal verbs in radiography materials *(Dec 2017 - July 2018)*			
Speaker-type	N (100)	Healthcare professional?	Medical comm writer?
L1	64	0	60
L2	36	25	0

106 respondents took part, with 100 finished questionnaires received. 64% indicated they were L1 speakers of English. As the respondents were contacted via a British-based healthcare communication forum, it is likely that the majority were L1 speakers of British English. Of these, 60 were healthcare communication writers. All but four of the L2 speakers worked or were training to work, in either radiography or biomedical science. I cannot know for sure which L2 respondent was working or training in healthcare as the question did not distinguish. It was also a question I added to the survey when I realised that this information was important. The first six

respondents did not answer this question. All L2 respondents had Italian, German, Hungarian or Romanian as their L1. Four of the respondents were trainers in medical English with Hungarian and Romanian as their L1s.

Having respondents who were knowledgeable about medicine in general, or who had knowledge specific to radiography was essential, as we have seen that this kind of knowledge is needed in order to distinguish a suggestion from an obligation. To my knowledge, all respondents were either very knowledgeable about radiography or, as medical communication writers, might be expected to have sufficient knowledge to appropriately interpret the sentence.

Instructions were provided asking the respondents to select the modal verb they preferred or considered most appropriate to complete the gap in 20 sentences. The sentences came from the patient information data in my corpus, so authenticity was controlled for, and four possible answers were provided. The sentences included distractors (using non-deontic modal verbs such as *can, could, will* and *might*) while the order of the answers was varied

Results

L1 and L2 difference

The results were very surprising. I did not expect to see *must* used with such frequency by either group. L1 speakers (over 90% medical communication writers) selected *must* over the other modal verbs. Remember that in my patient information corpus, *must* was the least used modal. *Should* and *need to* are used by L1 speakers with very similar frequency in my survey, though in patient information, *should* was far and away the most frequent modal, used three times as frequently as *need to,* the second most frequent modal.

In the L2 group of respondents, however, there seems to be no preference whatsoever, with all four modals being used with equal frequency. This was unexpected, and without further investigation I cannot be sure of the reasons. I have not yet examined the order in which modals were selected, though it may be that the modals are being used one after the other, as if they are synonymous, to avoid repetition. At the very least, I can say that some differences in use was revealed by my simple survey, a finding that warrants further investigation.

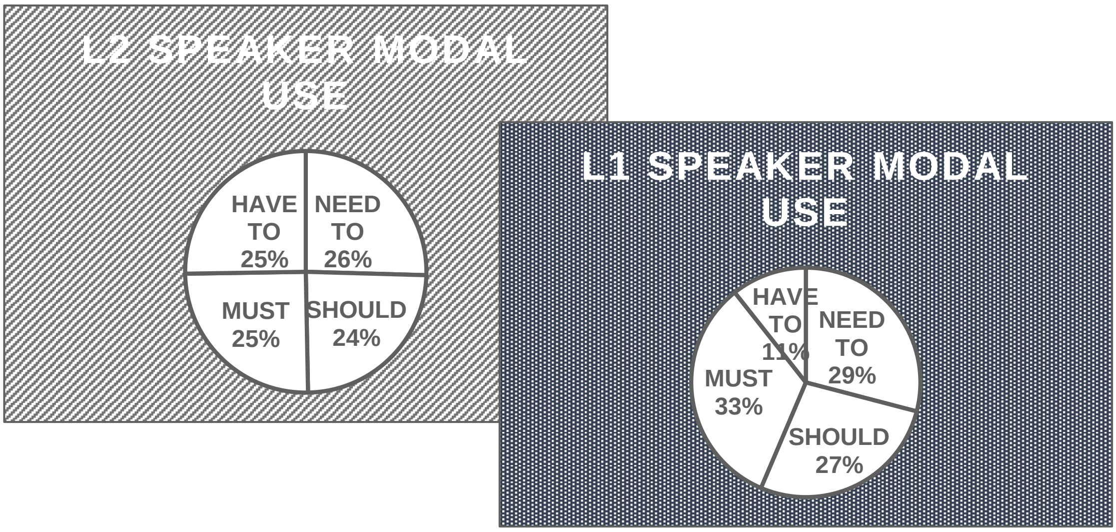


Figure 10 L1 and L2 Speaker difference in use of modals

Individual variation

When I examined the individual responses in the groups, I discovered great variation though it was particularly marked in the L1 group. Some L1 speakers never used *must*; some used it only once, preferring to use *should* or *need to* more frequently, while some respondents chose *must* most of the time.

This was an exploratory questionnaire that raised more questions than answers, and while I tried to control for medical knowledge by selecting only qualified and trainee medical professionals along with practising medical communication writers, it is possible that some healthcare communication writers and medical English trainers are not sufficiently knowledgeable about radiography to know if a statement referenced an action that was desirable or required. This highlights the difficult position patients may find themselves in when confronted by the same information.

The individual variation that I found was very interesting and could be explored further by asking respondents to provide their own response, rather than selecting it from a set of four. My survey asked people to complete a gapped sentence for ease and speed. Asking people to read statements using a range of deontic modal verbs and to rate the degree of obligation would also be a useful study. If knowledge of the topic could be controlled for, perhaps by using an entirely made-up scenario that follows recognisably ‘sensible’ rules, so that only the researchers knew what was permitted, we might see better some of the mechanisms that lie behind our use of these modal verbs. I suspect that one reason we do not see many studies of deontic modals is that capturing users’ perceptions of their range of meanings is not at all straightforward, which, my brief survey suggests is also subject to great individual variation.

The perception of negative and affirmative statements is also something that could be explored further. In my questionnaire, two sentences concerned the necessity to fast (i.e. not eat) before the exam. One sentence used a negative form and the verbs *eat* and *drink*; the second was an affirmative and used the more medical term *fast*. The huge individual variation was evident even here: *must not* was selected by just over 77% of the L1 speakers, most of whom were healthcare communication writers. The second sentence saw just under 30% selecting *must*, with the majority (47%) choosing *need to.* Perhaps *need to fast* sounds better than *must fast*? A survey that included interviews to find out what motivates people in their choice of modal verb would be interesting.

Though a small-scale questionnaire, the appearance of *must* as the most preferred modals for L1, and not *should*, is surprising, and contrasts with the frequency found in the patient information corpus. In patient information there is a very clear difference in frequency, with *should* being used eight times more frequently than *must*, and twice as frequently as *need to*. Further studies are needed to investigate the reasons healthcare information writers choose one deontic modal over another, and what meanings are perceived by readers of the materials: the patients. Studies that investigate the role of language and culture in the perception of obligations and requirements in healthcare materials are also needed, particularly as it is the multicultural, multilingual anglophone nations that produce the most patient information.

6.10 Conclusion

Patient information relies on a high frequency of modal verbs with *should* and *need to* used to express obligations and to instruct the patient. My findings, however, suggest that while patient information does express obligations, it is keen not to be seen to do so.

Obligations are phrased politely in patient information, very often using *should*, along with 3^{rd} person requirements or obligations (e.g. *your health professional should discuss…diabetics should be given a morning appointment*, etc.) *Need to* appears with surprising frequency, though in contrast to studies that report 1^{st} person and 3^{rd} person passivised structures to be the most common in writing, 75% of all occurrences of *need to* in patient information are with the 2^{nd} person. While this may suggest that it, too, is used to issue obligations and functions similarly to *should*, a closer look reveals that the majority of occurrences of *need to* refer to procedural necessities. There were no examples of strong obligations being expressed with this modal. There is a distinction in use, then, that sees *should* used with a variety of meanings and referencing different subjects while n*eed to* is more restricted. Interestingly, *need to* is never used to oblige the patient, unlike *should* (and far less frequently, *must* and *have to*).

Both *should* and *need to* can be used to minimise the voice of authority and to issue obligations and instructions that are presented as being for the good of the patient (Smith, 2003, p260). Glass (2015) claims *need to* does not possess a monolithic meaning at all, but is dependent on context and relationship between participants. In patient information, where the patient is increasingly referred to as a client or service user, and the hospital or professional a 'service provider', the voice of authority is minimised, and the tone is not overly authoritative. The strong 'we need you to' structure is never used and *need to* seems closer to external necessity in this register.

There is a pattern visible in the use of modal verbs for directives in patient information: *have to*, *need to* and *should* can all refer to procedural requirements, though the latter two can also refer to personal or non-procedural requirements. *Have* *to* is never used in this way. *Must* is the only modal used for strong obligation but is also the least used deontic modal in my data.

The use of *should* and *need to* as important expressions of obligation and necessity in patient information may be potentially problematic for patients, however. Low health literacy and language and cultural differences may result in the non-threatening, indirect modals *should* and *need to* being interpreted as suggestions, rather than instructions. Culture also influences the roles, and expectations of roles, of participants in medical interactions and thus how obligations - or suggestions - may be perceived.

The relationship between the patient and their healthcare professional is far more ambiguous today than it was a generation ago when it was very likely considered to be a straightforward non-expert/expert relationship. Patients in the 21st century are increasingly referred to as ‘clients’ or ‘customers’, and while the relationship may not be that of peers, it is increasingly a relationship where the patient expects to be treated as an equal partner in the healthcare decision-making process. Written patient information encourages this behaviour, avoiding bald obligation and encouraging patients to take responsibility for their own healthcare.

Some patients, however, are more familiar with a paternalistic style of medicine, which is still practised and where the role of doctor still has elevated status. e.g. in Asia (Clarmita et al., 2011) and former communist countries (Murgic et al., 2015). In cultures that are far less patient-centred, written information is less widely available and may well be considered irrelevant. Decisions are likely to be made by the health professional alone. It is reasonable to assume that patients from such countries will have different expectations of their role, the role of information and their expectations of the role of the doctor. As a result, the processing of obligations may be quite different depending on the perception of the role of patient and that of the doctor. For these patients, bald obligations may not be so face-threatening, while *should* may appear as little more than a suggestion. The fact remains, however, that we know very little about deontic modals in medicine, as they have yet to receive the same amount of interest as epistemic modals. More studies that look at the use and interpretation of deontic modals in medical information, both written and spoken, by patients with different linguistic backgrounds, could prove very fruitful in the quest to improve patient-provider communication.

7. Conclusions

At the start of this thesis, I set out to describe some of the lexical characteristics of patient information for radiography by applying three methodological approaches from the field of corpus linguistics. The motivation for my study lay in the fact that, in spite of the ubiquity of patient information and its growing importance in many healthcare systems, the lexical characteristics of the register had not been described. This absence in the literature also seems particularly hard to explain given the concerns relating to readability that surround discussions of patient health information and the growing concerns regarding patients' ignorance of radiography and radiation. Notwithstanding a handful of studies calling for a new linguistic approach to the writing and assessment of healthcare messages, lexical descriptions of the register are absent.

My decision to focus on keywords and lexical bundles was made primarily because the first two approaches had not yet been used with the register of patient information. Lexical bundle studies have provided plenty of evidence of the relationship that exists between the characteristics and communicative purpose of a register and the structure and distribution of its lexical bundles. The keyword approach, in turn, has shown itself to be very useful in healthcare communication studies as it can reveal hidden discourses, underlying beliefs and interesting communicative patterns that may not be otherwise apparent otherwise. As for deontic modal verbs, Vihla's (1999) study investigated modal verbs in a wide range of medical registers but she did not include patient information. My analysis was carried out on deontic modal verbs to partly fill the gap in the literature, but also because instructing patients is one of the primary functions of the information produced by hospital departments. If we understand how obligation and instruction are expressed in healthcare materials, it may lead to a better understanding of their effectiveness and why people sometimes do not comply with those requirements.

My analyses resulted in a wealth of information regarding the lexical characteristics of the register with clear implications for the both the development and evaluation of healthcare materials. My findings also support the notion that standard readability measures are inappropriate tools to use with healthcare materials. Some of the most important findings, and thus the reasons these measures are inappropriate, I will highlight in this concluding chapter. These findings will be followed by other aspects of interest revealed by my analyses, aspects which certainly merit further investigation, though which are outside the scope of the current study. I will also present some of the limitations to my research.

One of the most significant findings of my research is that, in terms of its lexico-grammatical characteristics, patient information has more in common with formal prose than conversation. This may seem a peculiar observation, given that the patient information that I analysed is written and thus as written discourse it may be expected to possess lexical qualities appropriate to written prose. However, we should remember the guides produced for patient information writers that exhort them to produce information that resembles a conversation (Plain English Campaign, 2001). At first glance much patient information does seem to be conversational in style, and many examples of patient information are laid out on the page as if it were a conversation, with questions as headings. However, closer analysis reveals that the structure of the text itself is not at all conversational, with two-thirds of the 4-word lexical bundles used in patient information those that are more commonly found in formal prose.

The same style guides tell writers to avoid the passive and to use *we* and *you* (NHS Toolkit, 2003), though in spite of the recommendations, the passive does appear with some regularity in the patient information that I studied. I found the passive used in 18% of the 4-word bundles in patient information, while a quarter of the occurrences of the modal *need to* are passive. One reason for the appearance of the passive, may be that avoiding it entirely is simply impractical advice for patient information writers. Some information is more naturally presented in the passive voice and advising writers to use the active voice continually may have an impact on how the information is read and comprehended. The passive seems more appropriate given that much of the information in patient information for radiography concerns what will be done to the patient, why it will be done, how it will be done and how long it will take. Using the active voice to present this kind of information is not natural. The use of the passive may explain some of the frequency of the subject pronoun *you* which is highly frequent in patient information, appearing at a rate of more than 14,000 per million words. *We* barely features, however, and is used at a rate of less than 700 times per million words. The focus in patient information seems to be predominantly on the patient and we may rightly ask whether this is always or entirely appropriate. Patients need to know who is instructing them, though the voice of authority (*we*) is almost entirely absent in the materials I examined. I will return to this finding later in this concluding chapter.

The characteristics of formal prose are also seen in the use of modals in my data. *May* is rare in spoken language and much more common in academic prose (Biber et al., 2002), though my analyses reveal it to be the most frequent modal in the corpus of patient information. *Have to*, on the other hand, the most frequent modal in conversation (Biber et al., 2002), is barely used in the patient information that I analysed. An incidental finding that relates to style but also to how conversational patient information really is, was that there were no contractions used anywhere in the corpus. *Can't* appeared as *cannot*; *won't* appeared as *will not*, *mustn't* appeared as *must not.* Contractions, of course, are a feature of spoken discourse. Consumer advice used contractions frequently, and also came across as the friendlier and less formal of the two registers. The generally accepted explanation for not using contractions in writing is that they appear informal, less scientific and/or authoritative, which, ironically, is precisely the type of writing patient information hopes to be. There does seem to be some evidence that negative contractions are problematic for readers with learning disabilities, but I am unaware of evidence that shows that the average reader has issues with contractions.

The findings relating to the lexico-grammatical characteristics of patient information can help us answer the research question that relates to the comprehensibility of patient information for radiography and, by extension, the appropriacy of readability assessments. As we have seen throughout this thesis, readability measures generally do not consider the structure of the sentence but consider the length of the sentence and the length of individual words. If two-thirds of the 4-word lexical bundles are of a structural type more common to academic prose, it suggests that in spite of shorter sentences and shorter words, the text may be less accessible for some patients. Many of these 4-word bundles were referential bundles, which are also more common in academic prose than conversation. They include passive structures such as *be used to take* and prepositional phrase structures, which predominate, such as *at the end of*. These types of bundle are far more frequent in more formal, information-dense text.

At the very least, patient information texts may have an academic, official voice, in spite of the simplified text, that may affect how these healthcare materials are perceived by readers. We should keep in mind that wanting to read information can be as important as being able to read it: to be effective, a patient information leaflet and healthcare materials in general must be ‘noticed, read, understood, believed and remembered’ (Protheroe, Estacio, Saidy-Kahn, 2015, p. 192, citing Ley, 1982).

Perhaps the inclusion of these structures is not surprising, however, when we consider the focus in procedural information on time and place, resulting in many references to time and place using phrasal structures. E.g. *after a couple of hours; within a few days; at the top of.* These types of structures are a principle characteristic of informational discourse, and one of the primary functions of patient information is to inform. The high-density of information in these healthcare materials suggests that certain structures - particularly the time and place structures - are inevitable. If that is the case, more research is needed on how these structures can remain without rendering the text too difficult to process or too formal in tone. One approach could be to time patients reading text with a low and high density of referential bundles to see if processing time is affected; message comprehension and retention can then be tested.

Another significant finding relating to the readability of patient information is the lack of discourse organising bundles that it appears to use. Discourse organising bundles function to organise the text by providing signposting and to support the reader (or listener). They play a significant role in the cohesion of a text, and cohesion, as we have seen, has a fundamental role in making text comprehensible. Discourse bundles also appears to have some relationship to how persuasive a text appears, as we saw in chapter 2, and how well the message of the text is remembered which is of great relevance to the topic of patient information. In spite of their function, which one would assume would be welcome in information-dense text, this category of bundle was the least used overall in the patient information, representing just 12% of the 4-word bundles. While rare overall, however, three discourse bundles were seen to be repeatedly used, and were among the first 5 most-frequent bundles in the corpus. Closer inspection revealed that these bundles were, in fact, framing instructions, a use which seems to highlight the fact that instruction is the principal function of patient information.

The relationship of discourse organising bundles specifically and lexical bundles more generally to the cohesion of a text warrants further investigation and may certainly have a bearing on its readability. Standard readability assessments, however, do not measure cohesion nor do they consider the type of lexical structure used (or the ratio of noun phrases to verb clauses, which in patient information are 2:1). Perhaps the lack of discourse organising bundles and the high number of referential/academic bundles confound the average patient. Perhaps they have no effect on comprehensibility, but until studies are carried out, we cannot say for sure. A similar experiment to that undertaken by Martinez (2002), on the effect of rhetorical structure on the comprehension and recall of unfamiliar text by ESP students, could be carried out on a group of patients, controlling for L1 and L2 speakers of English and type of patient information. An important finding from Martinez (2002) is that the use of rhetorical structure had a positive effect on the reader’s ability to comprehend and reproduce the information only when that structure was also recognised by the readers. When they failed to recognise the rhetorical structure in the text, the students in her study were still able to reproduce the information without having fully comprehended it What patients are able to recognise as rhetorical structure in healthcare materials we know nothing about, and while the headings-as-questions device many healthcare materials make use of is likely to be very helpful for patients navigating a leaflet, there seems little attention paid to cohesion within the text itself.

It should be remembered, too, that my corpus was made up of patient information for radiography. Different kinds of patient information exist, of course, some of it procedural such as the kind investigated here; some of it related to accessing services in the health system (a particularly ubiquitous kind of information in the NHS); some of it about specific health conditions; some of it is about medical devices and equipment that patients need to use. There have been no studies that I am aware of that investigate how the lexico-grammatical structure of these kinds of healthcare materials differ from each other, though there is evidence that the complexity of the texts differ, when standard readability measures are applied: Protheroe et al. (2005) found that readability levels of various kinds of materials (all found in medical centres and clinics in one region of the UK) differed, with some texts being judged more complex to read than others.

Another reason that standard readability measures are inappropriate (though certainly not the only reason) we first met in chapter 1: readability measures are not a reliable measure of lexical complexity. While my investigations did not concern the characteristics of the content words used in patient information, it is worth revisiting the topic briefly here. The length of a word is not always an indicator of its complexity and this seems particularly the case when we are considering shorter words. If we consider vocabulary used to talk about and to do medicine, we find medical vocabulary on one end of the spectrum, along with nomenclature and highly technical medical terms with Latin and Greek roots. General terms will be found at the other end of the spectrum while *influenza* might be found towards the middle. *Influenza,* a medical term, is often abbreviated to *flu* by lay people and understandable by the majority, if not all, patients. It is also probably close to the middle of the spectrum where we find the words that both patients and providers use, though used with different meanings or connotations. *Chronic* is a much-used example, meaning 'long-standing and persistent' by the medical professional and 'severe' in some varieties of English. All too often the misunderstanding goes undiscovered as both meanings relate to something unpleasant in need of a solution. Readability measures are likely to class *influenza* as a difficult word; *chronic* as an easier word, but neither judgement gives us the whole picture of how patients and providers are using and understanding these words.

As I show in this thesis, however, the complexity of medical vocabulary is just part of the story. When we consider the readability and the effectiveness of patient healthcare materials, the kinds of structure used in the text is also relevant. Readability measures do not distinguish between structural types, such as a verb clause in a longer sentence and a short noun phrase in a short, passive structure. Verb clauses are more common in spoken language, noun phrases more common in more formal written language, though standard readability measures will quite possibly judge the former to be more complex on the basis of length, when in fact it is the latter that is likely to be more complex to process. However conversational and accessible a leaflet may appear, with its cheery colours, large font and chatty headings, if the text itself has a structure that is more redolent of academic prose, it surely has a bearing on how the text will be received by the reader and understood.

Assessing the impact of certain structures in the context of healthcare materials is, I believe, a necessary next step and one which necessitates the involvement of the patient from the very beginning of the process. As we saw in chapter 2, there have been repeated calls for patients to be more involved in the production of healthcare materials, and not simply brought in when assessing the readability of the finished product. How patients respond to different types of structure, complex and simple, can be tested, as can the effect of certain types of structure on the cohesion of a text. Cohesion needs to take centre stage in the production and evaluation of healthcare materials and care taken that over-simplification does not result in a text that is stripped of the aspects that make it easier for a reader to process.

This brings us on to the role of health literacy in rendering standard readability measures inappropriate. Health literacy is, as we have seen (Zarcadoolas, 2011) a complex, multi-faced skill that goes far beyond the ability to read. The finding from the modal verb analysis that patient information relies heavily on *should* (and to a lesser extent *need to*) to instruct and oblige patients is an important one that becomes even more significant when viewed in the context of health literacy. Both of these modal verbs have been described as appealing to the needs of the patient, thereby avoiding any overt instructions, so perhaps they are seen as fulfilling the need to be patient-centred, to be friendly and non-alarming. *Should* is particularly problematic, however, as we have seen, as its meanings lie on a continuum with obligation at one end and mere suggestion at the other. There seems to be a vast amount of information to be dealt with a priori if the reader is to be fully confident of their interpretation of *should* when reading the patient information in my corpus.

As an example, taken from the context of radiography, a patient may well wonder what relevance summer fruits have, or a plate of clams, to a radiographic exam. When a patient who is booked in for an MRI scan reads that they must inform staff if they have ever experienced reactions to strawberries or shellfish, does it matter if *you should tell us* or *you must tell us* is used*?* The response to this question may be that it depends very much on whether the patient understand the reasons for asking. Knowing why the question is important will help the patient to decide whether *should* is to be taken as advice, as a suggestion or an instruction. As it happens, people who experience allergic reactions to shellfish and strawberries will likely experience the same reaction to the gadolinium contrast fluid that is used during an MRI. It will not kill the patient, but the reaction is unpleasant and if it can be avoided, so much better for the patient. When medical professionals use *should* in this context, they want their statement to be interpreted as an instruction, not as a suggestion.

We have seen that not only do patients frequently not know which exams use radiation, they also have a very poor understanding of the concomitant risks of radiation, suggesting that much background knowledge necessary to interpret certain types of instruction is lacking. Poor health literacy may mean that if the reason for the instruction is not understood by all patients it may not be responded to truthfully or accurately, if it is responded to at all. Our understanding of non-compliance (or non-adherence, depending on your choice of term) in healthcare will also benefit from better understanding the meanings that people attach to different modal verbs. ‘I didn’t know I had to’ and not ‘I didn’t want to’ may turn out to be the real reason why a patient did not follow what a healthcare professional considered were clear instructions.

It has also been suggested that patients who are more familiar with a paternalistic style of medicine may expect clearer instructions, and perhaps not only these people: older patients living in healthcare systems that pursue patient-centred medicine may also expect to receive instructions from their healthcare provider in the clearest, unambiguous way possible. Subtle, indirect structures for some people may be interpreted to be mere suggestions. There has been no research, to my knowledge, on the use of these words in healthcare materials, but my findings make it clear that further investigations are needed.

Language background may well be a factor in the interpretation of modal verbs, as my simple survey suggested. This is also an area that warrants further research, particularly as so much medicine is practised in a multicultural environment, not only in the UK but in many other countries. This aspect is also relevant in non-Anglophone societies of course, as English is the global lingua franca of medicine. The different meanings L2 speakers of English attach to *have to, must, should* and *need to* in the context of medicine also need to be examined. My survey results suggested there was little distinction, and all were used more or less with equal frequency. Does that also suggest that many L2 speakers of English understand these words in a similar way? Or, as many of the respondents were medical students, did they react as if they were being tested and so hedged their bets by using all the modal verbs? We need to find out. If the former, the implications are that some non-native speakers of English may not appreciate that modal verbs of obligation are not always – or even ever – fully interchangeable. There seem to be very few studies that have looked at the interpretation of deontic modal verbs and the perception of obligation in different registers, none at all in the context of healthcare materials. The interpretation of obligation becomes even more complex when we consider the absence of named authority in patient information that my research also revealed. The 'blurring of shouldness' suggests that contextual cues - the knowledge the reader brings to the text and to surrounding lexico-grammatical structure - take on a vital role in the interpretation of the degree of obligation. As we have seen, however, health literacy problems and certain structural characteristics of patient information may mean this support is absent.

And the story does not end there: my survey also revealed great individual variety in the use of deontic modal verbs by L1 speakers which was an entirely unexpected finding and one that most certainly warrants further investigation. This may suggest that we can never be too sure what a deontic modal is intended to mean without sufficient supporting context, which may cause few comprehension problems in familiar contexts but when the context is unfamiliar, being confident of the interpretation of a deontic modal will be more of a challenge. Health literacy issues can result in some people misinterpreting the significance of a deontic modal verb precisely because they lack the contextual cues needed to interpret the utterance. Contextual cues may also be absent because of the amount of text simplification that is common in standard healthcare materials.

My investigations also revealed other interesting findings to complement those described above. These findings do not relate directly to readability and my data did not allow any more than a cursory investigation, but I believe they are important enough to warrant further investigation.

One such finding related to the naming and referencing of healthcare professionals in the materials. There were many technical and medical terms in my keyword lists, underlining the complex nature of radiography, though the keywords that related to professionals in the lists suggested that certain internal inequalities and bias in medicine are mirrored in patient communication via the naming habits of medical professionals. The minimal use of the name of the profession that is most involved with the radiographic examinations, the radiographer and technologist, alongside the minimal use of nurse is hard to explain, particularly as the role of the doctor is so marked. Radiologist, doctor and physician are used considerably more frequently than radiographer or nurse. Why this may be so relates to the second significant finding I believe, which concerns the impact of gender on information seeking and information preference.

If it is the case, as reported by Seale et al (2006) and Seale and Charteris-Black (2008), that a patient's biological sex and/or gender may affect what kind of medical information they want to read, and if it is the case that older men in particular like naming experts and specialists and want to read about treatment and therapy and the details relating to procedures, then it seems that patient information for radiography, as it currently stands, is more appropriate to the needs of older men. Though tailored healthcare information is increasingly seen as an effective response to the very personal nature of the experience of (ill) health, it is only in the last few years that some healthcare materials have been tailored for sex and gender, and only in the last few years that public health policies have included sex and gender as a determinant of health. To my knowledge, the appropriacy of the information presented in procedural information has been assumed rather than investigated. We have assumed that radiography information needs to be the same for everybody, without stopping to think that different people may want quite different types of information. The few studies where patients were asked their views on leaflets for depression and rheumatoid arthritis, referred to in chapter 2, found a number of shortcomings in the content of the leaflets. Patients reported not finding enough information on certain topics, which were dealt with in a fairly superficial manner, while some aspects of living with depression or a chronic condition were never referred to at all. Though there are only a handful of studies like this, there is a growing awareness among professional patient information writers and communication agencies of the need to involve patients and advocacy groups in the developmental process.

Being more aware of the need to involve patients in the development of healthcare materials does not mean, however, that writers fully understand the impact of diversity on the content of that information. Where sex and gender are concerned, I discovered that some writers of patient communication are reluctant to consider sex and gender as factors that may affect what people want to know. The healthcare materials writers I contacted seemed entirely unaware of the growing literature that exists on the influence of sex on many serious health conditions, affecting signs and symptoms, diagnosis, treatment and outcomes. This lack of knowledge clearly restricts the approach to the development of patient information. This could prove to be a very exciting area for further research. Writing information should be a process that begins with the patient, rather than including them at the end of the process for their feedback, which often happens. Studies that look at the impact on the quality and content of patient information materials when patients have been involved at different stages would also be interesting. Controlling for age, sex and gender, as well as cultural background, could prove very fruitful.

Methodologically, there are some limitations to my study. With hindsight, while I wanted to capture some of the accessibility the internet brings in the development of my corpus, the lexical variation between the US and UK material that became evident in the results of the key word analysis, suggests that separating these corpora would have been methodologically more sensible. Not doing so meant that time was spent double-checking the source of the data while two of the categories of keywords - legal and treatment - related primarily to US material. There is also the question of the length of the texts that were selected on the basis of their content. Text length varied considerably, and the US-sourced material contained many more long examples than the UK material. The different sizes of the corpora may have resulted in different frequency counts for certain items, though I was attentive to any unusual clustering of the data and make mention in the text where relevant.

Using material from the UK and the US caused some problems, but at the same time, only using material from these countries is a limitation. I had intended to look at patient information from a number of different English-speaking countries. I had assumed that patient information leaflets for radiography would be readily available on websites, as they were in the UK and the US. This was not the case, and I found it difficult to find the equivalent in Australian, New Zealand, Canada and even Ireland. Five years have passed since I first started to gather material, and it is very likely that patient information is more readily available from these countries. The question of whether I would have been able to assess the quality of the information, if it had not come from a medical authority such as the NHS, is also pertinent. I needed to know that the information I analysed came from a reputable source and was an example of the kind of information patients would be expected to read. I can guarantee that applies to the material that makes up my corpus.

Another limitation relates to the adjusted frequency. My decision to use adjusted frequencies of per million was taken at the beginning of my research and was not remarked upon until a few weeks before submission. It was suggested to me, at this point, that an adjusted frequency of 1000 may have been more appropriate, given the size of my texts and my corpus. I sought further advice and it appears that while there is some debate among corpus linguists on this topic, many are of the opinion that small corpora necessitate small adjusted frequencies. However, I feel that an adjusted frequency of per million has little bearing on my findings, given the types of analyses that were performed, though on the few occasions where frequency results are reported, my data will be slightly inflated as a result.

To conclude, I have demonstrated in this thesis that approaching healthcare information from a linguistic perspective can yield many important insights that both support findings from the literature and introduce new avenues worthy of exploration. My research shows that there are characteristics that relate to the lexico-grammatical make-up of healthcare materials that are potentially every bit as important in the making-of-meaning for a reader as vocabulary is. The structure of a text can be subtle, and while simplifying is the aim of medical communication writers, too many untested assumptions have been made about what makes a healthcare text difficult to read: indeed, what many writers understand by ‘text simplification’ looks overly-simplified in the light of my findings, many of which relate to the importance and role of cohesion in reading. Along with tools specifically designed to measure the semantic complexity of healthcare materials, we need to be able to better evaluate cohesion, and also to better understand the impact of linguistic features that can contribute to cohesion, such as lexical bundles. When developing materials, writers need to keep health literacy uppermost in their minds, at least as far it may affect reader understanding of the specific health information leaflet in their hand. We also need to better understand the impact of ‘the blurring of shouldness’ in healthcare materials, as we may well find that an absence of authority is not only inappropriate but also ineffective in the context of healthcare information.

As there is every indication that radiographic examinations such as CT will become ever more frequent, it is vital that we understand how to develop information that patients will read, understand and act upon. As the availability and importance of healthcare materials in general, both digital and print, continues to increase, the quality of the message and the reader’s experience are of paramount importance.

Appendices

Appendix A

LEXICAL BUNDLES EXTRACTED AND CLEANED

BUNDLE	STRUCTURAL TYPE
1. you may be asked	STANCE INTENTION /PREDICTION
2. if you have any	DISCOURSE ORGANISERS /TOPIC INTRODUCTION?
3. you will be asked	STANCE INTENTION / PREDICTION
4. if there is any	DISCOURSE ORGANISING TOPIC INTRODUCTION
5. how do I get	DISCOURSE ORGANISER/TOPIC ELABORATION
6. It is important that	STANCE OBLIGATION/DIRECTIVE
7. a small amount of	REFERENTIAL TIME /QUANTITY
8. may be used to	STANCE ABILITY
9. a starting point for	REFERENTIAL / INT FRAMING
10. will be able to	STANCE ABILITY
11. of the inside of	REFERENTIAL / TIME/PLACE/TEXT REFERENCE
12. if you have a	DISCOURSE ORGANISER / TOPIC FOCUS INTRODUCTION
13. you may be given	STANCE INTENTION /PREDICTION
14. the best way to	REFERENTIAL /SPECIFICATION OF ATTRIBUTES / INTANGIBLE FRAMING
15. you will need to	STANCE/OBLIGATION/DIRECTIVE
16. you will feel a	STANCE INTENTION /PREDICTION

17. may also be necessary	STANCE INTENTION / PREDICTION
18. may need to be	STANCE/INTENTION /PREDICTION
19. You may also be	STANCE INTENTION /PREDICTION
20. at the same time	REFERENTIAL / TIME/PLACE/TEXT REFERENCE
21. Are you required to	STANCE OBLIGATION/DIRECTIVE
22. at the end of	REFERENTIAL TIME/PLACE
23. Who will you see	STANCE INTENTION /PREDICTION
24. If you have not	DISCOURSE ORGANISER? TOPIC ELABORATION
25. which is going to	STANCE INTENTION /PREDICTION
26. a slight chance of	STANCE EPISTEMIC
27. to the area of	REFERENTIAL / TIME/PLACE/TEXT REFERENCE
28. over a period of	REFERENTIAL /TIME /PLACE/ TEXT REFERENCE
29. with the use of	STANCE ABILITY
30. so as not to	DISCOURSE ORGANISERS /TOPIC ELABORATION
31. If you would like	STANCE DESIRE
32. may be necessary to	STANCE INTENTION /PREDICTION
33. for a few hours	REFERENTIAL /TIME /PLACE/ TEXT REFERENCE
34. it is safe to	STANCE /EPISTEMIC?
35. the area of your	REFERENTIAL /TIME /PLACE/ TEXT REFERENCE
36. to the site of	REFERENTIAL / TIME/PLACE/TEXT REFERENCE
37. you may need to	STANCE INTENTION /PREDICTION
38. you will not feel	STANCE INTENTION /PREDICTION

39. there is a risk	REFERENTIAL IDENTIFICATION/FOCUS
40. over the area of	REFERENTIAL /TIME /PLACE/ TEXT REFERENCE
41. it is possible to	STANCE EPISTEMIC
42. can be used to	STANCE/ ABILITY
43. any of the following	REFERENTIAL / IDENTIFICATION/FOCUS
44. that you are taking	REFERENTIAL /IDENTIFICATION/FOCUS
45. will make you feel	STANCE INTENTION /PREDICTION
46. it may be necessary	STANCE OBLIGATION/DIRECTIVE
47. at the site of	REFERENTIAL /TIME /PLACE/ TEXT REFERENCE
48. in the centre/center of	REFERENTIAL /TIME /PLACE/ TEXT REFERENCE
49. are a form of	REFERENTIAL /IDENTIFICATION/FOCUS
50. there may be some	STANCE INTENTION /PREDICTION
51. that is used to	STANCE ABILITY
52. for a few seconds	REFERENTIAL /TIME /PLACE/ TEXT REFERENCE
53. to be less than	REFERENTIAL SPECIFICATION OF ATTRIBUTES / QUANTITY
54. at the time of	REFERENTIAL /TIME /PLACE/ TEXT REFERENCE
55. may advise you to	STANCE INTENTION /PREDICTION
56. With the ability to	STANCE ABILITY
57. on the day of	REFERENTIAL /TIME /PLACE/ TEXT REFERENCE
58. may or may not	STANCE INTENTION /PREDICTION
59. for four hours before	REFERENTIAL /TIME /PLACE/ TEXT REFERENCE

60. from the area of	REFERENTIAL /TIME /PLACE/ TEXT REFERENCE
61. period of time before	REFERENTIAL /TIME /PLACE/ TEXT REFERENCE
62. sometimes referred to as	REFERENTIAL/ IDENTIFICATION/FOCUS
63. that may be used	STANCE ABILITY
64. that is located in	REFERENTIAL /TANGIBLE FRAMING
65. as soon as you	REFERENTIAL /TIME /PLACE/ TEXT REFERENCE
66. you will also have	STANCE/INTENTION PREDICTION
67. Why do you need	DISCOURSE ORGANISER TOPIC INTRODUCTION /FOCUS
68. can also be used	STANCE ABILITY
69. at the part of	REFERENTIAL/TIME/PLACE/TEXT
70. a specified period of	REFERENTIAL /TIME /PLACE/ TEXT REFERENCE
71. may be used during	STANCE ABILITY
72. have previously had a	DISCOURSE ORGANISER -TOPIC INTRODUCTION FOCUS
73. is also used to	STANCE ABILITY
74. for a few days	REFERENTIAL /TIME /PLACE/ TEXT REFERENCE
75. to make sure that	DISCOURSE ORGANISER - TOPIC ELABORATION/CLARIFICATION
76. how deeply you are	REFERENTIAL INTRANGIBLE FRAMING
77. will be moved into	STANCE INTENTION/PREDICTION
78. may take up to	STANCE INTENTION /PREDICTION
79. may be connected to	STANCE INTENTION /PREDICTION
80. will be placed into	STANCE INTENTION /PREDICTION
81. as a result of	REFERENTIAL SPECIFICATION OF ATTRIBUTES / INTANGIBLE FRAMING

82. in charge of your	REFERENTIAL /SPECIFICATION OF ATTRIBUTES / INTANGIBLE FRAMING
83. may be able to	STANCE ABILITY
84. we may have to	STANCE OBLIGATION/DIRECTIVE
85. or for people with	REFERENTIAL /IDENTITY/FOCUS
86. you may feel a	STANCE INTENTION /PREDICTION
87. in the form of	REFERENTIAL /SPECIFICATION OF ATTRIBUTES /TANGIBLE FRAMING
88. may also be used	STANCE ABILITY
89. If you do not	DISCOURSE ORGANISER TOPIC ELABORATION /CLARIFICATION
90. one or more of	REFERENTIAL /SPECIFICATION OF ATTRIBUTES / QUANTITY
91. does not have to	STANCE OBLIGATION/DIRECTIVE
92. If you need any	STANCE DESIRE?
93. in the treatment area	REFERENTIAL / TIME/PLACE/TEXT REFERENCE
94. can be treated with	STANCE ABILITY
95. at high risk for	STANCE EPISTEMIC
96. When will you get	DISCOURSE TOPIC INTRODUCTION/FOCUS,
97. please let us know	STANCE/OBLIGATION
98. will be shown where	STANCE INTENTION /PREDICTION
99. will also check the	STANCE INTENTION /PREDICTION
100. plenty of time to	REFERENTIAL /SPECIFICATION / QUANTITY
101. we will ask you	STANCE INTENTION/PREDICTION
102. the inner lining of	REFERENTIAL /TIME /PLACE/ TEXT REFERENCE
103. that need to be	STANCE OBLIGATION/DIRECTIVE

104.	Will try to keep	STANCE INTENTION /PREDICTION
105.	if you need to	STANCE OBLIGATION/DIRECTIVE
106.	so that it can	DISCOURSE ORGANISER TOPIC ELABORATION
107.	will be taken into	STANCE INTENTION /PREDICTION
108.	is a type of	DISCOURSE ORGANISER - TOPIC ELABORATION/CLARIFICATION
109.	In the case of	REFERENTIAL/SPECIFICATION/ INTANGIBLE

Appendix B

Survey of modal verb preference

The following sentences come from patient information leaflets for radiography.

Each sentence has a word missing. Select the word you prefer to fill the gap.

There is no right or wrong answer. Simply choose the word you prefer or that you think is the most appropriate choice.

Is English your first or primary language? * Yes or no

Are you a language teacher? Yes or no

Do you work in healthcare or medicine? (including healthcare communications but NOT as a language teacher) Yes or no

----

Now complete the following sentence. Remember, there is no right or wrong answer!

Your bowel ______________be empty before the examination

must	should	will	has to

When you arrive you ______________go to the reception desk in the department

need to	must	should	have to

If you stay in the department then you ______________use the special toilet for nuclear medicine patients.

need to	have to	must	should

Some diseases such as colds ___________trigger an asthma attack.

may	can	will	---

You ___________ smoke after midnight the day before the procedure.

should not	must not	can not	do not

You _____________tell the radiographer if you have breast implants.

must	should	have to	need to

There is a very small risk that inflating the colon with air __________ injure or perforate the bowel

could	might	will	can

During the exam, you ___________ lie still, but breathe normally as you move through the scanner.

have to	must	should	need to

Women __________always inform their physician and x-ray technologist if there is any possibility that they are pregnant

must should have to need to

If you have any allergies you __________tell the radiology staff before you have the examination.

need to have to must should

You ____________ eat or drink anything 6 to 12 hours before the procedure

must not can not should not ought not to

You _________ fast before the procedure.

need to should must have to

Some patients ___________take antibiotics before the procedure

have to need to must should

If you go home after the exam, you __________ arrange for someone to take you home by car or taxi and to stay with you overnight.

must need to should have to

Many things __________lead to the inflammation of the lungs and abnormal muscle tightening, these are known as triggers.

might can could will

In diabetic neuropathy, your feet or legs ________feel numb or unusually cold.

could might should will

Risks and complications are very unlikely, but possible. You __________know about them just in case they happen.

should must need to have to

Your visit to the clinic _____________mean spending two to three hours in the department

will could may can

If you___________cancel your appointment or change the date or time, please call us on the number below.

have to must ---- need to

You ____________stay in hospital for up to four hours after the procedure for us to observe you.

have to should need to must

References

Ädel, A., & Erman, B. (2012). Recurrent word combinations in academic writing by native and non-native speakers of English: A lexical bundles approach. http://doi.org/10.1016/j.esp.2011.08.004

Adepu, R., & Swamy, M. K. (2012). Development and Evaluation of Patient Information Leaflets (PIL) Usefulness. *Indian Journal of Pharmaceutical Sciences*, *74*(2), 174–8. http://doi.org/10.4103/0250-474X.103857

Adolphs, S., Brown, B., Carter, R., Crawford, P., & Sahota, O. (2004). Applying corpus linguistics in a health care context. *Journal of Applied Linguistics*, *1*(1), 9–28. http://doi.org/10.1558/japl.1.1.9.55871

Al-Harthy, N., Sudersanadas, K. M., Al-Mutairi, M., Vasudevan, S., Bin Saleh, G., Al-Mutairi, M., & Hussain, L. W. (2016). Efficacy of patient discharge instructions: A pointer toward caregiver friendly communication methods from pediatric emergency personnel. *Journal of Family & Community Medicine*, *23*(3), 155–60. http://doi.org/10.4103/2230-8229.189128

Allan, R., & Se, R. A. (2017). From Do You Know to I Don’t Know: An Analysis of the Frequency and Usefulness of Lexical Bundles in Five English Language Self-Study Books. *Corpus Pragmatics*, *1*, 351–372. http://doi.org/10.1007/s41701-017-0016-9

Angouri, J., & Sanderson, T. (2016). “You’ll find lots of help here” unpacking the function of an online Rheumatoid Arthritis (RA) forum. *Language and Communication*, *46*, 1–13. http://doi.org/10.1016/j.langcom.2015.10.001

Antaki, C., & Kent, A. (2012). Telling people what to do (and, sometimes, why): Contingency, entitlement and explanation in staff requests to adults with intellectual impairments. *Journal of Pragmatics*, *44*(6–7), 876–889. http://doi.org/10.1016/j.pragma.2012.03.014

Askehave, I. (1999). Askehave-Communicative Purpose as Genre Determinant. *Journal of Linguistics*, *23*(23), 13–24. Retrieved from https://pure.au.dk/ws/files/9940/H23_02.pdf

Askehave, I., & Zethsen, K. K. (2014). A comparative analysis of the lay-friendliness of Danish EU patient information leaflets from 2000 to 2012. *Communication and Medicine*. http://doi.org/10.1558/cam.v11i3.20700

Atkinson, D. (1992). *The Evolution of Medical Research Writing from 1735 to 1985: The Case of the Edinburgh Medical Journal*. Retrieved from http://applij.oxfordjournals.org/

Azam, N., & Harrison, M. (2011). Patients' perspectives on injuries. *Emergency Medicine Journal : EMJ*, *28*(7), 601–3. http://doi.org/10.1136/emj.2009.082032

Badarudeen, S., & Sabharwal, S. (2010). Assessing readability of patient education materials: Current role in orthopaedics. *Clinical Orthopaedics and Related Research*, *468*(10), 2572–2580. http://doi.org/10.1007/s11999-010-1380-y

Bagley, C., Hunter, R., & Bacarese-Hamilton, I. (2011, July). Patients' misunderstanding of common orthopaedic terminology: the need for clarity. *Annals of the Royal College of Surgeons of England*. http://doi.org/10.1308/003588411X580179

Baker, D., & Robson, J. (2012). Communication training for international graduates. *The Clinical Teacher*, *9*(5), 325–9. http://doi.org/10.1111/j.1743-498X.2012.00555.x

Barbieri, F. (2018). I don't want to and don't get me wrong. In J. Kopaczyk & Tyrkko Jukka (Eds.), *Applications of pattern driven methods in corpus linguistics* (pp. 251–276). Amsterdam: John Benjamins Publishing Company. http://doi.org/10.1075/scl.82.10bar

Barnes, R. K., Jepson, M., Thomas, C., Jackson, S., Metcalfe, C., Kessler, D., & Cramer, H. (2018). Using conversation analytic methods to assess fidelity to a talk-based healthcare intervention for frequently attending patients. *Social Science & Medicine*, *206*, 38–50. http://doi.org/10.1016/j.socscimed.2018.04.008

Bartels, R. D., Kelly, K. M., & Rothman, A. J. (2010). Moving beyond the function of the health behaviour: The effect of message frame on behavioural decision-making. *Psychology and Health*, *25*(7), 821–838. http://doi.org/10.1080/08870440902893708

Baumann, B. M., Chen, E. H., Mills, A. M., Glaspey, L., Thompson, N. M., Jones, M. K., & Farner, M. C. (2010). Patient Perceptions of Computed Tomographic Imaging and Their Understanding of Radiation Risk and Exposure. *YMEM*, *58*, 1–7.e2. http://doi.org/10.1016/j.annemergmed.2010.10.018

Bellés-Fortuño, B. (2018). Multimodality in medicine: How university medical students approach informative leaflets. *System*, *77*, 28–38. http://doi.org/10.1016/J.SYSTEM.2018.02.012

Berkhof, M., van Rijssen, H. J., Schellart, A. J. M., Anema, J. R., & van der Beek, A. J. (2011). Effective training strategies for teaching communication skills to physicians: An overview of systematic reviews. *Patient Education and Counseling*, *84*(2), 152–162. http://doi.org/10.1016/j.pec.2010.06.010

Berry, D., Knapp, P., & Raynor, D. (2002). Provision of information about drug side-effects to patients. *The Lancet*, *359*(9309), 853–854. http://doi.org/10.1016/S0140-6736(02)07923-0

Berry, D. C., Knapp, P., & Raynor, T. (2006). Expressing medicine side effects: Assessing the effectiveness of absolute risk, relative risk, and number needed to harm, and the provision of baseline risk information. *Patient Education and Counseling*, *63*(1–2), 89–96. http://doi.org/10.1016/J.PEC.2005.09.003

Bianco, A. V. (2016). VAGHEZZA, INDETERMINATEZZA E INCERTEZZA NEI FOGLIETTI ILLUSTRATIVI DEI MEDICINALI IN INGLESE E IN ITALIANO. http://doi.org/10.1285/i22390359v19p41

Biber, D. D., Conrad, S., Reppen, R., Byrd, P., & Helt, M. (2002). Speaking and Writing in the University : A Multidimensional Comparison. *TESOL Quarterly*, *36*(1), 8–48. http://doi.org/10.2307/3588359

Biber, D. (2012). Register as a Predictor of Linguistic Variation. *Corpus Linguistics and Linguistic Theory*, *8*(1), 9–37. http://doi.org/10.1515/cllt-2012-0002

Biber, D. (1988). *Variation across speech and writing*. Cambridge University Press. http://doi.org/10.1017/CBO9780511621024

Biber, D., & Conrad, S. (2009). *Register, Genre, and Style*. Cambridge University Press. http://doi.org/10.1017/CBO9780511814358

Biber, D., Conrad, S., & Reppen, R. (1998). *Corpus linguistics : investigating language structure and use*. Cambridge University Press. Retrieved from https://books.google.es/books?hl=en&lr=&id=2h5F7TXa6psC&oi=fnd&pg=PR9&dq=Corpus+linguistics&ots=8QmbvuynjA&sig=KG6umOH2ZARVgNkiulDakHqvMBw&redir_esc=y#v=onepage&q=Corpus linguistics&f=false

Biber, D., Johansson, S., Leech, G., Conrad, S., & Finegan, E. (1999). *Longman Grammar of Spoken and Written English*. Longman. http://doi.org/10.2307/3587792

Biber, D., & Barbieri, F. (2007). Lexical bundles in university spoken and written registers. *English for Specific Purposes*, *26*(3), 263–286. http://doi.org/10.1016/j.esp.2006.08.003

Biber, D., Conrad, S., & Cortes, V. (2004). If you look at ...: Lexical bundles in university teaching and textbooks. *Applied Linguistics*. http://doi.org/10.1093/applin/25.3.371

Biber, D., Conrad, S., & Reppen, R. (1996). Corpus-Based Investigations of Language Use. *Annual Review of Applied Linguistics*, *16*, 115. http://doi.org/10.1017/S0267190500001471

Biber, D., & Finegan, E. (2001). Intra-textual variation within medical research articles. In ed. by N. O. and P. de Haan (Ed.), *Variation in English: Multi-dimensional Studies* (pp. 108–123). Rodopi.

Bidmon, S., & Terlutter, R. (2015). Gender differences in searching for health information on the internet and the virtual patient-physician relationship in Germany: Exploratory results on how men and women differ and why. *Journal of Medical Internet Research*, *17*(6), e156. http://doi.org/10.2196/jmir.4127

Blackman, J., & Sahebjalal, M. (2014). Patient understanding of frequently used cardiology terminology. *British Journal of Cardiology*, *21*(1), 39. http://doi.org/10.5837/bjc.2014.007

Blalock, S. J., Sage, A., Bitonti, M., Patel, P., Dickinson, R., & Knapp, P. (2016). Communicating information concerning potential medication harms and benefits: What gist do numbers convey? *Patient Education and Counseling*, *99*(12), 1964–1970. http://doi.org/10.1016/j.pec.2016.07.022

Blomberg, F., Brulin, C., Andertun, R., & Rydh, A. (2010). Patients' Perception of Quality of Care in a Radiology Department: A Medical-Physical Approach. *Journal of Radiology Nursing*, *29*(1), 10–17. http://doi.org/10.1016/j.jradnu.2009.09.001

Bolderston, A., Palmer, C., Flanagan, W., & McParland, N. (2008). The experiences of English as second language radiation therapy students in the undergraduate clinical program: Perceptions of staff and students. *Radiography*, *14*(3), 216–225. http://doi.org/10.1016/j.radi.2007.03.006

Booth, L. A., & Manning, D. J. (2006). Observations of radiographer communication: An exploratory study using Transactional Analysis. *Radiography*, *12*(4), 276–282. http://doi.org/10.1016/j.radi.2005.09.005

Bostock, S., & Steptoe, A. (2012). Association between low functional health literacy and mortality in older adults: longitudinal cohort study. *BMJ (Clinical Research Ed.)*, *344*, e1602. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/22422872

Boundouki, G., Humphris, G., & Field, A. (2004). Knowledge of oral cancer, distress and screening intentions: longer term effects of a patient information leaflet. *Patient Education and Counseling*, *53*(1), 71–7. http://doi.org/10.1016/S0738-3991(03)00118-6

Bourhis, R. Y., Roth, S., & MacQueen, G. (1989). Communication in the hospital setting: a survey of medical and everyday language use amongst patients, nurses and doctors. *Social Science & Medicine (1982)*, *28*(4), 339–46. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/2705006

Boyle, C. M. (1970). Difference between patients' and doctors' interpretation of some common medical terms. *British Medical Journal*, *2*(5704), 286–289. http://doi.org/10.1136/bmj.2.5704.286

Buckton, C. H., Lean, M. E. J., & Combet, E. (2015). 'Language is the source of misunderstandings'–impact of terminology on public perceptions of health promotion messages. *BMC Public Health*, *15*(1), 579. http://doi.org/10.1186/s12889-015-1884-1

Budgell, B. (n.d.). Titles of Biomedical Articles: a corpus-based analysis. Retrieved from www.icmje.org

Burton, A. K., Waddell, G., Tillotson, K. M., & Summerton, N. (1999). Information and advice to patients with back pain can have a positive effect. A randomized controlled trial of a novel educational booklet in primary care. *Spine*, *24*(23), 2484–91. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/10626311

Campbell, D. A., & Johnson, S. B. (2001). Comparing syntactic complexity in medical and non-medical corpora. *Proceedings / AMIA ... Annual Symposium. AMIA Symposium*, 90–4. Retrieved from http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2243419&tool=pmcentrez&rendertype=abstract

Candlin, C. N., & Candlin, S. (2003). HEALTH CARE COMMUNICATION: A PROBLEMATIC SITE FOR APPLIED LINGUISTICS RESEARCH. *Annual Review of Applied Linguistics*, *23*(134). Retrieved from https://www.researchonline.mq.edu.au/vital/access/services/Download/mq:5617/DS01?view=true

Candlin, S. (2003). Issues Arising when the Professional Workplace is the Site of Applied Linguistic Research. *Applied Linguistics*, *24*(3), 386–394. http://doi.org/10.1093/applin/24.3.386

Carnet, D., & Magnet, A. (2006). Medical Journals, Letters to the Editor. In K. Brown (Ed.), *Encyclopedia of Language and Lingiustics*, (2nd ed., pp. 725–732). Elsevier Science.

Carter, R. A., & McCarthy, M. J. (1995). Grammar and the Spoken Language.pdf. *Applied Linguistics*, *16, 141–158*.

Chang, M. Y., Kim, J.-W., & Rhee, C.-S. (2015). The quality of health information on allergic rhinitis, rhinitis, and sinusitis available on the internet. *Allergy, Asthma & Immunology Research*, *7*(2), 141–7. http://doi.org/10.4168/aair.2015.7.2.141

Chapman, K., Abraham, C., Jenkins, V., & Fallowfield, L. (2003). Lay understanding of terms used in cancer consultations. *Psycho-Oncology*, *12*(6), 557–566. http://doi.org/10.1002/pon.673

Charnock, D., Shepperd, S., Needham, G., & Gann, R. (1999). DISCERN: An instrument for judging the quality of written consumer health information on treatment choices. *Journal of Epidemiology and Community Health*, *53*(2), 105–111. http://doi.org/10.1136/jech.53.2.105

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. http://doi.org/10.1016/j.esp.2007.04.003

Cherla, D. V., Sanghvi, S., Choudhry, O. J., Liu, J. K., & Eloy, J. A. (2012). Readability assessment of Internet-based patient education materials related to endoscopic sinus surgery. *The Laryngoscope*, *122*(8), 1649–54. http://doi.org/10.1002/lary.23309

Chin-Mo, C., & Yu-Chin, C. (2011). EU opt in and out under the framework of Lisbon treaty. *Tamkang Journal of International Affairs*, *14*(3), 95–128. Retrieved from http://dialnet.unirioja.es/servlet/dcart?info=link&codigo=2475568&orden=140891

Chu, J. T., Wang, M. P., Shen, C., Viswanath, K., Lam, T. H., & Chan, S. S. C. (2017). How, When and Why People Seek Health Information Online: Qualitative Study in Hong Kong. *Interactive Journal of Medical Research*, *6*(2), e24. http://doi.org/10.2196/ijmr.7000

Chur-Hansen, A., & Barrett, R. J. (1996). Teaching colloquial Australian English to medical students from non-English speaking backgrounds. *Medical Education*, *30*(6), 412–417. http://doi.org/10.1111/j.1365-2923.1996.tb00860.x

Chur-Hansen, A., Vernon-Roberts, J., & Clark, S. (1997). Language background, English language proficiency and medical communication skills of medical students. *Medical Education*, *31*(4), 259–263. http://doi.org/10.1111/j.1365-2923.1997.tb02922.x

Cilveti, L. D., Pérez, I. K. L., La, U. De, & Tenerife, L. (n.d.). Textual and language flaws : problems for Spanish doctors in producing abstracts in English, 61–79.

Claramita, M., Utarini, A., Soebono, H., Van Dalen, J., & Van der Vleuten, C. (2011). Doctor–patient communication in a Southeast Asian setting: the conflict between ideal and reality. *Advances in Health Sciences Education*, *16*(1), 69–80. http://doi.org/10.1007/s10459-010-9242-7

Clerehan, R., & Buchbinder, R. (2006). Toward a more valid account of functional text quality: The case of the patient information leaflet. *Text and Talk*. http://doi.org/10.1515/TEXT.2006.003

Clerehan, R., Buchbinder, R., & Moodie, J. (2005). A linguistic framework for assessing the quality of written patient information: Its use in assessing methotrexate information for rheumatoid arthritis. *Health Education Research*, *20*(3), 334–344. http://doi.org/10.1093/her/cyg123

Close, J., & Aarts, B. (2008). *Changes in the use of the modals HAVE TO, HAVE GOT TO and MUST*. Retrieved from https://www.ucl.ac.uk/english-usage/projects/verb-phrase/Close and Aarts ICEHL 15.pdf

Collins, P. (2009). Modals and quasi-modals in {W}orld {E}nglishes. *World Englishes*, *28*(3), 281–292. Retrieved from https://books.google.ch/books?id=GIDgo3L1MEsC&source=gbs_slider_cls_metadata_1_mylibrary

Collins, S. (2005). Explanations in consultations: the combined effectiveness of doctors' and nurses' communication with patients. *Medical Education*, *39*(8), 785–796. http://doi.org/10.1111/j.1365-2929.2005.02222.x

Connor, U., Antón, M., Goering, E., Lauten, K., Roach, P., Balunda, S., & Hayat, A. (2013). Listening to patients' voices: Linguistic indicators related to diabetes self-management. *Communication & Medicine*, *9*(1). http://doi.org/10.1558/cam.v9i1.1

Conrad, S. (2015). Register variation. In *The Cambridge Handbook of English Corpus Linguistics* (pp. 309–329). http://doi.org/10.1007/9781139764377.018

Conrad, S., & Biber, D. (2005). The Frequency and Use of Lexical Bundles in Conversation and Academic Prose. *Lexicographica*, *20/2004*, 56–71. http://doi.org/10.1515/9783484604674.56

Cortes, V. (2004). Lexical bundles in published and student disciplinary writing: Examples from history and biology. *English for Specific Purposes*, *23*(4), 397–423. http://doi.org/10.1016/J.ESP.2003.12.001

Cortes, V. (2006). Teaching lexical bundles in the disciplines: An example from a writing intensive history class. *Linguistics and Education*, *17*(4), 391–406. http://doi.org/10.1016/J.LINGED.2007.02.001

Cortes, V. (2013). The purpose of this study is to: Connecting lexical bundles and moves in research article introductions. *Journal of English for Academic Purposes*, *12*(1), 33–43. http://doi.org/10.1016/J.JEAP.2012.11.002

Coulter, A., & Ellins, J. (2007). Effectiveness of strategies for informing, educating, and involving patients. *BMJ (Clinical Research Ed.)*, *335*(7609), 24–7. http://doi.org/10.1136/bmj.39246.581169.80

Crawford, P., & Brown, B. (2011). Fast healthcare: Brief communication, traps and opportunities. *Patient Education and Counseling*, *82*(1), 3–10. http://doi.org/10.1016/j.pec.2010.02.016

Cross, C., & Oppenheim, C. (2006). A genre analysis of scientific abstracts. *Journal of Documentation*, *62*(4), 428–446. http://doi.org/10.1108/00220410610700953

Crosthwaite, P., Cheung, L., & Jiang, F. (Kevin). (2017). Writing with attitude: Stance expression in learner and professional dentistry research reports. *English for Specific Purposes*, *46*. http://doi.org/10.1016/j.esp.2017.02.001

Csomay, E. (2013). Lexical Bundles in Discourse Structure: A Corpus-Based Study of Classroom Discourse. *Applied Linguistics*, *34*(3), 369–388. http://doi.org/10.1093/applin/ams045

Curtis, J. R., Back, A. L., Ford, D. W., Downey, L., Shannon, S. E., Doorenbos, A. Z., … Engelberg, R. A. (2013). Effect of communication skills training for residents and nurse practitioners on quality of communication with patients with serious illness: a randomized trial. *JAMA*, *310*(21), 2271–81. http://doi.org/10.1001/jama.2013.282081

Cutting, J. (2007). *Vague language explored*. (Joan Cutting, Ed.)*Vague Language Explored*. http://doi.org/10.1057/9780230627420

Dahm, M. R. (2011). Exploring perception and use of everyday language and medical terminology among international medical graduates in a medical ESP course in Australia. *English for Specific Purposes*, *30*(3), 186–197. http://doi.org/10.1016/j.esp.2011.02.004

Dahm, M. R. (2018). A socio-cognitive investigation of English medical terminology: dynamic varieties of meaning. *Lexicography*, *4*(1), 81–103. http://doi.org/10.1007/s40607-018-0039-9

Dahm, M. R., & Yates, L. (2013). English for the Workplace: Doing Patient-Centred Care in Medical Communication. *TESL Canada Journal*, *30*(7), 21. Retrieved from http://teslcanadajournal.ca/index.php/tesl/article/view/1150

Dahm, M. R., Yates, L., Ogden, K., Rooney, K., & Sheldon, B. (2015). Enhancing international medical graduates' communication: The contribution of applied linguistics. *Medical Education*, *49*(8), 828–837. http://doi.org/10.1111/medu.12776

Davey, H. M., Butow, P. N., & Armstrong, B. K. (2003). Cancer patients' preferences for written prognostic information provided outside the clinical context. *British Journal of Cancer*, *89*(8), 1450–6. http://doi.org/10.1038/sj.bjc.6601287

De Bont, E. G. P. M., Alink, M., Falkenberg, F. C. J., Dinant, G. J., & Cals, J. W. L. (2015). Patient information leaflets to reduce antibiotic use and reconsultation rates in general practice: A systematic review. *BMJ Open*, *5*(6), e007612–e007612. http://doi.org/10.1136/bmjopen-2015-007612

DeCarrico, J., & Nattinger, J. R. (1988). Lexical phrases for the comprehension of academic lectures. *English for Specific Purposes*, *7*(2), 91–102. http://doi.org/10.1016/0889-4906(88)90027-0

Demark-Wahnefried, W., Peterson, B., McBride, C., Lipkus, I., & Clipp, E. (2000). Current health behaviors and readiness to pursue life-style changes among men and women diagnosed with early stage prostate and breast carcinomas. *Cancer*, *88*(3), 674–684. http://doi.org/10.1002/(SICI)1097-0142(20000201)88:3<674::AID-CNCR26>3.0.CO;2-R

Diaz, J. A., Griffith, R. A., Ng, J. J., Reinert, S. E., Friedmann, P. D., & Moulton, A. W. (2002). Patients' use of the Internet for medical information. *Journal of General Internal Medicine*, *17*(3), 180–5. http://doi.org/10.1046/J.1525-1497.2002.10603.X

Dickinson, D., Raynor, D. K., & Duman, M. (2001). Patient information leaflets for medicines: using consumer testing to determine the most effective design. *Patient Education and Counseling*, *43*(2), 147–59. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/11369148

Dieltjens, S. M., Heynderickx, P. C., Dees, M. K., & Vissers, K. C. (2014). Linguistic Analysis of Face-to-Face Interviews with Patients with An Explicit Request for Euthanasia, their Closest Relatives, and their Attending Physicians: the Use of Modal Verbs in Dutch. *Pain Practice*, *14*(4), 324–331. http://doi.org/10.1111/papr.12076

Dixon-Woods, M. (2001). Writing wrongs? An analysis of published discourses about the use of patient information leaflets. *Social Science & Medicine*, *52*(9), 1417–1432. http://doi.org/10.1016/S0277-9536(00)00247-1

Dong, S., Butow, P. N., Costa, D. S. J., Dhillon, H. M., & Shields, C. G. (2014). The influence of patient-centered communication during radiotherapy education sessions on post-consultation patient outcomes. *Patient Education and Counseling*, *95*(3), 305–312. http://doi.org/10.1016/j.pec.2014.02.008

Douglas, B. (2006). Stance in spoken and written university registers. *Journal of English for Academic Purposes*, *5*(2), 97–116. http://doi.org/10.1016/j.jeap.2006.05.001

Drew, P., Chatwin, J., & Collins, S. (2001). Conversation analysis: a method for research into interactions between patients and health-care professionals. *Health Expectations*, *4*(1), 58–70. http://doi.org/10.1046/j.1369-6513.2001.00125.x

Durrant, P. (2017). Lexical Bundles and Disciplinary Variation in University Students' Writing: Mapping the Territories. *Applied Linguistics*, *38*(2), 165–193. http://doi.org/10.1093/applin/amv011

Dutta-Bergman, M. (2003). Trusted online sources of health information: differences in demographics, health beliefs, and health-information orientation. *Journal of Medical Internet Research*, *5*(3), e21. http://doi.org/10.2196/jmir.5.3.e21

Eheman, C. R., Berkowitz, Z., Lee, J., Mohile, S., Purnell, J., Rodriguez, E. M., … Morrow, G. (2009). Information-seeking styles among cancer patients before and after treatment by demographics and use of information sources. *Journal of Health Communication*, *14*(5), 487–502. http://doi.org/10.1080/10810730903032945

Ehrlich, R. A., & Coakes, D. M. (2017). *Patient Care in Radiography: With an Introduction to Medical Imaging*. Retrieved from https://books.google.ch/books/about/Patient_Care_in_Radiography.html?id=A55QPgAACAAJ&pgis=1

Ek, S. (2013). Gender differences in health information behaviour: a Finnish population-based survey. *Health Promotion Internation*. http://doi.org/10.1093/heapro/dat063

Esfandiari, R., & Barbary, F. (2017). A contrastive corpus-driven study of lexical bundles between English writers and Persian writers in psychology research articles. *Journal of English for Academic Purposes*, *29*, 21–42. http://doi.org/10.1016/j.jeap.2017.09.002

Eysenbach, G., & Köhler, C. (2002). How do consumers search for and appraise health information on the world wide web? Qualitative study using focus groups, usability tests, and in-depth interviews. *Bmj*, *324*(7337), 573–577. http://doi.org/10.1136/bmj.324.7337.573

Fage-Butler, A. (2011). The discursive construction of risk and trust in patient information leaflets. *Hermes*, *46*, 61–74. Retrieved from www.medicines.org.uk

Fage-Butler, A. (2013). Improving patient information leaflets: Developing and applying an evaluative model of patient-centredness for text. *Communication and Medicine*, *10*(2), 105–115. http://doi.org/10.1558/cam.v10i2.105

Fage-Butler, A. M., & Nisbeth Jensen, M. (2016). Medical terminology in online patient-patient communication: evidence of high health literacy? *Health Expectations : An International Journal of Public Participation in Health Care and Health Policy*, *19*(3), 643–53. http://doi.org/10.1111/hex.12395

Fakult, P., Seggewi, F., & Ss, H. (2012). Current Changes in the English Modals – a Corpus-Based Analysis of Present-Day Spoken English.

Fehringer C, C. K. (2015). " You ’ ve got to sort of eh hoy the Geordie out ": Modals of obligation and necessity in 50 years of Tyneside English . English Language and Linguistics 2015 , 19 (Special Issue 2), 355-381 . Copyright : Date deposited :, *19*(2), 355–381.

Ferguson, G. (2001). If you pop over there: a corpus-based study of conditionals in medical discourse. *English for Specific Purposes*, *20*(1), 61–82. http://doi.org/10.1016/S0889-4906(99)00027-7

Fields, A. M., Freiberg, C. S., Fickenscher, A., & Shelley, K. H. (2008). Patients and jargon: are we speaking the same language? *Journal of Clinical Anesthesia*, *20*(5), 343–346. http://doi.org/10.1016/j.jclinane.2008.02.006

Fleischman, S. (2008). Language and Medicine. In *The Handbook of Discourse Analysis* (Vol. 19, pp. 470–502). Cambridge University Press. http://doi.org/10.1002/9780470753460.ch25

Flight, I. H., Wilson, C. J., Zajac, I. T., Hart, E., & McGillivray, J. A. (2012). Decision Support and the Effectiveness of Web-based Delivery and Information Tailoring for Bowel Cancer Screening: An Exploratory Study. *JMIR Research Protocols*, *1*(2), e12. http://doi.org/10.2196/resprot.2135

Flowerdew, J. (2015). John Swales's approach to pedagogy in Genre Analysis: A perspective from 25 years on. *Journal of English for Academic Purposes*, *19*, 102–112. http://doi.org/10.1016/j.jeap.2015.02.003

Flowerdew, L. (2005). An integration of corpus-based and genre-based approaches to text analysis in EAP/ESP: Countering criticisms against corpus-based methodologies. *English for Specific Purposes*, *24*(3), 321–332. http://doi.org/10.1016/j.esp.2004.09.002

Fox, N. J., Ward, K. J., & O'rourke, A. J. (2005). The "expert patient": empowerment or medical dominance? The case of weight loss, pharmaceutical drugs and the Internet. *Social Science & Medicine*, *60*, 1299–1309. http://doi.org/10.1016/j.socscimed.2004.07.005

Fox, N., & Ward, K. (2006). Health identities: from expert patient to resisting consumer. *Health: An Interdisciplinary Journal for the Social Study of Health, Illness and Medicine*, *10*(4), 461–479. http://doi.org/10.1177/1363459306067314

Frank, R. A. (2000). Medical communication: non-native English speaking patients and native English speaking professionals. *English for Specific Purposes*, *19*(1), 31–62. http://doi.org/10.1016/S0889-4906(98)00012-X

Friedman, D. B., & Hoffman-Goetz, L. (2006). A Systematic Review of Readability and Comprehension Instruments Used for Print and Web-Based Cancer Information. *Health Education & Behavior*, *33*(3), 352–373. http://doi.org/10.1177/1090198105277329

Fuertes Olivera, P. a. (2008). Advances in Medical Discourse Analysis: Oral and Written Contexts. *Iberica*, *15*(2), 186–191.

Gaissmaier, W., & Gigerenzer, G. (2008). Statistical illiteracy undermines informed shared decision making. *Zeitschrift Fur Evidenz, Fortbildung Und Qualitat Im Gesundheitswesen*, *102*(7), 411–413. http://doi.org/10.1016/j.zefq.2008.08.013

Gajšt, N. (2014). English Modal Verbs in Modal Verb Phrase Structures and Lexical Associations of Modality in 'Standard Terms and Conditions of Sale.' *Journal for Foreign Languages*, *4*(1–2), 243. http://doi.org/10.4312/vestnik.4.243-260

Gamble, M. (2014). Medical Malpractice in America: 15 Latest Statistics. Retrieved November 17, 2018, from https://www.beckershospitalreview.com/legal-regulatory-issues/medical-malpractice-in-america-15-latest-statistics.html

Garner, M., Ning, Z., & Francis, J. (2012). A framework for the evaluation of patient information leaflets. *Health Expectations : An International Journal of Public Participation in Health Care and Health Policy*, *15*(3), 283–294. http://doi.org/10.1111/j.1369-7625.2011.00665.x

Gasiorek, J., Van de Poel, K., & Blockmans, I. (2015). What do you do when you can't accommodate? Managing and evaluating problematic interactions in a multilingual medical environment. *Language & Communication*, *41*, 84–88. http://doi.org/10.1016/j.langcom.2014.10.005

Gaspari, F. (2013). A phraseological comparison of international news agency reports published online: Lexical bundles in the English-language output of ANSA, Adnkronos, Reuters and UPI. *VARIENG*, *13*.

Gelb, K., Pederson, A., & Greaves, L. (2012). How have health promotion frameworks considered gender? *Health Promotion International*, *27*(4), 445–452. http://doi.org/10.1093/heapro/dar087

Gemoets, D., Rosemblat, G., Tse, T., & Logan, R. (2004). Assessing readability of consumer health information: An exploratory study. *Studies in Health Technology and Informatics*, *107*, 869–873. http://doi.org/10.3233/978-1-60750-949-3-869

Ghafar Samar, R., Shokrpour, N., & Nasiri, E. (2018). The Applicability of Teaching Lexical Bundles on Medical Students' Writing Proficiency in an EFL Context. *Teaching English Language*, *12*(2), 27–44. http://doi.org/10.22132/TEL.2018.70728

Gillies, K., Huang, W., Skea, Z., Brehaut, J., & Cotton, S. (2014). Patient information leaflets (PILs) for UK randomised controlled trials: A feasibility study exploring whether they contain information to support decision making about trial participation. *Trials*, *15*(1). http://doi.org/10.1186/1745-6215-15-62

Gilmore, A., & Millar, N. (2018). The language of civil engineering research articles: A corpus-based approach. *English for Specific Purposes*, *51*, 1–17. http://doi.org/10.1016/J.ESP.2018.02.002

Giménez-Moreno, R., & Skorczynska, H. (2013). Corpus Analysis and Register Variation: A Field in Need of an Update. *Procedia - Social and Behavioral Sciences*, *95*, 402–408. http://doi.org/10.1016/J.SBSPRO.2013.10.662

Gisborne, N. (2007). Dynamic Modality. *Journal of Theoretical Linguistics, 4*(2), 44–61.

Glass, L. (2015). Grounding the social meaning of a strong necessity modal in its semantics (pp. 1–25).

Gledhill, C. (1995). *Scientific innovation and the phraseology of rhetoric: Posture, reformulation and collocation in cancer research articles* (Vol. Dissertati). Retrieved from http://www.isfla.org/Systemics/Print/Theses/Gledhill1995.pdf

Godolphin, W., Towle, A., & McKendry, R. (2001). Evaluation of the quality of patient information to support informed shared decision-making. *Health Expectations : An International Journal of Public Participation in Health Care and Health Policy*, *4*(4), 235–42. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/11703497

Gooberman-Hill, R., French, M., Dieppe, P., & Hawker, G. (2009). Expressing pain and fatigue: A new method of analysis to explore differences in osteoarthritis experience. *Arthritis Care and Research*, *61*(3), 353–360. http://doi.org/10.1002/art.24273

Goodwin, J. S. (2002). Language and Medicine: Thinking and Talking About Alzheimer's Disease. *The Gerontologist*, *42*(3), 293–296. http://doi.org/10.1093/geront/42.3.293

Gotti, M. (2015). Insights into medical discourse: diachronic and synchronic perspectives. *Esp Today-Journal of English for Specific Purposes at Tertiary Level*, *3*(1), 5–24. Retrieved from https://aisberg.unibg.it/retrieve/handle/10446/42776/44283/Gotti - Insights into medical discourse.pdf

Gozdz-Roszkowski, S. (2011). *Patterns of linguistic variation in American legal English : a corpus-based study.* (Vol. 22). Frankfurt am Main: Peter Lang.

Grabowski, Ł. (2013). Register Variation Across English Pharmaceutical Texts: A Corpus-driven Study of Keywords, Lexical Bundles and Phrase Frames in Patient Information Leaflets and Summaries of Product Characteristics. *Procedia - Social and Behavioral Sciences*, *95*, 391–401. http://doi.org/10.1016/j.sbspro.2013.10.661

Grabowski, Ł. (2014). On Lexical Bundles in Polish Patient Information Leaflets: A Corpus-Driven Study. *Http://Www.Ejournals.Eu/SPL/, Volume 9*(Issue 1), 21–43. http://doi.org/10.4467/23005920SPL.14.002.2186

Grabowski, Ł. (2015). Keywords and lexical bundles within English pharmaceutical discourse: A corpus-driven description. *English for Specific Purposes*, *38*, 23–33. http://doi.org/10.1016/j.esp.2014.10.004

Granhagen Jungner, J., Tiselius, E., Wenemark, M., Blomgren, K., Lützén, K., & Pergert, P. (2018). Development and evaluation of the Communication over Language Barriers questionnaire (CoLB-q) in paediatric healthcare. *Patient Education and Counseling*, *101*(9), 1661–1668. http://doi.org/10.1016/j.pec.2018.04.007

Greaves, F., Ramirez-cano, D., Millett, C., Darzi, A., & Donaldson, L. (2010). Machine learning and sentiment analysis of unstructured free-text information about patient experience online, 2010.

Grime, J. C., & Ong, B. N. (2007). Constructing osteoarthritis through discourse - A qualitative analysis of six patient information leaflets on osteoarthritis. *BMC Musculoskeletal Disorders*, *8*(1), 34. http://doi.org/10.1186/1471-2474-8-34

Grime, J., & Pollock, K. (2004). Information versus experience: A comparison of an information leaflet on antidepressants with lay experience of treatment. *Patient Education and Counseling*, *54*(3), 361–368. http://doi.org/10.1016/j.pec.2004.02.003

Hadjipavlou, M., Khan, S., & Rane, a. (2013). Readability of patient information leaflets for urological conditions and treatments. *Journal of Clinical Urology*, *6*(5), 302–305. http://doi.org/10.1177/2051415813489554

Hadlow, J., & Pitts, M. (1991). The understanding of common health terms by doctors, nurses and patients. *Social Science & Medicine (1982)*, *32*(2), 193–6. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/2014414

Halkett, G. K. B., & Kristjanson, L. J. (2007). Patients' perspectives on the role of radiation therapists. *Patient Education and Counseling*, *69*(1–3), 76–83. http://doi.org/10.1016/j.pec.2007.07.004

Halkett, G. K. B., Schofield, P., O'Connor, M., York, D., Jefford, M., Jiwa, M., … Aranda, S. (2012). Development and pilot testing of a radiation therapist-led educational intervention for breast cancer patients prior to commencing radiotherapy. *Asia-Pacific Journal of Clinical Oncology*, *8*(3), e1-8. http://doi.org/10.1111/j.1743-7563.2012.01520.x

Halkett, G. K. B., Short, M., & Kristjanson, L. J. (2009). How do radiation oncology health professionals inform breast cancer patients about the medical and technical aspects of their treatment? *Radiotherapy and Oncology : Journal of the European Society for Therapeutic Radiology and Oncology*, *90*(1), 153–9. http://doi.org/10.1016/j.radonc.2008.05.017

Hamarneh, A. (2015). Lack of language skills and knowledge of local culture in international medical graduates: Implications for the NHS. *Hospital Practice*

(1995), *43*(4), 208–11. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/26234814

Hamilton, C., Adolphs, S., & Nerlich, B. (2007). The meanings of "risk": A view from corpus linguistics. *Discourse and Society*, *18*(2), 163–181. http://doi.org/10.1177/0957926507073374

Hamilton, H. E., & Chou, W. S. (2010). *The Routledge Handbook of Language and Health Communication*. Routledge. http://doi.org/10.4324/9781315856971

Hamilton, J., & Woodward-Kron, R. (2010). Developing cultural awareness and intercultural communication through multimedia: A case study from medicine and the health sciences. *System*, *38*(4), 560–568. http://doi.org/10.1016/j.system.2010.09.015

Hamrosi, K. K., Aslani, P., & Raynor, D. K. (2014). Beyond needs and expectations: Identifying the barriers and facilitators to written medicine information provision and use in Australia. *Health Expectations*, *17*(2). http://doi.org/10.1111/j.1369-7625.2011.00753.x

Hamrosi, K. K., Raynor, D. K., & Aslani, P. (2013). Pharmacist and general practitioner ambivalence about providing written medicine information to patients-A qualitative study. *Research in Social and Administrative Pharmacy*, *9*(5). http://doi.org/10.1016/j.sapharm.2013.02.006

Hamrosi, K. K., Raynor, D. K., & Aslani, P. (2014). Pharmacist, general practitioner and consumer use of written medicine information in Australia: Are they on the same page? *Research in Social and Administrative Pharmacy*, *10*(4). http://doi.org/10.1016/j.sapharm.2013.10.002

Hamrosi, K. K., Raynor, D. K., & Aslani, P. (2014). Enhancing provision of written medicine information in Australia: pharmacist, general practitioner and consumer perceptions of the barriers and facilitators. *BMC Health Services Research*, *14*. http://doi.org/10.1186/1472-6963-14-183

Hancock, J. T., Woodworth, M. T., & Porter, S. (2013). Hungry like the wolf: A word-pattern analysis of the language of psychopaths. *Legal and Criminological Psychology*, *18*(1), 102–114. http://doi.org/10.1111/j.2044-8333.2011.02025.x

Hansberry, D. R., John, A., John, E., Agarwal, N., Gonzales, S. F., & Baker, S. R. (2014). A Critical Review of the Readability of Online Patient Education Resources From RadiologyInfo.Org. *American Journal of Roentgenology*, *202*(3), 566–575. http://doi.org/10.2214/AJR.13.11223

Hansberry, D. R., Ramchand, T., Patel, S., Kraus, C., Jung, J., Agarwal, N., … Baker, S. R. (2014). Are we failing to communicate? Internet-based patient education materials and radiation safety. *European Journal of Radiology*, *83*(9), 1698–1702. http://doi.org/10.1016/j.ejrad.2014.04.013

Harvey, K., & Brown, B. (2012). Health Communication and Psychological Distress: Exploring the Language of Self-harm. *Canadian Modern Language Review*. http://doi.org/10.3138/cmlr.1103

Harvey, K., Churchill, D., Crawford, P., Brown, B., Mullany, L., Macfarlane, A., … Health, M. A. (2008). Health communication and adolescents: what do their emails tell us? *Family Practice*, *25*(4), 304–311. http://doi.org/10.1093/fampra/cmn029

Harvey, K., Locher, M. A., & Mullany, L. (2013). "Can I Be at Risk of Getting AIDS?" A Linguistic Analysis of Two Internet Columns on Sexual Health *. *Linguistik Online*, *59*(2/13), 111–132. Retrieved from https://edoc.unibas.ch/28087/1/20131011103809_5257b8f1255f6.pdf

Haw, C., & Stubbs, J. (2011). Patient information leaflets for antidepressants: Are patients getting the information they need? *Journal of Affective Disorders*, *128*(1–2), 165–170. Retrieved from https://www.sciencedirect.com/science/article/pii/S0165032710004623

Hayes, E., Dua, R., Yeung, E., & Fan, K. (2018). Patient understanding of commonly used oral medicine terminology. *British Dental Journal*, *223*(11), 842. http://doi.org/10.1038/sj.bdj.2017.991

HAYES, S. C., & FARNILL, D. (1993). Medical training and English language proficiency. *Medical Education*, *27*(1), 6–14. http://doi.org/10.1111/j.1365-2923.1993.tb00222.x

Henderson, D., McGrath, P. D., & Patton, M. A. (2017). Experience of clinical supervisors of international medical graduates in an Australian district hospital. *Australian Health Review : A Publication of the Australian Hospital Association*, *41*(4), 365–371. http://doi.org/10.1071/AH15094

Herber, O. R., Gies, V., Schwappach, D., Thürmann, P., & Wilm, S. (2014). Patient information leaflets: informing or frightening? A focus group study exploring patients' emotional reactions and subsequent behavior towards package leaflets of commonly prescribed medications in family practices. *BMC Family Practice*, *15*(1). http://doi.org/10.1186/1471-2296-15-163

Herfs, P. G. P. (2014). Aspects of medical migration with particular reference to the United Kingdom and the Netherlands. *Human Resources for Health*, *12*(1), 59. http://doi.org/10.1186/1478-4491-12-59

Heritage, J., & Sefi, S. (1992). Dilemmas of advice: Aspects of the delivery and reception of advice in interactions between healh visitors and first-time mothers. *Talk at Work: Interaction in Institutional Settings*. Retrieved from http://www.sscnet.ucla.edu/soc/faculty/heritage/Site/Publications_files/DILEMMAS_OF_ADVICE.pdf

Hewett, D. G., Watson, B. M., & Gallois, C. (2015). Communication between hospital doctors: Underaccommodation and interpretability. *Language & Communication*, *41*, 71–83. http://doi.org/10.1016/j.langcom.2014.10.007

Hibbard, J. H., Peters, E., Dixon, A., & Tusler, M. (2007). Consumer competencies and the use of comparative quality information: It isn't just about literacy. *Medical Care Research and Review*, *64*(4), 379–394. http://doi.org/10.1177/1077558707301630

Hibbard, J. H., Peters, E., Dixon, A., & Tusler, M. (2007). Consumer Competencies and the Use of Comparative Quality Information. *Medical Care Research and Review*, *64*(4), 379–394. http://doi.org/10.1177/1077558707301630

Hinds, C., Streater, A., & Mood, D. (1995). Functions and preferred methods of receiving information related to radiotherapy. Perceptions of patients with cancer. *Cancer Nursing*, *18*(5), 374–84. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/7585492

Hinkel, E. (1995). The Use of Modal Verbs as a Reflection of Cultural Values. *TESOL Quarterly*, *29*(2), 325–343. http://doi.org/10.2307/3587627

Hirsh, D., Clerehan, R., Staples, M., Osborne, R. H., & Buchbinder, R. (2009). Patient assessment of medication information leaflets and validation of the Evaluative Linguistic Framework (ELF). *Patient Education and Counseling*, *77*(2), 248–254. http://doi.org/10.1016/j.pec.2009.03.011

Hoekje, B. J. (2007). Medical discourse and ESP courses for international medical graduates (IMGs). *English for Specific Purposes, 26*(3), 327–343. http://doi.org/10.1016/j.esp.2006.09.002

Hoffmann, K., Ristl, R., Heschl, L., Stelzer, D., & Maier, M. (2014). Antibiotics and their effects: what do patients know and what is their source of information? *The European Journal of Public Health*, *24*(3), 502–507. http://doi.org/10.1093/eurpub/ckt112

Hoffmann, T. C., & Del Mar, C. (2017). Clinicians' Expectations of the Benefits and Harms of Treatments, Screening, and Tests. *JAMA Internal Medicine*, *177*(3), 407. http://doi.org/10.1001/jamainternmed.2016.8254

Hoffmann, T. C., Légaré, F., Simmons, M. B., McNamara, K., McCaffery, K., Trevena, L. J., … Del Mar, C. B. (2014). Shared decision making: what do clinicians need to know and why should they bother? *The Medical Journal of Australia*, *201*(1), 35–39. http://doi.org/10.5694/mja14.00002

Hoffmann, T., McKenna, K., Worrall, L., & Read, S. J. (2007). Randomised trial of a computer-generated tailored written education package for patients following stroke. *Age and Ageing*, *36*(3), 280–6. http://doi.org/10.1093/ageing/afm003

Holmberg, O., Czarwinski, R., & Mettler, F. (2010). The importance and unique aspects of radiation protection in medicine. *European Journal of Radiology*, *76*(1), 6–10. http://doi.org/10.1016/j.ejrad.2010.06.031

Hsu, W. (2013). Bridging the vocabulary gap for EFL medical undergraduates: The establishment of a medical word list. *Language Teaching Research*, *17*(4), 454–484. http://doi.org/10.1177/1362168813494121

Hudson, R. (1994). About 37% of Word-Tokens are Nouns. *Language*, *70*(2), 331–339. http://doi.org/10.2307/415831

Hunston, S. (2006). Corpus Linguistics. In *Encyclopedia of Language & Linguistics*. http://doi.org/10.1016/B0-08-044854-2/00944-5

Hyden, L.-C., Mishler, E. G., Hydén, L., & Mishler, E. G. (2003). LANGUAGE AND MEDICINE. *Annual Review of Applied Linguistics*, *19*, 174–192. http://doi.org/10.1017/S0267190599190093

Hyland, K. (2015). Corpora and written academic English. In *The Cambridge Handbook of English Corpus Linguistics* (pp. 292–308). http://doi.org/10.1007/9781139764377.017

Hyland, K. (2002). Directives: Argument and engagement in academic writing. *Applied Linguistics*, *23*(2), 215–239. http://doi.org/10.1093/applin/23.2.215

Hyland, K. (2008). As can be seen: Lexical bundles and disciplinary variation. *English for Specific Purposes*, *27*(1), 4–21. http://doi.org/10.1016/j.esp.2007.06.001

Hyland, K. (2012). Bundles in Academic Discourse. *Annual Review of Applied Linguistics*, *3212*, 150–169. http://doi.org/10.1017/S0267190512000037

Ilić, N., Auchlin, A., Hadengue, A., Wenger, A., & Hurst, S. A. (2013). Informed Consent Forms in Oncology Research: Linguistic Tools Identify Recurrent Pitfalls. *AJOB Primary Research*, *4*(4), 39–54. http://doi.org/10.1080/21507716.2013.788101

Isaacman, D. J., Purvis, K., Gyuro, J., Anderson, Y., & Smith, D. (1992). Standardized instructions: do they improve communication of discharge information from the emergency department? *Pediatrics*, *89*(6 Pt 2), 1204–8. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/1594378

Jalali, Z. S., & Moini, M. R. (2014). Structure of Lexical Bundles in Introduction Section of Medical Research Articles. *Procedia - Social and Behavioral Sciences*, *98*, 719–726. http://doi.org/10.1016/j.sbspro.2014.03.473

Jalali, Z. S., Moini, M. R., & Arani, M. A. (2015). Structural and Functional Analysis of Lexical Bundles in Medical Research Articles: A Corpus-Based Study. *International Journal of Information Science and Management*, *13*(1), 51–69. Retrieved from http://ijism.ricest.ac.ir/index.php/ijism/article/viewFile/503/237

Jensen, J. D., King, A. J., Carcioppolo, N., & Davis, L. (2012). Why are Tailored Messages More Effective? A Multiple Mediation Analysis of a Breast Cancer Screening Intervention. *The Journal of Communication*, *62*(5), 851–868. http://doi.org/10.1111/j.1460-2466.2012.01668.x

Jin, J., Sklar, G. E., Min Sen Oh, V., & Chuen Li, S. (2008). Factors affecting therapeutic compliance: A review from the patient's perspective. *Therapeutics and Clinical Risk Management*, *4*(1), 269–86. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/18728716

Johansson, S. (2010). Modals and semi-modals of obligation in American English: Some aspects of developments from 1990 until the present day. *The Verb Phrase in English: Investigating Recent Language Change with Corpora*, 372–380. http://doi.org/10.1017/CBO9781139060998.016

Johnson, A., & Sandford, J. (2005). Written and verbal information versus verbal information only for patients being discharged from acute hospital settings to home: systematic review. *Health Education Research*, *20*(4), 423–9. http://doi.org/10.1093/her/cyg141

Jones S. (2016). The Noncompliant vs The Non-adherent Patient. Retrieved November 24, 2018, from https://www.capphysicians.com/articles/noncompliant-vs-non-adherent-patient

Juarez, G. (2011). Common english/spanish terminology use in radiology. *Journal of Radiology Nursing*, *30*(1), 9–14. http://doi.org/10.1016/j.jradnu.2010.12.001

Jung, M., Ramanadhan, S., & Viswanath, K. (2013). Effect of information seeking and avoidance behavior on self-rated health status among cancer survivors. *Patient Education and Counseling*, *92*(1), 100–106. http://doi.org/10.1016/j.pec.2013.02.008

Kalra, G., Bhugra, D. K., & Shah, N. (2012). Identifying and addressing stresses in international medical graduates. *Academic Psychiatry : The Journal of the American Association of Directors of Psychiatric Residency Training and the Association for Academic Psychiatry*, *36*(4), 323–9. http://doi.org/10.1176/appi.ap.11040085

Kampa, R. J., Pang, J., & Gleeson, R. (2006). Broken bones and fractures - an audit of patients' perceptions. *Annals of the Royal College of Surgeons of England*, *88*(7), 663–6. http://doi.org/10.1308/003588406X149192

Kandula, S., & Zeng-Treitler, Q. (2008). Creating a gold standard for the readability measurement of health texts. *AMIA ... Annual Symposium Proceedings. AMIA Symposium*, *2008*, 353–7. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/18999150

Kazemi, M., Kohandani, M., & Farzaneh, N. (2014). The Impact of Lexical Bundles on How Applied Linguistics Articles are Evaluated. *Procedia - Social and*

Behavioral Sciences, *98*, 870–875. http://doi.org/10.1016/J.SBSPRO.2014.03.494

Kecskes, I., & Kirner-Ludwig, M. (2017). “It Would Never Happen in My Country I Must Say”: A Corpus-Pragmatic Study on Asian English Learners’ Preferred Uses of Must and Should. *Corpus Pragmatics*, *1*(2), 91–134. http://doi.org/10.1007/s41701-017-0007-x

Kempe, C. B., Sullivan, K. A., & Edmed, S. L. (2013). CE the effect of varying diagnostic terminology within patient discharge information on expected mild traumatic brain injury outcome. *Clinical Neuropsychologist*, *27*(5), 762–778. http://doi.org/10.1080/13854046.2013.795245

Kenny, T. (1998). A PIL for every ill? Patient information leaflets (PILs): a review of past, present and future use. *Family Practice*, *15*(5), 471–479. http://doi.org/10.1093/fampra/15.5.471

Kershaw, A. (2003). Patient use of the Internet to obtain health information. *Nurs Times*, *99*(36), 30–32. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/14533432

Khan, J., & Laxmi, S. (2013). Does the cancer patient want to know? Results from a study in an Indian tertiary cancer center. *South Asian Journal of Cancer*, *2*(2), 57. http://doi.org/10.4103/2278-330X.110487

Kilgarriff, A. (2005). Language is never, ever, ever, random. *Corpus Linguistics and Linguistic Theory*, *1*(2), 263–275. http://doi.org/10.1515/cllt.2005.1.2.263

King Edward’s Hospital Fund for London. (1962). *Information booklets for patients : report of an enquiry by the Division of Hospital Facilities*. Retrieved from https://archive.kingsfund.org.uk/concern/published_works/

Knapp, P., Gardner, P. H., & Woolf, E. (2016). Combined verbal and numerical expressions increase perceived risk of medicine side-effects: A randomized controlled trial of EMA recommendations. *Health Expectations*, *19*(2), 264–274. http://doi.org/10.1111/hex.12344

Knight, D., Walsh, S., & Papagiannidis, S. (2017). I’m having a spring clear out: A corpus-based analysis of e-transactional discourse. *Applied Linguistics*, *38*(2), 235–257. http://doi.org/10.1093/applin/amv019

Kobayashi, L. C., Wardle, J., & von Wagner, C. (2014). Limited health literacy is a barrier to colorectal cancer screening in England: Evidence from the English Longitudinal Study of Ageing. *Preventive Medicine*, *61*, 100–105. http://doi.org/10.1016/j.ypmed.2013.11.012

Koch-Weser, S., DeJong, W., Rudd, R. E., Koch-Weser, S., DeJong, W., & Rudd, R. E. (2009). Medical word use in clinical encounters. *Health Expectations : An*

International Journal of Public Participation in Health Care and Health Policy, *12*(4), 371–82. http://doi.org/10.1111/j.1369-7625.2009.00555.x

Koch-Weser, S., Rudd, R. E., & Dejong, W. (2010). Quantifying Word Use to Study Health Literacy in Doctor-Patient Communication. *J Health Commun.*, *15*(6), 590–602.

Kopaczyk, J. (2013). Formulaic discourse across Early Modern English medical genres. *Meaning in the History of English: Words and Texts in Context*, *148*, 257–300. http://doi.org/10.1075/slcs.148.12kop

Koteyko, N., Nerlich, B., Crawford, P., & Wright, N. (2008). “Not rocket science” or “no silver bullet”? Media and government discourses about MRSA and cleanliness. *Applied Linguistics*, *29*(2), 223–243. http://doi.org/10.1093/applin/amn006

Koteyko, N., Nerlich, B., Crawford, P., & Wright, N. (2008). “Not rocket science” or “No silver bullet”? Media and Government Discourses about MRSA and Cleanliness. *Applied Linguistics*, *292*, 223–243. http://doi.org/10.1093/applin/amn006

Krankl, J. T., Shaykevich, S., Lipsitz, S., & Lehmann, L. S. (2011). Patient predictors of colposcopy comprehension of consent among English- and Spanish-speaking women. *Women’s Health Issues : Official Publication of the Jacobs Institute of Women’s Health*, *21*(1), 80–5. http://doi.org/10.1016/j.whi.2010.07.009

Lampert, A., Wien, K., Haefeli, W. E., & Seidling, H. M. (2016). Guidance on how to achieve comprehensible patient information leaflets in four steps. *International Journal for Quality in Health Care*, *28*(5), 634–638. http://doi.org/10.1093/intqhc/mzw077

Landmark, A. M. D., Gulbrandsen, P., & Svennevig, J. (2015). Whose decision? Negotiating epistemic and deontic rights in medical treatment decisions. *Journal of Pragmatics*, *78*, 54–69. http://doi.org/10.1016/j.pragma.2014.11.007

Landmark, A. M. D., Svennevig, J., Gerwing, J., & Gulbrandsen, P. (2017). Patient involvement and language barriers: Problems of agreement or understanding? *Patient Education and Counseling*, *100*(6), 1092–1102. http://doi.org/10.1016/j.pec.2016.12.006

Lawrence, H., Poggenpoel, M., & Myburgh, C. (2011). Experiences of being a therapy radiographer. *Health SA Gesondheid*, *16*(1), 7 pages. http://doi.org/10.4102/hsag.v16i1.596

Lee, D. Y. (2001). Genres, Registers, Text Types, Domains, and Styles: Clarifying the Concepts and Navigating a Path Through the Bnc Jungle. http://doi.org/10.1016/S1364-6613(00)01594-1

Leech, G. (1971). *{M}eaning and the {English} {V}erb*. Retrieved from https://books.google.ch/books?id=ISisAgAAQBAJ&source=gbs_slider_cls_metadata_1_mylibrary

Leech, G. (2004). A new Gray's Anatomy of English grammar. *English Language and Linguistics*, *8*(1), 121–147. http://doi.org/10.1017/S1360674304001273

Leech, G. (2004). Developing Linguistic Corpora: a Guide to Good Practice: Metadata for corpus work. Retrieved from https://ota.ox.ac.uk/documents/creating/dlc/chapter1.htm

Legato, M. J., Johnson, P. A., & Manson, J. E. (2016). Consideration of Sex Differences in Medicine to Improve Health Care and Patient Outcomes. *JAMA*, *316*(18), 1865. http://doi.org/10.1001/jama.2016.13995

Lei, L., & Liu, D. (2016). A new medical academic word list: A corpus-based study with enhanced methodology. *Journal of English for Academic Purposes*, *22*, 42–53. http://doi.org/10.1016/J.JEAP.2016.01.008

León-Araúz, P., & Reimerink, A. (2015). ScienceDirect Signs and symptoms in the psychiatric domain: a corpus analysis. *Procedia - Social and Behavioral Sciences*, *173*, 285–292. http://doi.org/10.1016/j.sbspro.2015.02.067

Lerner, E. B., Jehle, D. V, Janicke, D. M., & Moscati, R. M. (2000). Medical communication: do our patients understand? *The American Journal of Emergency Medicine*, *18*(7), 764–6. http://doi.org/10.1053/ajem.2000.18040

Levin, M. E. (2006). Different use of medical terminology and culture-specific models of disease affecting communication between Xhosa-speaking patients and English-speaking doctors at a South African paediatric teaching hospital. *South African Medical Journal*, *96*(10), 1080–1084.

Lewis, M. A., & Newton, J. T. (2006). An evaluation of the quality of commercially produced patient information leaflets. *British Dental Journal*, *201*(2), 114–7; discussion 100. http://doi.org/10.1038/sj.bdj.4813846

Leydon, G. M., Boulton, M., Moynihan, C., Jones, A., Mossman, J., Boudioni, M., & McPherson, K. (2000). Cancer patients' information needs and information seeking behaviour: in depth interview study. *BMJ (Clinical Research Ed.)*, *320*(7239), 909–13. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/10742000

Lindström, A., & Weatherall, A. (2015). Orientations to epistemics and deontics in treatment discussions. *Journal of Pragmatics*, *78*, 39–53. http://doi.org/10.1016/j.pragma.2015.01.005

Little, P., Dorward, M., Warner, G., Moore, M., Stephens, K., Senior, J., & Kendrick, T. (2004). Randomised controlled trial of effect of leaflets to empower patients in consultations in primary care. *BMJ (Clinical Research Ed.)*, *328*(7437), 441. http://doi.org/10.1136/bmj.37999.716157.44

Lövestam, E., Fjellström, C., Koochek, A., & Andersson, A. (2014). The power of language on patient-centredness: linguistic devices in the dietetic notes of patient records. *International Journal of Applied Linguistics*, *25*(2), n/a-n/a. http://doi.org/10.1111/ijal.12064

Lucas, P., Lenstrup, M., Prinz, J., Williamson, D., Yip, H., & Tipoe, G. (1997). Language as a barrier to the acquisition of anatomical knowledge. *Medical Education*, *31*(2), 81–86. http://doi.org/10.1111/j.1365-2923.1997.tb02463.x

Lucassen, E. B. (2015). Treatment of Multiple Sclerosis in Switzerland and the United States: What can be Learned from our Differences? *Journal of Multiple Sclerosis*, *02*(04), 1–1. http://doi.org/10.4172/2376-0389.1000e107

Lustria, M. L. A., Cortese, J., Brown, L. L., Davis, R., Mahabi, V., Bae, B. J., & Plotnikova, K. A. (2008). All computer-tailored online health interventions are not created equal. *Proceedings of the American Society for Information Science and Technology*, *44*(1), 1–6. http://doi.org/10.1002/meet.1450440361

Lustria, M. L. A., Noar, S. M., Cortese, J., Van Stee, S. K., Glueckauf, R. L., & Lee, J. (2013). A meta-analysis of web-delivered tailored health behavior change interventions. *Journal of Health Communication*, *18*(9), 1039–1069. http://doi.org/10.1080/10810730.2013.768727

Lysanets, Y., Morokhovets, H., & Bieliaieva, O. (2017). Stylistic features of case reports as a genre of medical discourse. *Journal of Medical Case Reports*, *11*(1), 83. http://doi.org/10.1186/s13256-017-1247-x

Łyszczarz, B. (2017). Gender bias and sex-based differences in health care efficiency in Polish regions. *International Journal for Equity in Health*, *16*(1), 8. http://doi.org/10.1186/s12939-016-0501-y

Maat, H. P., & Klaassen, R. (1994). Side Effects of Side Effect Information in Drug Information Leaflets. *Journal of Technical Writing and Communication*, *24*(4), 389–404. http://doi.org/10.2190/LEK9-VUJP-L3B9-V2LD

Maat, H. P., & Lentz, L. (2011). Using sorting data to evaluate text structure: An evidence-based proposal for restructuring patient information leaflets. *Technical Communication*, *58*(3), 197–216.

Macdonald, M. N. (2002). Pedagogy, pathology and ideology: The production, transmission and reproduction of medical discourse. *Discourse and Society*, *13*(4), 447–467. http://doi.org/10.1177/0957926502013004453

Mack, J. E., Meltzer-Asscher, A., Barbieri, E., & Thompson, C. K. (2013). Neural correlates of processing passive sentences. *Brain Sciences*, *3*(3), 1198–214. http://doi.org/10.3390/brainsci3031198

Maggs, F. M., Jubb, R. W., & Kemm, J. R. (1996). Single-blind randomized controlled trial of an educational booklet for patients with chronic arthritis. *British Journal of Rheumatology*, *35*(8), 775–7. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/8761192

Marco, M. J. L. (2000). Collocational frameworks in medical research papers: a genre-based study. *English for Specific Purposes*, *19*(1), 63–86. http://doi.org/10.1016/S0889-4906(98)00013-1

Marshall, A. L., Leslie, E. R., Bauman, A. E., Marcus, B. H., & Owen, N. (2003). Print versus website physical activity programs: a randomized trial. *American Journal of Preventive Medicine*, *25*(2), 88–94. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/12880874

Martin, J. R. (2014). Evolving systemic functional linguistics: beyond the clause. *Functional Linguistics*, *1*(1), 3. http://doi.org/10.1186/2196-419X-1-3

Martínez, A. C. L. (2002). Empirical examination of EFL readers' use of rhetorical information. English for Specific Purposes, 21(1), 81–98. https://doi.org/10.1016/s0889-4906(00)00029-6

Mathers, S. A., McKenzie, G. A., & Robertson, E. M. (2011). A necessary evil: The experiences of men with prostate cancer undergoing imaging procedures. *Radiography*, *17*(4), 284–291. http://doi.org/10.1016/J.RADI.2011.06.005

Mathers, S. a, Chesson, R. a, & Mckenzie, G. A. (2009). The information needs of people attending for computed tomography (CT): what are they and how can they be met? *Patient Education and Counseling*, *77*(2), 272–8. http://doi.org/10.1016/j.pec.2009.03.030

Matthiessen, C. (201AD). Applying systemic functional linguistics in healthcare contexts. *Text&Talk*, *33(4-5)*(437–467). Retrieved from https://www-degruyter-com.openathens-proxy.swan.ac.uk/downloadpdf/j/text.2013.33.issue-4-5/text-2013-0021/text-2013-0021.pdf

Mayberry Margaret, & Mayberry John. (2002). Consent with understanding: a movement towards informed decisions. *Clinical Medicine Royal College of Physicians Journal*, *2*, 523–6. Retrieved from www.bailii.org

Maynard, D. W., & Heritage, J. (2005). Conversation analysis, doctor-patient interaction and medical communication. *Medical Education*, *39*, 428–435. http://doi.org/10.1111/j.1365-2929.2005.02111.x

Mazzi, D. (2016). "It is natural for you to be afraid…": On the discourse of web-based communication with patients. *Language Learning in Higher Education*, *6*(1). http://doi.org/10.1515/cercles-2016-0011

McCabe, M., Morgan, F., Curley, H., Begay, R., & Gohdes, D. M. (2005). The informed consent process in a cross-cultural setting: is the process achieving the intended result? *Ethnicity & Disease*, *15*(2), 300–4. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/15825977

McCartney, M. (2013). Patient information leaflets: “a stupid system”. *BMJ (Clinical Research Ed.)*, *347*, f4748. http://doi.org/10.1136/bmj.f4748

McEnery, A., Xiao, Z., & Tono, Y. (2006). *Corpus-based Language Studies: An Advanced Resource Book*. *Corpus-based Language Studies: An Advanced Resource Book* ((Routledge, Vol. 86). London: London: Routledge. http://doi.org/10.1016/j.jvir.2007.12.450

McEnery, T., & Hardie, A. (2011). *Corpus linguistics: Method, theory and practice*. *Corpus Linguistics: Method, Theory and Practice*. Cambridge: Cambridge University Press. http://doi.org/10.1017/CBO9780511981395

McEnteggart, G. E., Naeem, M., Skierkowski, D., Baird, G. L., Ahn, S. H., & Soares, G. (2015). Readability of online patient education materials related to IR. *Journal of Vascular and Interventional Radiology*, *26*(8), 1164–1168. http://doi.org/10.1016/j.jvir.2015.03.019

McGrath, P., Henderson, D., & Holewa, H. (2013). Language issues: an important professional practice dimension for Australian International Medical Graduates. *Communication & Medicine*, *10*(3), 191–200. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/25233557

McMullan, M. (2006). Patients using the Internet to obtain health information: How this affects the patient–health professional relationship. *Patient Education and Counseling*, *63*(1–2), 24–28. http://doi.org/10.1016/j.pec.2005.10.006

Meade, C. D., & Smith, C. F. (1991). Readability formulas: Cautions and criteria. *Patient Education and Counseling*, *17*(2), 153–158. http://doi.org/10.1016/0738-3991(91)90017-Y

Meadows, L. M., Mrkonjic, L. A., O’Brien, M. D., & Tink, W. (2007). The importance of communication in secondary fragility fracture treatment and prevention. *Osteoporosis International : A Journal Established as Result of Cooperation between the European Foundation for Osteoporosis and the National Osteoporosis Foundation of the USA*, *18*(2), 159–66. http://doi.org/10.1007/s00198-006-0213-z

Meuter, R. F. I., Gallois, C., Segalowitz, N. S., Ryder, A. G., & Hocking, J. (2015). Overcoming language barriers in healthcare: A protocol for investigating safe and effective communication when patients or clinicians use a second language. *BMC Health Services Research*, *15*(1), 371. http://doi.org/10.1186/s12913-015-1024-8

Michalski, K., Farhan, N., Motschall, E., Vach, W., & Boeker, M. (2017). Dealing with foreign cultural paradigms: A systematic review on intercultural challenges of international medical graduates. *PloS One*, *12*(7), e0181330. http://doi.org/10.1371/journal.pone.0181330

Millar, N., & Budgell, B. S. (2008). The language of public health - A corpus-based analysis. *Journal of Public Health*, *16*(5), 369–374. http://doi.org/10.1007/s10389-008-0178-9

Minton, T. D. (2015). In Defense of the Passive Voice in Medical Writing. *Keio J Med*, *64*(1), 1–10. http://doi.org/10.2302/kjm.2014-0009-RE

Mira, J. J., Orozco-Beltrán, D., Pérez-Jover, V., Martínez-Jimeno, L., Gil-Guillén, V. F., Carratala-Munuera, C., … Asencio-Aznar, A. (2013). Physician patient communication failure facilitates medication errors in older polymedicated patients with multiple comorbidities. *Family Practice*, *30*(1), 56–63. http://doi.org/10.1093/fampra/cms046

Moore, R. J., & Hallenbeck, J. (2010). Narrative Empathy and How Dealing with Stories Helps: Creating a Space for Empathy in Culturally Diverse Care Settings. *Journal of Pain and Symptom Management*, *40*(3), 471–476. http://doi.org/10.1016/j.jpainsymman.2010.03.013

Morony, S., Webster, A. C., Buchbinder, R., Kirkendall, S., McCaffery, K. J., & Clerehan, R. (2018). A Linguistic Analysis of Health Literacy Demands of Chronic Kidney Disease Patient Education Materials. *HLRP: Health Literacy Research and Practice*, *2*(1), e1–e14. http://doi.org/10.3928/24748307-20171227-01

Moss, B., & Roberts, C. (2005). Explanations, explanations, explanations: how do patients with limited English construct narrative accounts in multi-lingual, multi-ethnic settings, and how can GPs interpret them? *Family Practice*, *22*(4), 412–8. http://doi.org/10.1093/fampra/cmi037

Mottram, D. R., & Reed, C. (1997). Comparative evaluation of patient information leaflets by pharmacists, doctors and the general public. *Journal of Clinical Pharmacy and Therapeutics*, *22*(2), 127–134.

Mullany, L., Smith, C., Harvey, K., & Adolphs, S. (2015). Am i anorexic? Weight, eating and discourses of the body in online adolescent health communication. *Communication and Medicine*, *12*(2–3), 211–223. http://doi.org/10.1558/cam.16692

Munsour, E. E., Awaisu, A., Hassali, M. A. A., Darwish, S., & Abdoun, E. (2017). Readability and Comprehensibility of Patient Information Leaflets for Antidiabetic Medications in Qatar. *Journal of Pharmacy Technology*, *33*(4), 128–136. http://doi.org/10.1177/8755122517706978

Murgic, L., Hébert, P. C., Sovic, S., & Pavlekovic, G. (2015). Paternalism and autonomy: views of patients and providers in a transitional (post-communist) country. *BMC Medical Ethics*, *16*(1), 65. http://doi.org/10.1186/s12910-015-0059-z

Murphy, F. (2001). Understanding the humanistic interaction with medical imaging technology. *Radiography*, *7*(3), 193–201. http://doi.org/10.1053/radi.2001.0328

Murray, D. E. (1988). The context of oral and written language: A framework for mode and medium switching. *Language in Society*, *17*(3), 351–373. http://doi.org/10.1017/S004740450001294X

Myhill, J. (1997). Should and ought: the rise of individually oriented modality in American English. *English Language and Linguistics*, *1*(01), 3–23. http://doi.org/10.1017/S1360674300000332

Myhill, J. (1995). Change and continuity in the functions of the American English modals. *Linguistics*, *33*(2), 157–212. http://doi.org/10.1515/ling.1995.33.2.157

Nagel, K., Wizowski, L., Duckworth, J. A., Cassano, J., Hahn, S. A., & Neal, M. (2008). Using plain language skills to create an educational brochure about sperm banking for adolescent and young adult males with cancer. *Journal of Pediatric Oncology Nursing*, *25*(4), 220–226. http://doi.org/10.1177/1043454208319973

Nair, K., Dolovich, L., Cassels, A., McCormack, J., Levine, M., Gray, J., … Burns, S. (2002). What patients want to know about their medications. Focus group study of patient and clinician perspectives. *Canadian Family Physician Medecin de Famille Canadien*, *48*, 104–10. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/11852597

Narumol, J., Arunrot, P., & Krska, J. (2015). Survey of patients' experiences and their certainty of suspected adverse drug reactions. *International Journal of Clinical Pharmacy*, *37*(1), 168–174. http://doi.org/10.1007/s11096-014-0060-5

Nerlich, B., & Halliday, C. (2007). Avian flu: The creation of expectations in the interplay between science and the media. *Sociology of Health and Illness*, *29*(1), 46–65. http://doi.org/10.1111/j.1467-9566.2007.00517.x

Nesselhauf, N. (2004). *Collocations in a Learner Corpus*. *John Benjamins Publishing* (Vol. 14). Amsterdam: John Benjamins Publishing Company. http://doi.org/10.1017/S0272263107270068

NHS (2003) *Toolkit for producing patient information*. Retrieved from https://www.uea.ac.uk/documents/246046/0/Toolkit+for+producing+patient+information.pdf

Nissen, T., & Wynn, R. (2014). The history of the case report: a selective review. *JRSM Open*, *5*(4), 205427041452341. http://doi.org/10.1177/2054270414523410

Nokkonen, S. (2014). The register variation of need to in Spoken British English. *Neuphilologische Mitteilungen*, *115*(1), 63–94. http://doi.org/10.3366/cor.2010.0003

Nutbeam, D. (2000). Health literacy as a public health goal: a challenge for contemporary health education and communication strategies into the 21st century. *Health Promotion International*, *15*(3), 259–267. http://doi.org/10.1093/heapro/15.3.259

Nuyts, J., Byloo, P., & Diepeveen, J. (2010). On deontic modality, directivity, and mood: The case of Dutch mogen and moeten. *Journal of Pragmatics*, *42*(1), 16–34. http://doi.org/10.1016/j.pragma.2009.05.012

Nwogu, K. N. (1997). The medical research paper: Structure and functions. *English for Specific Purposes*, *16*(2), 119–138. http://doi.org/10.1016/S0889-4906(97)85388-4

O’Connell, R. L., Hartridge-Lambert, S. K., Din, N., St John, E. R., Hitchins, C., & Johnson, T. (2013). Patients’ understanding of medical terminology used in the breast clinic. *The Breast*, *22*(5), 836–838. http://doi.org/10.1016/J.BREAST.2013.02.019

Ogden, J., Boden, J., Caird, R., Chor, C., Flynn, M., Hunt, M., Khan, K., M., & K., Swade, S., & Thapar, V. (1999). (1999). You’re Depressed, No I’m not: GPs’ and patients’ different models of depression. *British Journal of General Practice, 49*, 123–124.

O’Grady, C., Dahm, M. R., Roger, P., & Yates, L. (2014). Trust, talk and the dictaphone: Tracing the discursive accomplishment of trust in a surgical consultation. *Discourse and Society*, *25*(1), 65–83. http://doi.org/10.1177/0957926513496354

Ong, L. M., de Haes, J. C., Hoos, a M., & Lammes, F. B. (1995). Doctor-patient communication: a review of the literature. *Social Science & Medicine (1982), 40*(7), 903–18. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/18666041

Ownby, R. L. (2005). Influence of vocabulary and sentence complexity and passive voice on the readability of consumer-oriented mental health information on the Internet. *AMIA ... Annual Symposium Proceedings. AMIA Symposium*, *2005*, 585–9. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/16779107

Pan, F., Reppen, R., & Biber, D. (2016). Comparing patterns of L1 versus L2 English academic professionals: Lexical bundles in Telecommunications research journals. *Journal of English for Academic Purposes*, *21*, 60–71. http://doi.org/10.1016/j.jeap.2015.11.003

Pandiya, A. (2010). Readability and comprehensibility of informed consent forms for clinical trials. *Perspectives in Clinical Research*, *1*(3), 98–100. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/21814628

Papautsky, E. L., & Shalin, V. L. (2017). Strategies of Using Textual Information: Implications for Delivery of Health Information to Patients and Families. In *Proceedings of the 2017 International Symposium on Human Factors and Ergonomics in Health Care* (pp. 79–83). http://doi.org/10.1177/2327857917061018

Patha, P. (2006). This Is Very Important - University of Helsinki Research Portal - University of Helsinki. In F. S.-M. Maurizio Gotti (Ed.), *Advances in Medical Discourse Analysis : Oral and Written Contexts* (pp. 357–381). Peter Lang. Retrieved from https://tuhat.helsinki.fi/portal/en/publications/this-is-very-import(af8a1b4a-4ec7-4609-961e-f4f315a4ff35).html

Paul, F., Jones, M. C., Hendry, C., & Adair, P. M. (2007). The quality of written information for parents regarding the management of a febrile convulsion: a randomized controlled trial. *Journal of Clinical Nursing*, *16*(12), 2308–22. http://doi.org/10.1111/j.1365-2702.2007.02019.x

Peacock, M. (2014). Modals in the construction of research articles: A cross-disciplinary perspective. *Iberica*, (27), 143–164.

Peckham, T. J. (1994). Doctor, have I got a fracture or a break? *Injury*, *25*(4), 221–222. http://doi.org/10.1016/0020-1383(94)90065-5

Pennebaker, J. W., & King, L. A. (1999). Linguistic styles: Language use as an individual difference. *Journal of Personality and Social Psychology*, *77*(6), 1296–1312. http://doi.org/10.1037/0022-3514.77.6.1296

Pérez-sabater, C., Pena-Martinez, G., Turney, E., & Monero-Fleta, B. (2008). A Spoken Genre Gets Written: Online Football Commentaries in English, French, and Spanish. *Written Communication*, *25*(2), 235–261. http://doi.org/10.1177/0741088307313174

Perkins, M. R. (1982). The core meanings of the English modals. *Journal of Linguistics*, *18*(02), 245. http://doi.org/10.1017/S0022226700013608

Phillips, S. P. (2005). International Journal for Equity in Health Defining and measuring gender: A social determinant of health whose time has come. *International Journal for Equity in Health*, *4*(11). http://doi.org/10.1186/1475-9276-4

Pieterse, A. H., Jager, N. A., Smets, E. M. A., & Henselmans, I. (2013). Lay understanding of common medical terminology in oncology. *Psycho-Oncology*, *22*(5), 1186–1191. http://doi.org/10.1002/pon.3096

Pilnick, A., Hindmarsh, J., & Gill, V. T. (2010). *Communication in Healthcare Settings: Policy, Participation and New Technologies*. *Communication in Healthcare Settings: Policy, Participation and New Technologies*. Wiley-Blackwell.

Pires, C. M., & Cavaco, A. M. (2014). Exploring the perspectives of potential consumers and healthcare professionals on the readability of a package insert: A case study of an over-the-counter medicine. *European Journal of Clinical Pharmacology*, *70*(5). http://doi.org/10.1007/s00228-014-1645-6

Pires, C., Vigário, M., & Cavaco, A. (2015). Readability of medicinal package leaflets: A systematic review. *Revista de Saude Publica*. http://doi.org/10.1590/S0034-8910.2015049005559

Prabhu, A. V., Hansberry, D. R., Agarwal, N., Clump, D. A., & Heron, D. E. (2016). Radiation Oncology and Online Patient Education Materials: Deviating From NIH and AMA Recommendations. *International Journal of Radiation Oncology Biology Physics*, *96*(3), 521–528. http://doi.org/10.1016/j.ijrobp.2016.06.2449

Protheroe, J., Estacio, E. V., & Saidy-Khan, S. (2015). Patient information materials in general practices and promotion of health literacy: An observational study of their effectiveness. British Journal of General Practice, 65(632), e192–e197. https://doi.org/10.3399/bjgp15X684013

Raynor, D. K., Savage, I., Knapp, P., & Henley, J. (2004). We are the experts: People with asthma talk about their medicine information needs. *Patient Education and Counseling*. http://doi.org/10.1016/S0738-3991(03)00126-5

Rayson, P. (2015). Computational tools and methods for corpus compilation and analysis. In *The Cambridge Handbook of English Corpus Linguistics* (pp. 32–49). http://doi.org/10.1007/9781139764377.003

Regitz-Zagrosek, V. (2012). Sex and gender differences in health. Science & Society Series on Sex and Science. *EMBO Reports*, *13*(7), 596–603. http://doi.org/10.1038/embor.2012.87

Research, G. V. (2018). What are the factors influencing ehealth market growth? Retrieved October 19, 2018, from www.Healthworkscollective.com

Rice, R. E. (2006, January). Influences, usage, and outcomes of Internet health information searching: Multivariate results from the Pew surveys. *International Journal of Medical Informatics*. http://doi.org/10.1016/j.ijmedinf.2005.07.032

Richard Watson Todd. (2017). An Opaque Engineering Word List. *English for Specific Purposes*, *45*. Retrieved from https://ac.els-cdn.com/S0889490616300862/1-s2.0-S0889490616300862-main.pdf?_tid=dcd9ef3d-7547-4960-a39a-f970e3dcd976&acdnat=1533633466_28c34cf5fe4608f02acca6ad9bd00933

Roberts, C., Moss, B., Wass, V., Sarangi, S., & Jones, R. (2005). Misunderstandings: a qualitative study of primary care consultations in multilingual settings, and educational implications. *Medical Education*, *39*(5), 465–75. http://doi.org/10.1111/j.1365-2929.2005.02121.x

Roberts, C., & Sarangi, S. (2003). Uptake of Discourse Research in Interprofessional Settings: Reporting from Medical Consultancy. *Applied Linguistics*, *24*(3), 338–359+421. http://doi.org/10.1093/applin/24.3.338

Roeland, E., Cain, J., Onderdonk, C., Kerr, K., Mitchell, W., & Thornberry, K. (2014). When Open-Ended Questions Don’t Work: The Role of Palliative Paternalism in Difficult Medical Decisions. *Journal of Palliative Medicine*, *17*(4), 415–420. http://doi.org/10.1089/jpm.2013.0408

Rooks, R. N., Wiltshire, J. C., Elder, K., BeLue, R., & Gary, L. C. (2012). Health information seeking and use outside of the medical encounter: Is it associated with race and ethnicity? *Social Science and Medicine*, *74*(2), 176–184. http://doi.org/10.1016/j.socscimed.2011.09.040

Rosenkrantz, A. B., & Flagg, E. R. (2015). Survey-Based Assessment of Patients’ Understanding of Their Own Imaging Examinations. *Journal of the American College of Radiology*, *12*(6), 549–555. http://doi.org/10.1016/J.JACR.2015.02.006

Røst, T. B., Huseth, O., & Grimsmo, A. (2008). Lessons from Developing an Annotated Corpus of Patient Histories. *Journal of Computer Science and Engineering*, *2*(2), 162–179.

Royal College of Emergency Medicine. (2016). *The Royal College of Emergency Medicine Best Practice Guideline Management of Radiology Results in the Emergency Department*. Retrieved from https://www.rcem.ac.uk/docs/College Guidelines/5q. Management of Radiology Results in the Emergency Department (February 2016).pdf

Rozmovits, L., & Ziebland, S. (2004). What do patients with prostate or breast cancer want from an Internet site? A qualitative study of information needs. *Patient Education and Counseling*, *53*(1), 57–64. http://doi.org/10.1016/S0738-3991(03)00116-2

Rutten, L. J. F., Squiers, L., & Hesse, B. (2006). Cancer-related information seeking: Hints from the 2003 Health Information National Trends Survey (HINTS). *Journal of Health Communication*, *11*(SUPPL. 1), 147–156. http://doi.org/10.1080/10810730600637574

Ryhänen, A. M., Johansson, K., Virtanen, H., Salo, S., Salanterä, S., & Leino-Kilpi, H. (2009). Evaluation of written patient educational materials in the field of diagnostic imaging. *Radiography*, *15*(2), e1–e5. http://doi.org/10.1016/j.radi.2008.04.006

Sakai, H., Katsumata, N., & Takahashi, M. (2017). Providing written information increases patient satisfaction: a web-based questionnaire survey of Japanese cancer survivors. *Japanese Journal of Clinical Oncology*, *47*(7), 611–617. http://doi.org/10.1093/jjco/hyx043

Salager-meyer, F. (1983). The Lexis of Fundamental English:Classificatory Framework and Rhetorical Function (A Statistical Approach). *Reading in a Foreign Language*.

Salager-Meyer, F. (1990). Discoursal fiaws in Medical English abstracts: A genre analysis per research- and text-type*. *Text*, *10*(4), 365–384.

Salager-Meyer, F. (1994). Hedges and textual communicative function in medical English written discourse. *English for Specific Purposes*, *13*(2), 149–170. http://doi.org/10.1016/0889-4906(94)90013-2

Salager-Meyer, F. (1992). A text-type and move analysis study of verb tense and modality distribution in medical English abstracts. *English for Specific Purposes*, *11*(2), 93–113. http://doi.org/10.1016/S0889-4906(05)80002-X

Scantlebury, A., Booth, A., & Hanley, B. (2017). Experiences, practices and barriers to accessing health information: A qualitative study. *International Journal of Medical Informatics*, *103*, 103–108. http://doi.org/10.1016/j.ijmedinf.2017.04.018

Schiffrin, D., Tannen, D., & Hamilton, H. E. (2008). *The {Handbook} of {Discourse} {Analysis}*.

Schulz, P. J., & Nakamoto, K. (2013). Health literacy and patient empowerment in health communication: The importance of separating conjoined twins. *Patient Education and Counseling*. http://doi.org/10.1016/j.pec.2012.09.006

Scott, M. (1997). PC ANALYSIS OF KEY WORDS -AND KEY KEY WORDS. *System*, *25*(2), 233–245. http://doi.org/10.1016/S0346-251X(97)00011-0

Scott, M., & Tribble, C. (2006). *Key words of individual texts Aboutness and style*. John Benjamin.

Seale, C., Charteris-Black#, J., & Charteris-Black, J. (2008). The interaction of age and gender in illness narratives. *Ageing and Society*, *28*(7), 1025–1045. http://doi.org/10.1017/S0144686X0800737X

Seale, C., Ziebland, S., & Charteris-Black, J. (2006). Gender, cancer experience and internet use: A comparative keyword analysis of interviews and online cancer support groups. *Social Science & Medicine*, *62*(10), 2577–2590. http://doi.org/10.1016/j.socscimed.2005.11.016

Segal, J. Z. (2009). Internet Health and the 21st-Century Patient: A Rhetorical View. *Written Communication*, *26*(4), 351–369. http://doi.org/10.1177/0741088309342362

Semino, E., Demjén, Z., Hardie, A., Payne, S., & Rayson, P. (2017). *Metaphor, Cancer and the End of Life: A Corpus-Based Study*. *Metaphor, Cancer and the*

End of Life: A Corpus-Based Study. Routledge. http://doi.org/10.4324/9781315629834

Semino, E., & Demjén, Z. (2017). The Cancer Card: Metaphor, Intimacy, and Humor in Online Interactions about the Experience of Cancer. In B. Hampe (Ed.), *Metaphor* (pp. 181–199). Cambridge: Cambridge University Press. http://doi.org/10.1017/9781108182324.011

Shaw, J., & Baker, M. (2004). “Expert patient”—dream or nightmare? *Bmj*, *328*(7442), 723–724. http://doi.org/10.1136/bmj.328.7442.723

Shea-Budgell, M. A., Kostaras, X., Myhill, K. P., & Hagen, N. A. (2014). Information needs and sources of information for patients during cancer follow-up. *Current Oncology*, *21*(4), 165. http://doi.org/10.3747/co.21.1932

Sheard, C., & Garrud, P. (2006). Evaluation of generic patient information: effects on health outcomes, knowledge and satisfaction. *Patient Education and Counseling*, *61*(1), 43–7. http://doi.org/10.1016/j.pec.2005.02.004

Shergill, I., Bahl, K., Farjad, M., Phipps, C., & Fowlis, G. (2010). Patient information leaflets for Transrectal Ultrasound guided prostate biopsy: Results of North Thames deanery survey. *BMC Research Notes*, *3*(1), 27. http://doi.org/10.1186/1756-0500-3-27

Shim, H., Ailshire, J., Zelinski, E., & Crimmins, E. (2018). The Health and Retirement Study: Analysis of Associations Between Use of the Internet for Health Information and Use of Health Services at Multiple Time Points. *Journal of Medical Internet Research*, *20*(5), e200. http://doi.org/10.2196/jmir.8203

Short, C. E., James, E. L., Plotnikoff, R. C., & Girgis, A. (2011). Efficacy of tailored-print interventions to promote physical activity: a systematic review of randomised trials. *International Journal of Behavioral Nutrition and Physical Activity*, *8*(1), 113. http://doi.org/10.1186/1479-5868-8-113

Silver, M. P. (2015). Patient Perspectives on Online Health Information and Communication With Doctors: A Qualitative Study of Patients 50 Years Old and Over. *Journal of Medical Internet Research*, *17*(1), e19. http://doi.org/10.2196/jmir.3588

Sinclair, J., & Carter, R. (2004). *Trust the text : language, corpus and discourse*. Routledge. Retrieved from https://www.routledge.com/Trust-the-Text-Language-Corpus-and-Discourse/Sinclair-Carter/p/book/9780203594070

Singh, N., Mohacsy, A., Connell, D. A., & Schneider, M. E. (2017). A snapshot of patients’ awareness of radiation dose and risks associated with medical imaging examinations at an Australian radiology clinic. *Radiography*, *23*(2), 94–102. http://doi.org/10.1016/J.RADI.2016.10.011

Sinha, A., Porter, T., & Wilson, A. (2018). The Use of Online Health Forums by Patients With Chronic Cough: Qualitative Study. *Journal of Medical Internet Research*, *20*(1), e19. http://doi.org/10.2196/jmir.7975

Skelton, J. R., & Hobbs, F. D. (1999). Descriptive study of cooperative language in primary care consultations by male and female doctors. *BMJ (Clinical Research Ed.)*, *318*(7183), 576–579. http://doi.org/10.1136/bmj.318.7183.576

Skelton, J. R., Kai, J., & Loudon, R. F. (2008). Cross-cultural communication in medicine: questions for educators. *Medical Education*, *35*(3), 257–261. http://doi.org/10.1111/j.1365-2923.2001.00873.x

Skelton, J. (1997). The representation of truth in academic medical writing. *Applied Linguistics*, *18*(2), 121–140. http://doi.org/10.1093/applin/18.2.121

Skelton, J. R., Wearn, A. M., & Hobbs, F. D. R. (2002). A concordance-based study of metaphoric expressions used by general practitioners and patients in consultation. *The British Journal of General Practice : The Journal of the Royal College of General Practitioners*, *52*(475), 114–8. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/11887876

Skelton, J. R., Wearn, A. M., & Hobbs, F. D. R. (2002). “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. http://doi.org/10.1093/fampra/19.5.484

Skelton, J. R., & Whetstone, J. (2012). English for Medical Purposes and Academic Medicine: looking for common ground. *Ibérica*, *24*(24), 87–102. Retrieved from http://www.redalyc.org/articulo.oa?id=287024476008

Skelton, J. (2008). *Language and clinical communication : this bright Babylon*. Radcliffe Pub.

Skelton, J. R., & Hobbs, F. D. R. (1999). Concordancing: Use of language-based research in medical communication. *Lancet*, *353*(9147), 108–111. http://doi.org/10.1016/S0140-6736(98)02469-6

Skjeggestad, E., Gerwing, J., & Gulbrandsen, P. (2017). Language barriers and professional identity: A qualitative interview study of newly employed international medical doctors and Norwegian colleagues. *Patient Education and Counseling*, *100*(8), 1466–1472. http://doi.org/10.1016/j.pec.2017.03.007

Skjeggestad, E., Norvoll, R., Sandal, G. M., & Gulbrandsen, P. (2017). How do international medical graduates and colleagues perceive and deal with difficulties in everyday collaboration? A qualitative study. *Scandinavian Journal of Public Health*, *45*(4). http://doi.org/10.1177/1403494817698286

Slade, D., Scheeres, H., Manidis, M., Iedema, R., Dunston, R., Stein-Parbury, J., … McGregor, J. (2008). Emergency communication: the discursive challenges facing emergency clinicians and patients in hospital emergency departments. *Discourse & Communication*, *2*(3), 271–298. http://doi.org/10.1177/1750481308091910

Sloane, C., Whitley, S. A., Anderson, C., & Holmes, K. (2010). *Clark's Pocket Handbook for Radiographers*. CRC Press. Retrieved from https://books.google.com/books?id=6i7SBQAAQBAJ&pgis=1

Smith, S., & Duman, M. (2009, December). The state of consumer health information: An overview. *Health Information and Libraries Journal*. Wiley/Blackwell (10.1111). http://doi.org/10.1111/j.1471-1842.2009.00870.x

Squiers, L., Finney Rutten, L. J., Treiman, K., Bright, M. A., & Hesse, B. (2005). Cancer patients' information needs across the cancer care continuum: Evidence from the Cancer Information Service. *Journal of Health Communication*. http://doi.org/10.1080/10810730500263620

Stevenson, F., McNulty, G., & Leontowitsch, M. (2012). Negotiating roles in pharmacy practice: Interactions across linguistic and cultural barriers. *Communication and Medicine*, *9*(1), 83–94. http://doi.org/10.1558/cam.v9i1.83

Sudore, R. L., & Schillinger, D. (2009). Interventions to Improve Care for Patients with Limited Health Literacy. *Journal of Clinical Outcomes Management : JCOM*, *16*(1), 20–29. Retrieved from http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2799039&tool=pmcentrez&rendertype=abstract

Sustersic, M., Jeannet, E., Cozon-Rein, L., Maréchaux, F., Genty, C., Foote, A., … Bosson, J. L. (2013). Impact of information leaflets on behavior of patients with gastroenteritis or tonsillitis: A cluster randomized trial in french primary care. *Journal of General Internal Medicine*, *28*(1), 25–31. http://doi.org/10.1007/s11606-012-2164-8

Swee Heng, C., Kashiha, H., & Tan, H. (2014). Lexical Bundles: Facilitating University "Talk" in Group Discussions. *English Language Teaching*, *7*(4). http://doi.org/10.5539/elt.v7n4p1

Taavitsainen, I., & Pahta, P. (2000). *Conventions of Professional Writing The Medical Case Report in a Historical Perspective*. *JEngL* (Vol. 28). Retrieved from http://journals.sagepub.com.openathens-proxy.swan.ac.uk/doi/pdf/10.1177/00754240022004875

Tailor, A., & Ogden, J. (2009). Avoiding the term 'obesity': An experimental study of the impact of doctors' language on patients' beliefs. *Patient Education and Counseling*, *76*(2), 260–264. http://doi.org/10.1016/j.pec.2008.12.016

Tausczik, Y. R., & Pennebaker, J. W. (2010). The psychological meaning of words: LIWC and computerized text analysis methods. *Journal of Language and Social Psychology*. http://doi.org/10.1177/0261927X09351676

The Medicines and Healthcare products Regulatory Agency. (2012). *BEST PRACTICE GUIDANCE ON PATIENT INFORMATION LEAFLETS EXPLANATORY MEMORANDUM*. Retrieved from http://www.ema.europa.eu/docs/en_GB/document_library/Template_or_form/200

Tian, C., Champlin, S., Mackert, M., Lazard, A., & Agrawal, D. (2014). Readability, suitability, and health content assessment of web-based patient education materials on colorectal cancer screening. *Gastrointestinal Endoscopy*, *80*(2). http://doi.org/10.1016/j.gie.2014.01.034

Tognini-Bonelli, E. (2001). *Corpus Linguistics at Work*. *Studies in corpus linguistics*. Amsterdam: John Benjamins. http://doi.org/10.1017/CBO9781107415324.004

Tong, V., Raynor, D. K., & Aslani, P. (2017). User testing as a method for identifying how consumers say they would act on information related to over-the-counter medicines. *Research in Social and Administrative Pharmacy*, *13*(3), 476–484. http://doi.org/10.1016/j.sapharm.2016.06.001

Tong, V., Raynor, D. K., & Aslani, P. (2018). Receipt and use of spoken and written over-the-counter medicine information: insights into Australian and UK consumers' experiences. *International Journal of Pharmacy Practice*, *26*(2), 129–137. http://doi.org/10.1111/ijpp.12382

Tutty, L., & O'Connor, G. (1999). Patient information leaflets: Some pertinent guidelines. *Radiography*, *5*(1), 11–14. http://doi.org/10.1016/S1078-8174(99)90003-2

Ukkola, L., Oikarinen, H., Henner, A., Haapea, M., & Tervonen, O. (2017). Patient information regarding medical radiation exposure is inadequate: Patients' experience in a university hospital. *Radiography*, *23*(4), e114–e119. Retrieved from https://www.sciencedirect.com/science/article/pii/S1078817417300524

Van Berkel, J., & Gerritsen, M. (2012). Patient Information Leaflets in Flanders and the Netherlands: Unnecessary Differences? In P. Heynderickx, S. Dieltjens, G. Jacobs, P.Gillaerts, & E. De Groot (Eds.), *The language factor in international business: New perspectives on research, teaching and practice* (pp. 151–171). Bern: Peter Lang. Retrieved from https://marinelgerritsen.files.wordpress.com/2017/11/van-berkel-gerritsen-2012.pdf

Van Rosse, F., De Bruijne, M., Suurmond, J., Essink-Bot, M.-L., & Wagner, C. (2016). Language barriers and patient safety risks in hospital care. A mixed methods study. *International Journal of Nursing Studies*, *54*, 45–53. http://doi.org/10.1016/j.ijnurstu.2015.03.012

Vassallo, T., Mizzi, A., Depasquale, R., Maher, M., & Rainford, L. (2018, October). The development of patient information leaflets incorporating patient diversity considerations: Varicocele embolisation and fluoroscopy guided joint injection examinations. *Radiography*, pp. S20–S27. http://doi.org/10.1016/j.radi.2018.06.006

Vaughan, E., & O'Keeffe, A. (2015). Corpus Analysis. In *The International Encyclopedia of Language and Social Interaction* (pp. 1–17). Hoboken, NJ, USA: John Wiley & Sons, Inc. http://doi.org/10.1002/9781118611463.wbielsi168

Verma, A., Griffin, A., Dacre, J., & Elder, A. (2016). Exploring cultural and linguistic influences on clinical communication skills: a qualitative study of International Medical Graduates. *BMC Medical Education*, *16*(1), 162. http://doi.org/10.1186/s12909-016-0680-7

Vihla, M. (1999). *Medical Writing: Modality in Focus Volume 28 of Language and Computers: Studies in Practical Linguistics* (28th ed.). Rodopi.

Villafranca, A., Kereliuk, S., Hamlin, C., Johnson, A., & Jacobsohn, E. (2017). The Appropriateness of Language Found in Research Consent Form Templates: A Computational Linguistic Analysis. *PloS One*, *12*(2), e0169143. http://doi.org/10.1371/journal.pone.0169143

Visschers, V. H. M., Meertens, R. M., Passchier, W. W. F., & de Vries, N. N. K. (2009). Probability Information in Risk Communication: A Review of the Research Literature. *Risk Analysis*, *29*(2), 267–287. http://doi.org/10.1111/j.1539-6924.2008.01137.x

Vom, J., & Williams, I. (2017). Justification of radiographic examinations: What are the key issues? *Journal of Medical Radiation Sciences*, *64*(3), 212–219. http://doi.org/10.1002/jmrs.211

Vrana, S. R., Vrana, D. T., Penner, L. A., Eggly, S., Slatcher, R. B., & Hagiwara, N. (2018). Latent Semantic Analysis: A new measure of patient-physician communication. *Social Science & Medicine*, *198*, 22–26. http://doi.org/10.1016/J.SOCSCIMED.2017.12.021

Waisman, Y., Siegal, N., Siegal, G., Amir, L., Cohen, H., & Mimouni, M. (2005). Role of diagnosis-specific information sheets in parents' understanding of emergency department discharge instructions. *European Journal of Emergency Medicine : Official Journal of the European Society for Emergency Medicine*, *12*(4), 159–62. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/16034260

Wang, L.-W. W., Miller, M. J., Schmitt, M. R., & Wen, F. K. (2013). Assessing readability formula differences with written health information materials: Application, results, and recommendations. *Research in Social and Administrative Pharmacy*, *9*(5), 503–516. http://doi.org/10.1016/j.sapharm.2012.05.009

Ward, J. (2009). A basic engineering English word list for less proficient foundation engineering undergraduates. *English for Specific Purposes*, *28*(3), 170–182. http://doi.org/10.1016/J.ESP.2009.04.001

Waring, M. E., McManus, D. D., Amante, D. J., Darling, C. E., & Kiefe, C. I. (2018, June 30). Online health information seeking by adults hospitalized for acute coronary syndromes: Who looks for information, and who discusses it with healthcare providers? *Patient Education and Counseling*. Elsevier. http://doi.org/10.1016/j.pec.2018.06.016

Watson Todd, R. (2017). An opaque engineering word list: Which words should a teacher focus on? *English for Specific Purposes*, *45*, 31–39. http://doi.org/10.1016/j.esp.2016.08.003

Webber, P. (2005). Interactive features in medical conference monologue. *English for Specific Purposes*, *24*(2), 157–181. http://doi.org/10.1016/J.ESP.2004.02.003

Weidner, M. (2015). Telling somebody what to tell: “Prosze mi powiedzieć” in Polish doctor-patient interaction. *Journal of Pragmatics*, *78*, 70–83. http://doi.org/10.1016/j.pragma.2015.01.006

Weingarten, S. R., Henning, J. M., Badamgarav, E., Knight, K., Hasselblad, V., Gano, A., & Ofman, J. J. (2002). Interventions used in disease management programmes for patients with chronic illnesswhich ones work? Meta-analysis of published reports. *BMJ*, *325*(7370), 925. http://doi.org/10.1136/bmj.325.7370.925

Westerman, S., & Wenger, N. K. (2016). Women and heart disease, the underrecognized burden: sex differences, biases, and unmet clinical and research challenges. *Clinical Science*, *130*(8), 551–563. http://doi.org/10.1042/CS20150586

White, P., Smith, H., Webley, F., & Frew, A. (2004). A survey of the quality of information leaflets on hayfever available from general practices and community pharmacies. *Clinical and Experimental Allergy : Journal of the British Society for Allergy and Clinical Immunology*, *34*(9), 1438–43. http://doi.org/10.1111/j.1365-2222.2004.02055.x

Williams, I. A. (1999). *Results Sections of Medical Research Articles: Analysis of Rhetorical Categories for Pedagogical Purposes*. *English for Specific Purposes* (Vol. 18). Retrieved from https://ac.els-cdn.com/S0889490698000039/1-s2.0-S0889490698000039-main.pdf?_tid=a4aedacb-b9b1-42ef-aa79-00ce63eaa899&acdnat=1540129496_a572425de1be180ae7ff22eeebbbab5f

Williams, N., & Ogden, J. (2004). The impact of matching the patient’s vocabulary: a randomized control trial. *Family Practice*, *21*(6), 630–635. http://doi.org/10.1093/fampra/cmh610

Williams, S. B., Huo, J., Dafashy, T. J., Ghaffary, C. K., Baillargeon, J. G., Morales, E. E., … Kamat, A. M. (2017). Survival differences among patients with bladder cancer according to sex: Critical evaluation of radical cystectomy use and delay to treatment. *Urologic Oncology: Seminars and Original Investigations*, *35*(10), 602.e1-602.e9. http://doi.org/10.1016/j.urolonc.2017.05.022

Winterbottom, A., Conner, M., Mooney, A., & Bekker, H. L. (2007). Evaluating the quality of patient leaflets about renal replacement therapy across UK renal units. *Nephrology, Dialysis, Transplantation : Official Publication of the European Dialysis and Transplant Association - European Renal Association*, *22*(8), 2291–6. http://doi.org/10.1093/ndt/gfm095

Wong, S. S. M., Bekker, H. L., Thornton, J. G., & Gbolade, B. A. (2003). Choices about abortion method: assessing the quality of patient information leaflets in England and Wales. *BJOG : An International Journal of Obstetrics and Gynaecology*, *110*(3), 263–6. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/12628264

Wright, P. (1999). Writing and information design of healthcare materials. In Christopher N Candlin and Ken Hyland (Ed.), *Writing: Texts, Processes and Practices - Google Books* (p. 330). Retrieved from https://books.google.ch/books?id=NWTJAwAAQBAJ&pg=PA85&lpg=PA85&dq=patricia+wright,+writing+and+information+design+of+healthcare+materials&source=bl&ots=9p3S-bVEfh&sig=ZyeeNll0bGOZAuINCrg5mjUSovs&hl=en&sa=X&ved=2ahUKEwi48biMgMXdAhWQgVwKHVPRCcAQ6AEwAnoECAgQ

Yagoda, B. (2006). You need to read this: How need to vanquished have to, must, and should. *Slate Magazine*. Retrieved from http://www.slate.com/articles/life/the_good_word/2006/07/you_need_to_read_this.html

Yang, A., Zheng, S., & Ge, G. (2015). Epistemic modality in English-medium medical research articles: A systemic functional perspective. *English for Specific Purposes*, *38*, 1–10. http://doi.org/10.1016/J.ESP.2014.10.005

Yang, M.-N. (2015). A nursing academic word list. *English for Specific Purposes*, *37*, 27–38. http://doi.org/10.1016/j.esp.2014.05.003

Yates, L., Dahm, M. R., Roger, P., & Cartmill, J. (2016). Developing rapport in inter-professional communication: Insights for international medical graduates. *English for Specific Purposes*, *42*, 104–116. http://doi.org/10.1016/j.esp.2016.01.006

Zarcadoolas, C. (2011). The simplicity complex: Exploring simplified health messages in a complex world. *Health Promotion International*, *26*(3), 338–350. http://doi.org/10.1093/heapro/daq075

Zarcadoolas, C., Pleasant, A., & Greer, D. S. (2005). Understanding health literacy: an expanded model. *Health Promotion International*, *20*(2), 195–203. http://doi.org/10.1093/heapro/dah609

ZARCADOOLAS, C., PLEASANT, A., & Greer, D. S. (2003). Elaborating a definition of health literacy: A commentary. *Journal of Health Communication*, *8*(May), 119–120. http://doi.org/10.1080/713851982

Zarcadoolas, C., Sealy, Y., Levy, J., Dresser, M., Ponieman, D., Young, S. M., … Silver, L. (2011). Health literacy at work to address overweight and obesity in adults: The development of the obesity action kit. *Journal of Communication in Healthcare*, *4*(2), 88–101. http://doi.org/10.1179/175380611X13082296179316

ZHAO, J., & WU, T. (2013). A genre analysis of medical abstracts by Chinese and English native speakers. *Journal of Medical Colleges of PLA*, *28*(1), 60–64. http://doi.org/10.1016/S1000-1948(13)60018-0

Zheng, J., & Yu, H. (2017). Readability Formulas and User Perceptions of Electronic Health Records Difficulty: A Corpus Study. *Journal of Medical Internet Research*, *19*(3), e59. http://doi.org/10.2196/jmir.6962

Zummo Marianna Lye. (2016). Exploring web-mediated communication: A genre-based linguistic study for new patterns of doctor-patient interaction in online environment. *Communication and Medicine*, *12*(2/3). http://doi.org/10.1558/cam.31897