# DISCUSSING SCIENCE IN THE PUBLIC SPHERE: A CORPUS-ASSISTED STUDY OF WEB-BASED INTERACTION CONCERNING THE MEASLES, MUMPS AND RUBELLA (MMR) TRIPLE VACCINE

by

Deborah Orpin

A thesis submitted to the University of Birmingham for the degree of DOCTOR OF PHILOSOPHY

Department of English

School of English, Drama, and American and Canadian Studies

University of Birmingham

December 2014

## Abstract

This thesis reports a study into aspects of the discourse concerning the measles, mumps and rubella (MMR) triple vaccine. The aim of the thesis is to contribute to knowledge about the ways in which debates about science are enacted in the public sphere. The study uses a corpus-assisted discourse studies (CADS) approach to examine key lexico-grammatical patterns in the *JABS* corpus, a corpus of texts gathered from the website of the vaccine-critical Justice Awareness and Basic Support (JABS) group. The aim of the study is to discover how participants on the JABS website discussion forum draw on discursive resources to achieve their rhetorical goals. Comparison is made with the typical lexico-grammatical patterns in the *NHSvax* corpus, a corpus comprising texts from NHS immunization websites. The study finds that, although there are several areas of similarity between the two corpora, the *JABS* corpus data contains greater evidence of evaluative lexis, a higher frequency of nouns which express evaluations of the status of discursive objects. These resources are used to reformulate and reframe propositions which originate in the medical-scientific domain. Narratives of vaccine damage are also frequently used to express warrants for expertise.

# Dedication

This is dedicated to Caroline and Maya.

# Acknowledgements

Many people have lent me their help and support along the way. I wish to thank my mother, Pearl Orpin, and my sister, Liz Kidd, for tending to my every need last summer so that I could get on with the job of writing this thesis. Maya deserves a vote of thanks for her patience with me over the past few years.

Particular thanks go to Andy Dickinson, for writing the JABS Cleanup program for me, and to Gill Francis for her insightful comments and suuport in the earlier stages of this project. I also wish to thank friends and colleagues from Birmingham and Wolverhampton Universities for their help and support, especially Luba Atherton, Mike Cunningham, Tom Dickins, and Emma Franklin.

I would also like to thank Nick Groom, my Director of Studies, for his assistance and his calming manner. Above all, I wish to thank my supervisor, Alison Sealey, for her support and inspiration over the years, and for her unstinting generosity.

# CONTENTS

## CHAPTER SEVEN: INTERACTIVITY AND INTERTEXTUALITY IN THE JABS CORPUS

# LIST OF FIGURES AND TABLES

**CHAPTER** **PAGE**

## CHAPTER THREE



## CHAPTER FOUR

## CHAPTER FIVE

## CHAPTER SIX

## CHAPTER SEVEN

## LIST OF APPENDICES PAGE

# CHAPTER ONE

# INTRODUCTION

## 1.1 General aim of the thesis

This thesis aims to contribute to enhancing understanding of the ways in which debates about health and science are enacted in the public sphere. Specifically, it aims to discover the ways in which discursive resources are drawn on to make claims and to express warrants for expertise in a corpus of texts taken from a vaccine-critical website.[1] The focus of the study is the discourse around vaccine safety, in particular the safety and risks associated with the measles, mumps and rubella (MMR) combined vaccine. The thesis uses the MMR debate as a case study to illustrate key aspects of public discourse about scientific issues. It examines language use in the *JABS* corpus, a corpus of texts gathered from the JABS website[2] between 2005 and 2008 comparing it, where appropriate, with language use in the *NHSvax* corpus, a corpus comprising data from the Department of Health (DoH) websites, 'MMR The facts', and 'NHS immunisation', and from the DoH-sponsored Health Protection Agency (HPA) immunisation pages.[3] The *NHSvax* texts were gathered between 2005 and 2008. The thesis employs a corpus assisted discourse studies (CADS) approach: a textlinguistic-oriented

[1] The term *vaccine-critical* is used in a lot of the social science literature on public attitudes to vaccination to refer to interest groups which resist public policy on vaccination. Use of the term *vaccine-critical* is considered preferable to *anti-vaccine* because it more accurately reflects the full range of stances towards vaccination (Hobson-West, 2007).
[2] http://www.jabs.org.uk.
[3] http://www.mmrthefacts.nhs.uk; http://www.immunisation.nhs.uk; http://www.hpa.org.uk. Set up in 2004, 'MMR The facts' ceased to be updated in 2008. From 2008, users were directed to the 'NHS immunisation' site. The HPA site closed down in 2013, when the HPA became part of Public Health England.

approach to discourse analysis, which uses corpus linguistics methods so that reliable generalizations can be made on the basis of observations of a large amount of data.[4]

## 1.2 Definitions and explanations of key terms

As described in the section above, the focus of this project is discourse and the analytical approach taken is a form of discourse analysis. It is opportune at this juncture to define *discourse*, since it is defined in various ways in the literature. It is also useful to look at the ways in which discourse analysis is understood, since it is used in a variety of disciplines and executed differently in each. A full discussion of the different meanings of *discourse* (and *discourses*) and of the various approaches to discourse analysis is presented in Chapter Three. For current purposes, I provide below definitions of the terms *discourse* and *discourses*, as they are used in this study. I also define other key terms used in this study which are relevant to the study of discourse and explain how they relate to one another. The terms are *text, ideology, intertextuality* and *interdiscursivity*. I also define *narrative*. I finally give a brief explanation of the differences between what I see as two main approaches to discourse analysis.

*Discourse* is used as an uncountable and as a countable noun. When used as an uncountable noun in this project, it is understood in two main ways:

- as 'language in use' or as 'the pragmatic process of meaning negotiation' (Widdowson, 2004: 8);

[4] The term *Corpus Assisted Discourse Studies* (*CADS*) was coined by Partington (2004a) to refer to studies which combine the quantitative methods of corpus linguistics with the qualitative approach of discourse analysis. He sometimes uses it in a broad sense to refer to any studies which adopt this method of analysis (e.g. Hardt-Mautner, 1995; Stubbs and Gerbig, 1993; Krishnamurthy, 1996; Baker and McEnery, 2005), but more often, he uses it to refer to the specific version of corpus informed discourse analysis developed by researchers at the Universities of Siena and Bologna. I use it in its broader sense.

- as language use associated with particular contexts, for example, the language used in the media (*media discourse*) or language use typical of political contexts (*political discourse*) (Baker and Ellece, 2011: 31).

When defined the first way, *discourse* is often contrasted with *text*. If *discourse* is the process of interaction and meaning negotiation, *text* is its product (Widdowson, 2004: 8). The explanation below elaborates on this distinction:

> The discourse analyst treats his data as the record (text) of a dynamic process in which language was used as an instrument of communication in a context by a speaker/writer to express meanings and achieve intentions (discourse).
>
> (Brown and Yule, 1983: 26)

As a countable noun, *discourse* (and *discourses*) is used in this project as it is generally understood in Critical Linguistics and Critical Discourse Analysis, that is, in a broadly Foucauldian way. The term is therefore used to refer to specific ways of representing the world, in which particular representations of reality are expressed through the use of particular words, phrases, metaphors, and so on. Used in this sense, *discourse(s)* is related to the notion of ideology, a basic definition of which is given by Baker and Ellece (2011):

> Ideology can generally be thought of as the set of ideas, beliefs and aims that a person or group holds.
>
> (Baker and Ellece, 2011: 57)

The connection between discourse and ideology is explained by Fairclough (1992) in the following way:

> I shall understand ideologies to be significations/constructions of reality, which are built into various dimensions of the forms/meanings of discursive practices.
>
> (Fairclough, 1992: 87)

The notion of *intertextuality* and the related notion of *interdiscursivity* are central concepts in critical approaches to discourse analysis. These terms are used in this project as they are used by Fairclough (1992). *Intertextuality* is defined 'the property texts have of being full of snatches of other texts' (ibid.: 84). Fairclough (1992), though, makes a distinction between two types of intertextuality: *manifest intertextuality*, 'where explicit other texts are overtly drawn upon within a text' (ibid.: 85), and *interdiscursivity*, which consists of 'relations between different discursive formations or … different types of discourse' (ibid.: 47). This distinction is observed in this thesis.

In Chapter Seven, I refer to health narratives. A useful definition of narrative in the context of public health is:

> …any cohesive and coherent story with an identifiable beginning, middle, and end that provides information about scene, characters, and conflict; raises unanswered questions or unresolved conflict; and provides resolution.
>
> (Hinyard and Kreuter, 2007: 778)

As for approaches to discourse analysis, I make one basic distinction in the current chapter and that is between the approach typically used in communication studies and that more commonly used in applied linguistics. The former approach has been likened to content analysis (Maingueneau, 1999: 180): in other words, inferences about meaning are drawn on the basis of close reading of texts with selected samples of text used as illustrations. In contrast, in the applied linguistic approach to discourse analysis, conclusions are supported by rigorous lexico-grammatical analysis.

Having defined key terms and given brief explanations of the ways in which they are understood in this project, I now move on to describing the background to the project, before presenting a survey of relevant literature and setting out a rationale for the research questions guiding the study. The research questions and research objectives are then described in detail, and the chapter ends with a summary of the contents of the rest of the thesis.

## 1.3 Social background to the thesis

The initial motivation for undertaking this study was sparked by interest in what is often referred to as the MMR controversy. The controversy followed the publication in the *Lancet* of the now retracted article by Wakefield *et al*. (1998) reporting the discovery, among a group of children, of a syndrome characterized by a form of inflammatory bowel disease and regressive autism. A temporal association between administration of the MMR vaccine and the onset of symptoms had been reported by a significant proportion of the parents or physicians of the children. Suggestions of a link between MMR, the single measles vaccine or measles virus and regressive autism and bowel disease were not new, having previously been implicated in the development of regressive autism (Fudenberg, 1996; Miller *et al.*, 1997) or the inflammatory bowel condition known as Crohn's disease (Ekbom *et al.*, 1994; Thompson *et al.*, 1995). Interestingly, Wakefield and his colleagues stated clearly that their research 'did not prove an association between measles, mumps, and rubella vaccine and the syndrome described' (Wakefield *et al.*, 1998: 641). They did, however, imply that a causal link had not been disproved and called for further research. A further complication to the issue was introduced at a press conference held by members of the research team to announce the findings of the research. At this press conference, Dr Andrew Wakefield recommended that parents opt for single vaccines rather than MMR.

Numerous epidemiological studies followed, all of which reported finding no causal link between MMR and autism (Peltola *et al.*, 1998; DeStefano and Chen, 1999; Taylor *et al.*, 1999; Dales, Hammer and Smith, 2001; Fombonne and Chakrabarti, 2001). The small handful of scientific studies claiming a possible link between the vaccine and the syndrome were vastly outnumbered by those indicating no evidence for a link. Despite this, the MMR vaccine became associated in many parents' minds with a risk of autism (Boyce, 2007: 169-170) and uptake of MMR in the UK fell from 91.7% in 1997 to 78.9% in 2003,[5] below the level required to assure 'herd immunity'.[6][7] Fears were exacerbated by the suggestion that the mercury-based vaccine preservative thiomersal[8] might cause autism (Holton *et al.*, 2012: 9-10), even though the MMR vaccine has never contained it.[9] The MMR uptake figures for 2003 indicate that, even at the height of the vaccine scare, the majority of parents continued to accept the vaccine. There is evidence, though, that a substantial proportion of parents consented to the vaccine despite believing many of the 'MMR myths' the Department of Health claimed were prevalent at the time (Poltorak *et al.*, 2005).

One of the responses of the NHS to the fall in uptake was to set up the 'MMR The facts' website. In 2008, this was supplemented, and later replaced, by the 'NHS immunisation' website .The setting up of a dedicated website was part of a wider campaign to counter the effects of adverse publicity but it also marked recognition of the potential influence of

[5] Statistics from Health Protection Agency website: www.hpa.org.uk/cdr/archives, accessed 03/03/2008.
[6] *Herd immunity* is defined as 'a form of immunity that occurs when the vaccination of a significant portion of a population (or herd) provides a measure of protection for individuals who have not developed immunity' http://www.vaccinestoday.eu/vaccines/what-is-herd-immunity; accessed 22/04/2015.
[7] The impact of this situation is still felt today: in 2013 a measles outbreak occurred in South Wales and many health professionals attributed the outbreak to the fact that herd immunity had been compromised by the fall in uptake of the vaccine a decade earlier.
[8] Also known as *thimerosal*.
[9] Source: http://www.nhs.uk/Conditions/vaccinations/Pages/mmr-questions-answers.aspx#thiomersal; accessed 02/07/2014.

vaccine-critical websites.[10] While conducting preliminary investigations for this research in April 2005, I found that the top-ranked return on a Google search on *MMR* was the link to 'MMR The facts'. The top-ranked hit which linked to a vaccine-critical site was the link to the website of the JABS group.[11] The JABS group was particularly influential in the anti-MMR campaign. Founded in 1994, it campaigns for greater awareness of the risks of injury from vaccines and for improvements in the vaccine damage reporting and compensation systems. Although the group is concerned with the potential risks of any or all vaccines, its main focus was, and, to a certain extent, continues to be, the MMR vaccine. The JABS group was founded by Jackie Fletcher, the mother of a child who suffered brain damage ten days after receiving the MMR vaccine. JABS worked closely with Andrew Wakefield, and he and Fletcher formed part of a delegation which, in 1997, secured a meeting with the then Public Health Minister, Tessa Jowell (Boyce, 2007: 11). According to information posted on the JABS website, the group asked the minister to consider suspending administration of MMR pending medical assessments of specific children involved in multi-party legal action against MMR vaccine manufacturers. Fletcher (2012), in an article posted on the JABS website, claims that Jowell 'said she was prepared to look at all new scientific evidence' but that none of details that JABS provided of 1,200 children were examined individually and clinical histories were not taken.

What first sparked my interest in investigating the MMR debate was the mismatch between what the vast majority of scientists claimed and what many members of the public appeared to

[10] It has been estimated that accessing a vaccine-critical website for as little as ten minutes can have a marked influence on people's perception of risks from vaccination (Betsch *et al.*, 2010). Zimmerman *et al.* (2005) identify 78 English language vaccine-critical sites at the time of the MMR controversy.

[11] According to Hobson-West (2005: 93), there were ten vaccine-critical groups active in Britain at the height of the MMR controversy. JABS was one of seven of these groups which made strategic use of the internet. It was among the three vaccine-critical groups which she judges as having had a particularly high media profile at the time (the other two being AAA, Action against Autism, and Informed Parent).

believe. As I started researching the subject, I became especially interested in the beliefs of those parents who resisted, or were sceptical about the merits of, vaccination with MMR. As a discourse analyst, I became interested in discovering how they articulated claims about MMR and what persuasive resources they drew on. The JABS website offered a rich source of vaccine-critical texts: news reports, 'alternative' medical articles and documents authored by JABS activists. Crucially, though, it offered access to the voices of parents. In 2005, the JABS website included a message board where members could seek advice or air their concerns. In 2006, the message board was replaced with a discussion forum: a technologically more advanced form of message board which offers a greater range of affordances. The contents of the JABS website offered the ideal source of data for this study. The DoH-sponsored online material, meanwhile, offered a source of data representing the voice of the mainstream medical-scientific authorities and the government.

## 1.4 Overview of relevant research and rationale for the current project

### 1.4.1 Studies related to the MMR debate in the fields of medical sociology and anthropology and health communication

Since the controversy surrounding the MMR vaccine has obvious implications for public health policy, a substantial body of research has been undertaken in the field of medical sociology and anthropology and heath communication into the reasons behind the fall in uptake. Although such research is not strictly relevant to a discourse analysis study of MMR, these studies offer insights into the sorts of beliefs that parents typically expressed about MMR and the way in which the debate was framed in the media. Conclusions from such studies provide a useful starting point for the current research. What emerges from research in these fields is that parents hold a particularistic view of their child's immune system and of

the risks posed by vaccines which is at odds with the mainstream medical-epidemiological view of risk and immunity (Hobson-West, 2003; Petts and Niemeyer, 2004; Poltorak *et al.*, 2005; Cassell *et al.*, 2006; Hilton, Petticrew and Hunt, 2006). Their knowledge of their child's health and their own and family members' experiences with vaccination are also critical factors in guiding decisions (Poltorak *et al.*, 2005; Cassell *et al.*, 2006). Attitudes towards risk are fundamentally bound up with the degree of trust the individual holds towards the medical authorities and the government (Casiday, 2007; Hobson-West, 2007). Belief in alternative forms of healthcare, such as homeopathy, although not widespread among the general population, was found to be a predictor for vaccine refusal (Poltorak *et al.*, 2005; Cassell *et al.*, 2006).

Mass media publications and broadcasts represented the main source of information for parents at the time and it has been suggested that biased reporting played a key role in influencing public opinion (Begg *et al.*, 1998; Pareek and Pattison, 2000; Evans *et al.*, 2001). The media typically framed the MMR issue as a controversy between two equally balanced groups of scientists and, by so doing, gave undue prominence to anti-MMR arguments (Lewis and Speers, 2003; Speers and Lewis, 2004; Boyce, 2007; Clarke, 2008; Dixon and Clarke, 2013). Furthermore, the issue was most frequently framed as a political or public policy issue rather than a scientific issue (Boyce, 2007; Weberling McKeever, 2013). Of course, not all parents accepted the media frame uncritically. The majority negotiated the meaning, drawing on past experience, information from other sources, and relationships with family and friends, but many continued to mistrust science and the government (Boyce, 2007: 184-186). But media influence cannot be considered the decisive factor in influencing parents' vaccination decisions (Bellaby, 2003; Petts and Niemeyer, 2004). A degree of resistance to vaccination

has been a feature of public life since the smallpox vaccine was introduced in the 19th century and vaccine scares have arisen, and continue to arise, at isolated intervals ever since.[12] Bellaby (2003) suggests that parents' attitudes to MMR can only properly be understood if one considers the social and historical context in which parents were acting and that memories of the government's handling of the BSE-vCJD affair would have predisposed the public to mistrust the authorities.[13]

The literature surveyed here brings to light a number of important issues of relevance to the current research:

- issues of risk and uncertainty are central to the MMR debate;
- people's perceptions of risk are related to the degree of trust they hold in sources of authority;
- lay people have been found to hold different views about risk, the immune system and (in some cases) health from the mainstream medical-scientific establishment;
- media reporting often reframed the MMR issue as a political issue and in much media reporting, parents were afforded the status of experts;
- although influential in shaping public attitudes, media reporting was by no means the decisive factor in forming parents' beliefs about MMR;

[12] Some vaccine scares have happened for good reason. In the 1950s, it was found that the oral polio vaccine (OPV) could result in vaccine-associated paralytic poliomyelitis (VAPP). In the 1960s, some batches of the injected polio vaccine (IPV) were found to be contaminated with Simian Virus 40 (SV40), which has been implicated in causing kidney disease and cancer in humans (Shah, 2004). Other vaccine scares have less justification. In the 1980s, the swine-flu vaccine was linked with Guillain-Barré syndrome and the whooping cough vaccine was believed by some to cause whooping cough. More recently, the mercury-based preservative, thimerosal, used in some vaccines, has been linked with autism (Epstein, 2005). The introduction in the UK of the human papilloma virus (HPV) vaccine for girls has also been met with suspicion (Hilton *et al.*, 2010).

[13] In the late 1980s, it was suspected that Bovine Spongiform Encephalopathy (BSE) could cause variant Creuzfeld Jacob disease (vCJD) in humans. Following assurances of its safety, the British government eventually admitted in 1996 that eating meat from cattle infected with BSE could cause death from vCJD.

- parents form their beliefs and opinions on MMR drawing on different sources of information and on the basis of their own experience, the experience of family and friends, and through interaction with others;
- parents form their beliefs and opinions on MMR against the backdrop of a collective public memory of previous health scares, such as the BSE-vCJD scare or other vaccine scares.

### 1.4.2 Discourse studies of the MMR debate

There are relatively few applied linguistic studies of the MMR debate in the field of discourse analysis. A small number of studies of the MMR issue follow the 'communication studies' approach, as described in Section 1.2 above. Hobson-West (2005; 2007) focuses on the talk of vaccine-critical groups, analysing interview, document and website data, O'Dell and Brownlow (2005) examine news articles and reader comments on the BBC website, and Richardson (2005)[14] looks at MMR talk on Usenet newsgroup threads. Much of the research in this field highlights similar phenomena to those identified in many of the medical sociological/anthropological and communication studies referred to in Section 1.4.1 above. The centrality to the MMR debate of notions of risk and trust and the foregrounding of parents as experts are two examples. Vaccine-critical groups, in particular, foreground as experts the parents of vaccine damaged children (Hobson-West, 2005). But what also emerges is the strategic use of particular discourses. In internet-based media texts and comments sections, for example, scientific discourse is used strategically by writers on both sides of the ideological divide (O'Dell and Brownlow, 2005). Vaccine-critical groups reframe risk and

[14] Although Richardson's background is in applied linguistics, this particular book is aimed at a wide audience. She explains that the research described in the book is 'heavily influenced by the linguistic study of discourse, but with a light touch' (Richardson, 2005: 7).

trust for strategic purposes and sometimes draw on the discourse of 'alternative' medicine to further their arguments (Hobson-West, 2005; 2007).

Discourse analysis studies of the MMR issue in the field of applied linguistics are even fewer than those in the field of communication studies. When the MMR debate has been mentioned in applied linguistic works, it is sometimes for merely illustrative purposes. Hunston (2011: 26-27), for example, provides a brief analysis of a passage from an NHS website in order to illustrate the relation of expressions of epistemic status to evaluation. Jones (2013) devotes a chapter to the MMR issue in his exploration of health and risk discourse from an applied linguistic perspective, but his work is concerned chiefly with interaction in clinical settings and, as far as MMR is concerned, he analyses only a few fragments of text. The one study in this field which offers a thorough lexico-grammatical analysis of a set of MMR-related texts is Rundblad, Chilton and Hunter's (2006) comparison of the discursive strategies used in two scientific and four media articles to establish credibility. The main purpose of Rundblad, Chilton and Hunter's (ibid.) study is to test a framework for analysing texts concerning health communication. As such, it is a very small scale study. It offers valuable insights into how one might approach the analysis of scientific and popular scientific discourse, but sheds little light on the MMR controversy in general.

The studies reviewed in this section highlight the following issues relevant to the current study:

- scientific discourse has been found to be used in media texts for strategic purposes to further both pro- and anti-MMR arguments;

- notions of risk and trust, of central importance in the MMR debate, are reframed and used for strategic purposes in the talk of vaccine-critical groups;
- vaccine-critical groups sometimes draw on 'alternative' medical discourse in their arguments;
- parents, especially the parents of vaccine damaged children, are afforded the status of experts in many contexts where MMR is discussed;
- there is little applied linguistic research to date on the MMR debate.

It is clear that there is a gap in current research regarding linguistically oriented studies of MMR related vaccine-critical talk among lay people. Furthermore, there are as yet no thorough linguistic analyses of MMR related texts based on large quantities of data. This thesis aims to fill that gap by providing a detailed and rigorous linguistic analysis of MMR related vaccine-critical talk in the *JABS* corpus.

### 1.4.3 Applied linguistic discourse analysis studies of scientific discourse in the public sphere

The aim of this thesis is not simply to examine the MMR issue as an end in itself but to use the MMR debate as a case study to further understanding of the nature of public debate about health and science. The thesis is situated in the field of textlinguistic-oriented discourse analysis. This section presents an overview of work in this field on scientific discourse in order to highlight the contribution this thesis makes to the discipline. Before surveying the literature in this field, though, it is useful to make some distinctions between different types of linguistically oriented discourse analysis studies of scientific discourse. For current purposes, we can divide them into three categories. The first consists of those which are concerned with

describing the discursive practices found in scientific writing in academia[15] (e.g. Hunston, 1994; Hyland, 1998; Charles, 2006b; Groom, 2007), while the second comprises those whose main objective is to uncover the ideological messages in science-related texts produced for consumption by a non-expert audience. Studies in the latter category look at the representation of, for example, climate change (e.g. Bell, 1994; Carvalho, 2005; Grundmann and Krishnamurthy, 2010), biotechnology (e.g. Hellsten, 2003; Cook, 2004; Cook, Pieri and Robbins, 2004; Cook, Robbins and Pieri, 2006; Motion and Doolin, 2007; Yamaguchi, 2007; Henderson, Weaver and Cheney, 2007; Leitch and Davenport, 2007), the SARS outbreak (Richardson, 2005; Chiang and Duann, 2007; Joye, 2010), avian flu (Heffernan, Misturelli and Thomson, 2011), the environment (Alexander, 2010) or the representation of science in general (Taylor, 2010). Such studies highlight the particular lexical and grammatical choices made. The third category comprises those which identify themselves as studies of scientific popularization (e.g. Fahnestock, 1986; Myers, 1994; Beacco *et al.*, 2002; Calsamiglia and Lopez Ferrero, 2003; Calsamiglia and van Dijk, 2004; Ciapuscio, 2003; Gűlich, 2003; Moirand, 2003; Myers, 2003; Hyland, 2010; Luzon, 2013). Like the studies belonging to the second category, these analyse texts on science–related topics which are produced for a non-expert audience. These studies, too, consider the ways in which scientific issues are represented and framed. However, their main objective is to shed light on the particular discursive practices at work in popularization genres as distinct from scientific genres. As such, there are also similarities in approach between these and studies of writing in academic disciplines.

[15] 'Scientific' here is used in its broadest sense to refer to academic writing in general.

Many applied linguistic discourse analysis studies into scientific debate in the public sphere, whether those concerned solely with representations of science or those interested in the discursive practices of popularization, use texts from traditional printed media as sources of data. Many of these analyse news reports (e.g. Beacco *et al.*, 2002; Calsamiglia and Lopez Ferrero, 2003; Moirand, 2003; Calsamiglia and van Dijk, 2004; Carvalho, 2005; Cook, Robbins and Pieri, 2006; Grundmann and Krishnamurthy, 2010) or popular science reports (e.g. Hyland, 2010) and some consider the documents produced by organizations (e.g. Henderson, Weaver and Cheney, 2007). Studies such as these draw valuable conclusions about the sorts of messages about scientific topics that are presented for public consumption but offer no insight into the views or arguments of lay people. A little over a decade ago, Myers (2003) suggested that, for a fuller understanding of scientific debate in the public sphere, researchers needed to tap into a wide range of types of texts and to access a range of voices. A number of studies have emerged since then which address this issue. Some use interview and focus group data to gain insight into the views of a range of stakeholders, such as scientists (Cook, Pieri and Robbins, 2004; Cook, 2004; Motion and Doolin, 2007; Yamaguchi, 2007), journalists and other professional groups (Yamaguchi, 2007), or members of the public (Cook, 2004; Myers, 2004). Others draw on data from internet sources, examples being Richardson's (2001; 2003; 2005) studies of risk talk in Usenet newsgroups and Luzon's (2013) examination of science talk in blogs. Also relevant here is Myers' (2010) analysis of public debate on a range of topics (not necessarily science topics) in online blogs. Vayreda and Antaki's (2011) study of a discussion forum thread on swine flu vaccination is also worth mentioning, although this study is situated in the sub-discipline of discursive psychology rather than text linguistics. There is, however, still need for further studies into lay people's talk about science. This thesis aims to contribute to the study of scientific debate

in the public sphere by presenting an analysis of lay people's talk in data from a website which has a message board and discussion forum.

### 1.4.4 Applied linguistic research into health communication

There is a growing body of applied linguistic research into health communication. Unsurprisingly, most focuses on interaction in clinical settings (e.g. Sarangi and Roberts, 1999; Candlin and Candlin, 2003). Some notable recent studies use corpus linguistic methods to examine interaction in online settings (Adolphs *et al.*, 2004; Harvey *et al.*, 2007; Atkins and Harvey, 2010; Seale *et al.*, 2010; Harvey, 2012; Harvey, Locher and Mullany, 2013). The latter group of studies shed valuable light on the ways in which people represent their beliefs about health and the body and to that extent, they are relevant to the aims of the current study. However, they differ from the current study in that they examine interactions between lay people and healthcare professionals, where members of the public pose questions which are answered by qualified doctors. In the *JABS* corpus data, parents do indeed post questions in which they seek advice, but their respondents are other parents. Also, as a vaccine-critical website, much of the material posted on the JABS site has an overtly ideological purpose. The intention is often to exchange information and opinions as part of the campaign to influence public policy. The current study is concerned more in discovering the ways in which linguistics and intertextual resources are drawn on for the purposes of persuasion.

### 1.4.5 Rationale underpinning the thesis

The discussion so far has brought to light a gap in applied linguistic research into the MMR debate and has suggested that more research is called for in the field of scientific popularization into lay people's discourse in web-based communication. It has also

highlighted a number of pertinent features relating to the beliefs and concerns of parents and to the ways in which claims about MMR, or vaccination in general, are transformed as they travel in the public sphere. It was found that parents tend to hold different ideas from medical professionals about health, risk and the immune system. Parents reported that they seek information on MMR from a variety of sources: mass-media sources, medical sources, family and friends, and so on. It was also suggested that particular socio-historical factors (such as memories of other vaccine scares or the BSE-vCJD affair) might impact on parents' vaccination decisions. But parents also base their decisions on their knowledge of their child's, or other family members', history with vaccination. This suggests that lay people's talk about vaccination is likely to consist of a complex intertextual mix, characterized by different 'discourses' about health, risk and the immune system and, perhaps, personal anecdotes and narratives of health scares. But there are likely to be other intertextual influences in the mix. Mass-media publications and broadcasts were found to represent the main source of information. But the media typically framed the MMR stories not as a science issue but as a political or public policy issue or as controversy between competing groups of scientists. Strategic reframing of issues of risk and trust was found to be a feature of the discourse of vaccine-critical groups. These groups were also found to make strategic use of scientific discourse and to emphasize the expertise of parents.

The findings summarized above echo some of the findings in recent research into scientific popularization. It has long been recognized by scholars in this field that when scientific claims move from their source genre to a popularization genre, they undergo particular lexico-grammatical and rhetorical transformations (Fahnestock, 1986; Myers, 1994). Recent research shows that the processes by which scientific claims are disseminated in the public sphere are

more complex than first thought. Scientific claims circulate in a network of intertextually related genres (Solin, 2004) and meanings are negotiated during the course of interaction (Ciapuscio, 2003; Gűlich, 2003) or emerge as scientific claims move between discourse communities (Beacco *et al.*, 2002; Moirand, 2003). Nor is a strict division between the expert and the lay-person uncritically accepted nowadays. Instead, it is accepted that 'lay' people have their own forms of expertise, either actively acquired or gained through personal experience, and thus have particular persuasive resources they can draw on (Myers, 2003: 268-269). Myers (ibid.), in fact, cites the use by the media of parents as experts in the MMR debate as an example of the persuasive potential of lay people's arguments. The ways in which beliefs about health, risk and immunity are expressed and how expertise is performed form a major focus of this study.

As mentioned in Section 1.3 above, the JABS website was chosen as a data source for this project because it offers access to the voices of lay people. It also offers access to a wide range of views and ideological positions. JABS is a *reformist vaccine-critical* group, rather than a *radical* one (Hobson-West, 2005). *Reformist vaccine critical* groups are led by parents who believe their children have suffered vaccine damage and who campaign for reform to the compensation system and increased awareness of vaccine risks. Not all members of reformist vaccine-critical groups are opposed to vaccination per se. *Radical vaccine-critical* groups are those whose members do not necessarily have experience of vaccine damage but who oppose the use of vaccines and are often in favour of 'alternative' medicine and opposed to big pharmaceutical companies. Although not all JABS members are opposed to vaccination in general, most support Andrew Wakefield and find his hypothesis of a causal connection between MMR and autism, if not convincing, certainly plausible. Between 2005 and 2008,

when the data was collected, the message board and discussion forum moderators allowed people to post who were openly critical of Wakefield's hypothesis. The presence of these 'pro-science' posters grew from 2007 to the end of 2008, leading to much heated debate on some of the forum threads. The *JABS* corpus data is therefore ideal for finding out what discourses people deploy and what sorts of persuasive resources people draw on when arguing about MMR online.

We considered above the sort of intertextual mix we might expect to find in the *JABS* corpus. We now need to consider another aspect of argumentation. In short, in order to make an argument persuasive, a writer needs to establish credibility and to align the reader with his or her point of view. This involves deploying various lexico-grammatical resources which signal attitudinal meaning. It also involves attribution. A powerful tool for uncovering attitudinal meaning is the study of *evaluation*.[16] I discuss evaluation in full in Chapter Three. For now, I follow Hunston (2000; 2011) in highlighting expressions of *status* (expressions which signal the writer's judgement of the epistemic status of a discursive object) and *value* (expressions which encode a judgement of the qualities of an object) as performing key functions in the expression of evaluation.

Obviously, the *JABS* corpus comprises web-based data. The technical features of new media offer the user types of discursive strategies not possible with traditional print based media, for example, the incorporation of texts or links to other websites to provide explanations or lend authority to their claims (Luzon, 2013). Using corpus assisted methods to analyse discussion

[16] This phenomenon is variously termed *stance* (e.g. Biber and Finegan, 1988; 1989), *APPRAISAL* (Martin and White, 2005), or *evaluation* (e.g. Hunston, 2000) (among other things). For current purposes, I use the term *evaluation*.

forum data raises its own challenges: the discussion forum allows users to copy large sections of texts or whole texts into their posts. Since the concordance line offers only a glimpse of a text, it can be hard for the discourse analyst to distinguish between the voice of the forum user and that of an external writer. This thesis attempts to address some of the challenges posed conducting a corpus assisted discourse analysis of discussion forum data.

## 1.5 Objectives and research questions

The main objective of this thesis is to advance understanding of public debates about health and science by providing a detailed lexico-grammatical examination of the ways in which contributors on the vaccine-critical JABS website draw on discursive resources to construct persuasive arguments. A further objective is to explore some of the opportunities afforded by, and to address some of the methodological challenges involved in, conducting a corpus assisted discourse analysis of web-based texts including internet discussion forum threads.

The research questions guiding this thesis are:

- How are beliefs about risk, health and the immune system expressed in the *JABS* corpus and to what extent are linguistic realizations of ideational content in the *JABS* corpus similar to and different from those in the *NHSvax* corpus?
- How are discursive objects and sources of authority evaluated in the *JABS* and *NHSvax* corpora and what are the similarities and differences?
- To what extent, and how, are different discourses used strategically in the *JABS* corpus to further arguments?
- How do interactants on the *JABS* website discussion forum express their warrant for expertise?

The empirical objectives are:

- To describe the typical lexico-grammatical features used to express notions of health, risk and immunity in the *JABS* and *NHSvax* corpora and to identify similarities and differences between the two;
- To identify the typical lexico-grammatical features used to realize assessments of status and value in the *JABS* and *NHSvax* corpus and to identify similarities and differences between the two;
- To identify common discursive strategies and to describe the functions they fulfil.

A further objective is methodological, and that is to assess what opportunities and challenges are afforded by carrying out a corpus assisted discourse study of interactive web-based texts and to suggest how challenges might be addressed.

## 1.6 Structure of the thesis

The rest of the thesis is structured as follows. Chapter Two examines literature on the MMR issue from the point of view of risk and as an example of the ways in which scientific claims travel in the public sphere. It begins with a discussion of the literature on risk and literature on MMR and risk, before considering the ways in which notions of risk are expressed linguistically. It moves on to discuss theories about the processes of scientific popularization, notions of expertise, and the functions of intertextuality in the context of web-based interaction. Chapter Three sets out the theories of language and principles of discourse and corpus analysis which inform the project. It begins with a survey of CADS studies. It then goes on to discuss definitions of *discourse* and *text*, before problematizing the Foucauldian

notion of discourse and evaluating approaches to studying discourse. It argues for adopting a realist rather than constructionist approach to interpreting the relationship between discourse and the world. The chapter continues by evaluating the strengths and weaknesses of the Sinclairean approach to corpus linguistics, and argues for following this approach. The chapter then examines approaches to analysing evaluation in language and considers corpus approaches to evaluation. Chapter Four discusses methodological issues in corpus linguistics and CADS before describing the process of constructing the *JABS* and *NHSvax* corpora and the method of analysis followed in this project. It also describes salient features of the *JABS* data, focussing in particular on features of the discussion forum data, such as the number of forum members and the average frequency of posting. The chapter discusses some of the opportunities and challenges posed by the task of analysing a large corpus composed mainly of discussion forum data using corpus linguistics methods. Chapter Five examines the main lexico-grammatical patterns used in the *NHSvax* corpus in order to gain an idea of the ways in which the MMR causal hypothesis, health, immunity and risk are talked about in mainstream medical-scientific discourse. Chapter Six examines the ways in which notions of health, immunity, risk and causation are expressed in the *JABS* corpus, comparing the common lexico-grammatical patterns found in this corpus with the findings from the *NHSvax* corpus. Chapter Seven examines the ways in which intertextuality is exploited in the *JABS* corpus, how attribution is expressed, and how claims to knowledge are framed. Chapter Eight concludes the thesis by summarizing the main findings and considering the contribution to knowledge this thesis has made, its limitations, and implications for future research.

# CHAPTER TWO

# THE DISCOURSE OF SCIENCE, HEALTH AND RISK IN THE PUBLIC SPHERE

## 2.1 Introduction

The aim of this thesis is to discover some of the ways in which discursive resources are drawn on by the writers represented in the *JABS* corpus to construct persuasive arguments related to the safety and risks of the MMR vaccine. It also aims to discover the similarities and differences between language use in the *JABS* and *NHSvax* corpora. In so doing, the project seeks to contribute to the wider study of public debate about scientific issues. This involves examining the ways in which scientific claims about health and risk are expressed, and how their meanings are negotiated, in the various articles, information documents and forum discussion threads, which make up the *JABS* corpus. It also involves examining how writers attempt to establish credibility: how they signal that they have a right to contribute to a discussion and convey the idea that their views are authoritative.

The aims of this chapter are to evaluate relevant literature on the MMR controversy, on scientific and popular scientific discourse, risk discourse and the study of evaluative language. Section 2.2 considers the MMR issue from the point of view of risk, discussing literature on risk discourse and the MMR debate, and examining the ways notions of risk are expressed in language. Section 2.3 discusses theories related to the communication of scientific claims in the public sphere. It starts with a consideration of what is meant by the public sphere, moves

on to discuss theories about the processes of popularization, before considering the ways in which expertise is expressed and intertextuality is used in online forums. Section 2.4 concludes the chapter.

## 2.2 Risk, science, health and the MMR debate

### 2.2.1 Introduction to Section 2.2

The MMR debate can be seen very much as an expression of the preoccupation with risk which characterises modern society (Beck, 1992; Giddens, 1991). Anxiety about health risks is particularly prominent in modern life, possibly exacerbated by the emphasis in contemporary public health policy on personal responsibility for health (Mythen, 2004: 140). Risk, nowadays, is viewed as arising from internal decision-making, so that the individual in modern society is sensitive to the knowledge that certain courses of action (or inaction) entail risk and may result in undesirable outcomes (Beck, 1992: 155). Modern society expects the individual to put their trust in experts and institutions, or *expert systems*. At the same time, there is a general decline in trust in expert systems, caused, it is suggested, by the uncertainties created by modern life (Giddens, 1991). Once the possibility was raised that the MMR vaccine might entail a risk of a child developing regressive autism and bowel disease, it was understandable that many parents started to fear that vaccination with MMR was too hazardous to subject their child to.

It is because concern with risk lies at the heart of the MMR debate that much sociological and health communication research into the topic has focused on risk. The majority of studies in this field examine what people say about MMR and risk. Most of these are interested in the

views of parents (Ramsay *et al.*, 2002; Petts *et al.*, 2003; Petts and Niemeyer, 2004; Casiday, 2005; 2007; Poltorak *et al.*, 2005; Casiday *et al.*, 2006; Cassell *et al.*, 2006; Leask *et al.*, 2006), although Hobson-West (2003; 2005; 2007) focuses on vaccine-critical groups. The current study has much in common with these, so that their findings are relevant to its aims. The *JABS* corpus offers access to the voices of people who are active vaccine-critical campaigners, as well as to the voices of parents who are merely sceptical or uncertain about MMR risk. However, this study is concerned more with how opinions are expressed than what is expressed. There are a number of studies which look at how representations of MMR risks are constructed, but most are concerned with the effects of the processes of news production on the message (Lewis and Speers, 2003; Speers and Lewis, 2004; Boyce, 2007; Clarke, 2008; 2011; Dixon and Clarke, 2013). Insights from studies such as these are, of course, relevant to the current project, since the mass media plays such an important role in public debate. Furthermore, media texts make up a large proportion of the *JABS* corpus. But this project is more similar in its aims to Richardson's (2005) study of the ways in which arguments about MMR risks are put forward in newsgroup interaction. This study is concerned with discovering how available lexico-grammatical and intertextual resources are exploited in the *JABS* corpus. The rest of this section therefore discusses the literature on parental attitudes to the risks of MMR and on vaccine-critical discourse, before moving on to discussing the findings on media representations of MMR risks, on representations of risk in online forums, before finally considering the ways in which risk and causation are expressed linguistically.

### 2.2.2 Parents' attitudes towards MMR

What emerges from the literature on parental attitudes towards MMR is that, to a certain extent, vaccination decisions are based on an assessment of the relative risks of the vaccination compared with the diseases, although this is by no means the whole picture (Casiday, 2007). Some of the literature identifies a deficit in public understanding of risk as a major cause for the fall in uptake (e.g. Burgess, Burgess and Leask, 2006). What is claimed is the idea that the standard technical model of risk, based on a statistical calculation of probability (Lupton, 1999: 5-6), is poorly understood by the public, whose assessment of risk is based on 'lay' understandings. It is true that parents tend to hold a more particularised view of risk, connected to the idea that each child's immune system is unique (Poltorak *et al.*, 2005; Cassell *et al.*, 2006), and that they generally base their assessment of risk on factors such as the individual child's health history and the family history of vaccination (Poltorak *et al.*, 2005; Cassell *et al.*, 2006). But that is not to say that parents are incapable of understanding epidemiological models of risk. Parents are able to engage with information about risk presented to them in terms of statistical probabilities (Petts and Niemeyer, 2004; Casiday, 2007), although the parent's level of education is a factor here (Petts and Niemeyer, 2004).

But factors concerning identity and trust also play a role. As well as considering their assessment of the relative risks of the diseases and the vaccine, parents were found to take into consideration their ideas about what it means to be a 'good' parent (Casiday, 2007). Part of the job of the responsible parent was seen as the obligation to make the right decision about vaccination, in the midst of uncertainty and contradictory information (ibid.). Understandably, many parents fear they may live to regret making the wrong decision (Leask *et al.*, 2006).

There is evidence, too, of tension between parents' sense of responsibility towards their own child and societal concerns for the public good (Leask *et al.*, 2006; Casiday, 2007). Several studies highlight trust as a key factor in decision making (Petts and Niemeyer, 2004; Casiday, 2005; 2007; Casiday *et al.*, 2006; Leask *et al.*, 2006). Many parents cited a lack of trust in medical experts, in particular doctors and the medical authorities, and often turned to other trusted sources for advice, such as family and friends (Petts and Niemeyer, 2004; Poltorak *et al.*, 2005). The lack of trust in authorities is exacerbated by the perceived contradiction between, on the one hand, the promotion of patient choice, and, on the other, a policy which denies parents the choice of single vaccines (Poltorak *et al.*, 2005). Sometimes medical beliefs guide vaccination decisions. Belief in 'alternative' approaches to healthcare, such as homeopathy, is found to be a predictor for non-compliance with vaccination policy (Poltorak *et al.*, 2005; Cassell *et al.*, 2006).

### 2.2.3 Vaccine-critical groups and arguments about MMR risks

Research in this field does not report finding expressions of 'lay' understandings of risk at odds with expert understanding. Instead, vaccine-critical groups are found to make strategic use of risk discourse and to reframe the discussion of risk. These groups frequently present risks in terms of unknowns. They claim there have been insufficient trials of vaccines, point to inadequacies in the reporting system for adverse events, and suggest there is insufficient evidence of the effects of mass immunisation (Hobson-West, 2005: 109-112). Criticising epidemiological models of health on the grounds that they overlook individual particularities, they claim medical-science has insufficient knowledge of how the body works, that the one-size-fits-all policy of mass vaccination overlooks the individual who may suffer from an adverse effect to vaccination, and they call for a healthcare delivery based on a calculation of

risk tailored to the individual rather than calculated on population-based statistics (ibid.: 111-117). Some of the other arguments they put forward focus on bias in the ways in which vaccination messages are communicated by the authorities. They argue that risk discourse is manipulated, that knowledge about the risks of vaccination is concealed from the public and that claims about the risks of diseases are exaggerated, and they counter the dominant narrative that mass immunisation has been an unqualified success, emphasising instead the risks from dangers such as autism and other chronic diseases (ibid.: 119-130). They also sometimes refer to some of the wilder claims which have been made about vaccines, such as the supposed link between vaccination and infertility or between the polio vaccine and HIV (ibid.: 131-133).

As well as reframing risk in these ways, vaccine-critical groups make strategic use of the connection in people's minds between risk and trust. They represent uncritical trust in authority in a negative way, portraying 'blind faith' as risky behaviour in itself and presenting personal responsibility and empowerment as alternatives (Hobson-West, 2007). Questions of trust are also used to challenge the reliability of experts. The power of pharmaceutical companies (often referred to as *big pharma*) comes in for particular criticism. Because they are seen as profiting from government vaccination policy, they are not trusted, and the reliability of any medical expert who is perceived as having links with a pharmaceutical company is seen as compromised (Hobson-West, 2005: 141-143). Vaccine-critical groups have an ambivalent attitude towards science, though. JABS is not among those groups classed as 'radical' vaccine-critical groups (Hobson-West, 2005), but some JABS members do hold radical views. Radical groups are typically mistrustful of science and warn against glorifying science (ibid.: 163). Science is often contrasted with nature (ibid.: 164) and many of the

fundamental precepts of epidemiology are challenged, in particular the germ theory of disease and the theory of antibody response (ibid.: 167-170). Reformist groups are more likely to argue that the wrong type of science is being used and tend to contrast epidemiological research with clinical research or research into genetics (ibid., 2005: 171-174).

### 2.2.4 Risk discourse and the MMR debate in the media

Consideration of the ways in which the media reported the issue of risk and the MMR vaccine is relevant to the aims of this project because media texts are frequently used on the JABS site as a discursive resource. Furthermore, the influence of the media in promoting messages about the causal hypothesis cannot be underestimated. Although only a small proportion of parents are influenced by media stories to the extent that they change their decisions on vaccination (e.g. Poltorak *et al.*, 2005; Boyce, 2007), media reports were found to be the main source from which members of the public gleaned information on the supposed MMR-autism link (Boyce, 2007: 87-88). Media reporting tends to amplify perceptions of risk. One of the main ways in which this happened in the case of the MMR controversy was through 'false balance' (Lewis and Speers, 2003; Speers and Lewis, 2004; Boyce, 2007; Clarke, 2008; Dixon and Clarke, 2013). In the interests of presenting a balanced view of the matter, journalists tended to give equal weight to pro- and anti-MMR arguments, despite the fact that the number of scientists who believed that MMR was safe greatly outnumbered those who supported the causal hypothesis. This resulted in 'overbalancing', that is, over-representing a minority argument. The MMR issue was a prime candidate for over-balancing since this phenomenon frequently occurs when 'new evidence challenges the status quo or when maverick scientists use the media to secure coverage' (Boyce, 2007: 74). Some articles at the time suffered from 'under-balancing', that is, they represented only one set of studies or

claims. This is estimated to have been a feature of 52% of media stories on MMR in 2002, with 32% putting forward only anti-MMR arguments and 20% underbalanced in favour of pro-MMR arguments (Boyce, 2007: 75). In 'balanced' reports, pro-MMR voices were often given minimal space (ibid.: 76). The over-representation of anti-MMR voices gave the impression that the science of vaccination was uncertain and made the MMR-autism link appear plausible (Clarke, 2008; Dixon and Clarke, 2013). As is typical of scientific popularization writing, claims about MMR were expressed with a higher degree of certainty than was warranted (Boyce, 2007: 26-27), while, in interviews, journalists often demanded of scientists greater certainty about facts than science is capable of assuring (ibid.: 29).

News values also have an effect in amplifying messages about risk (Boyce, 2007). Stories about health risks, such as the MMR story, are considered more newsworthy if they can be framed as a controversy (Kitzinger, 1999: 63). The presentation of a story as a controversy lends itself to the narrative device, common in journalistic practice, of setting up binary oppositions (Seale, 2002: 29-30). Two main binary oppositions characterised MMR reporting: a) the juxtaposition of supporters of the MMR-autism hypothesis and the rest of the medical establishment; b) the juxtaposition of MMR vaccine and single vaccines (Boyce, 2007: 26). Science stories are also more newsworthy if a human face can be put to the story, if there is evidence of conflict between stakeholders, or if there is the possibility of attaching blame to the government or other institution (Kitzinger, 1999: 63). The media therefore foregrounded Andrew Wakefield as a maverick scientist, standing up against the medical establishment and the government (Boyce, 2007). The possibility of framing the story as a political issue made the MMR controversy particularly attractive. Science stories which can be framed as political issues are more newsworthy because they enable the journalist to avoid lengthy descriptions

of scientific detail (Boyce, 2007: 45-46). In her analysis of 285 MMR-related news stories from 2002, in fact, Boyce (2007: 59-60) found that only 13% foregrounded Wakefield's research, while by far the majority (59%) represented the issue as a political or public policy matter, often focusing on the extent to which politicians could be trusted.

Strategic use of sources also helped promote Andrew Wakefield's arguments. The voices of parents and pressure groups opposed to MMR were often given a more privileged position than sources with scientific expertise. Although Boyce (2007) found that, in the news reports she surveyed, scientists and health professionals accounted for a similar percentage of the sources used,[17] it became clear that, by routinely juxtaposing the scientific experts with parents and pressure group representatives, the lived experience and opinions of the latter were represented as equivalent to scientific expertise (ibid.: 98). Furthermore, in 26% of broadcast reports, parents and pressure group representatives were used as the first source in a story, thus allowing them to frame the debate and putting pro-MMR voices, often scientists, in the position of having to defend the science (ibid.: 99; 143). Lending support to the claim that the MMR issue was often framed as a political issue, it was found that media publications which were critical of the then Labour government used more anti- than pro-MMR sources, while the converse was the case with publications sympathetic to the government (ibid.: 81).[18]

### 2.2.5 Risk and the MMR-autism issue in online forums

The discussion forum of the JABS website fulfils similar functions to those of newsgroups in that it provides a forum for the exchange of news and a means for establishing and

[17] Among the sources used, 23% were scientists and health professionals and 22% were parents and pressure group representatives Boyce (2007: 98).
[18] In the *Mail on Sunday*, for example, 65% of sources used were anti-MMR and 19% pro-MMR, while in the *Guardian*, 74% were pro-MMR and 14% anti-MMR.

maintaining relationships (Richardson, 2001). One could argue that the value of the JABS forum in fulfilling an interpersonal function is greater than that of the newsgroups which Richardson (2001; 2003; 2005) examined, since JABS exists as a campaign group, not just as an online phenomenon. The newsgroups which she examines are similar to the JABS forum in that they are concerned with health risks, whether related to the link between BSE and vCJD (Richardson, 2001), the risk of contracting brain tumours from mobile phone technology (Richardson, 2003; 2005), the SARS outbreak (Richardson, 2005), or the hypothesised MMR-autism link (ibid.). Discussions about the risks of MMR on newsgroups are characterised by expressions of uncertainty (Richardson, 2005), although, in Richardson's (ibid.) data, roughly the same proportion of writers in her corpus of 1,203 messages expressed confidence in the safety of the MMR vaccine (330 messages) as those who expressed doubts (291 messages). People who talked down the risks of MMR, often drew on commonplaces, such as comparisons with driving a car (ibid.: 145-146).[19] Those who chose to talk up the risk, drew comparisons with other health risks, such as smoking or eating beef (Richardson, 2005).

### 2.2.6 Expressing notions of risk

Risk is at the heart of the MMR debate. It is therefore helpful to see how notions of risk are expressed in the *JABS* corpus. The discussion above referred to the technical understanding of risk as opposed to lay understandings. To reiterate, in its technical uses, *risk* is understood as the degree of probability of loss or injury resulting from a particular hazard (Kaplan and Garrick, 1981). It is related to, but distinct from, the concept of uncertainty. But it is questionable whether there is such a clear distinction between lay and expert uses of *risk*. In

[19] Commonplaces are defined as 'short evocations of standard arguments that will work in many situations' (Myers and Mcnaghten, 1998, cited in Myers, 2005: 536) and which consist of references to socially shared experiences and beliefs (Myers, 2005: 536).

some respects, lay notions of risk are not wholly dissimilar from technical notions. As Kaplan and Garrick (ibid.) observe:

> In analyzing risk we are attempting to envision how the future will turn out if we undertake a certain course of action (or inaction). Fundamentally, therefore, a risk analysis consists of an answer to the following three questions:
>
> What can happen? (i.e., What can go wrong?)
> How likely is it that that will happen?
> If it does happen, what are the consequences?
>
> (Kaplan and Garrick, 1981: 12-13)

It is precisely this sort of consideration which many parents make when deciding whether to opt for the MMR vaccination for their child. A key difference lies in the fact that technical calculations of the probability of an undesirable outcome are based on the frequency of similar past events. Epidemiologists therefore base their calculations of the risk of adverse events arising from vaccination on data relating to the general population. Although risk calculations are normally expressed as a single number, they are based on assessments of a range of factors. One thing which is important to stress is that, for risk analysts, risk is always relative, never absolute, (ibid.: 25). Finally, risk analysis involves consideration of potential costs and benefits (ibid.). We might therefore safely predict that, when used in technical contexts, *RISK* collocates with words such as *relative,* or *cost(s)* and *benefit(s).*

On the face of it, there are differences in the ways in which notions of risk are expressed in technical and non-technical contexts. However, corpus analyses have indicated that the distinction is not as absolute as much of the literature on risk research suggests. The RISK frame, which applies to the lexeme *RISK* and to semantically related terms such as *DANGER*, *HAZARD*, and so on, has two sub-frames, CHANCE and HARM (Fillmore and Atkins, 1992:

80).[20] The activation of one or the other sub-frame depends on whether the co-textual environment does or does not explicitly represent the state of being at risk as resulting from someone's action' (ibid.). Expressions such as *at risk* or *in danger* or *RUN the risk* represent the individual as being in a hazardous situation, while expressions such as *TAKE the/a risk* or *PUT (someone) at* represent the individual as being faced with a choice and as facing the possibility that harm may result from the actions of the individual or another person or thing (ibid.: 80). Semantically, it denotes ideas about choice and possibilities, but uncertainty is a core part of the meaning of the CHANCE element (ibid.: 81). We therefore see that the concept of risk in general use is broader than in technical contexts, but it does cover some of the same semantic space. Fillmore and Atkins (1992: 85) also find a number of expressions containing *risk* which relate to the Actor's awareness of, and calculation of, risks. They cite expressions such as *KNOW THE RISK, UNDERSTAND THE RISK, APPRECIATE THE RISK, CALCULATE THE RISKS* and *BALANCE THE RISKS.*

Hamilton, Adolphs and Nerlich (2007) explicitly challenge the assumption that there is a clear-cut distinction between technical and lay meanings of *risk*. They examine the uses of *risk* in conversations belonging to five different categories of interaction in the CANCODE corpus of spoken English: the *intimate* category (family and partners), the *socio-cultural* category (friends who socialize), the *professional* category (workplace interaction), the *transactional* category (service encounters, which include insurance encounters and some medical interviews), and the *pedagogical* category (teaching situations, with an emphasis on medical lectures and seminars). Although findings on all five categories are relevant, particular focus will be given here to the intimate, socio-cultural and pedagogic categories.

[20] *Frames* are defined as 'cognitive structures' (Fillmore and Atkins, 1992: 75).

These are of particular relevance to this study, because the first two represent informal interaction between family, friends and colleagues, and the last concerns discourse from the medical-scientific register. The authors find similarities and differences between all five categories in the co-textual environments of *risk*. In all categories, expressions denoting assessments of risk were evident. In the intimate and socio-cultural categories, such assessments were characterised by expressions of hypotheticality or uncertainty, and discussion of risk typically referred to the actions, or situations, of specific individuals. Modal markers such as *might, shall, think, just, always, could, should* and *would* occurred frequently. Assessments of degrees of risk were expressed and assessments ranged from zero risk, through low to high risk. These were signalled by expressions of negative polarity, such as *there isn't*, or by adverbs and adjectives such as *low, small, kind (of), less, big, more* (ibid.: 173-174). In the pedagogic category, *risk* was, unsurprisingly, typically used to describe the possible effects of an action or health condition (ibid.: 177). In the corpus examples the authors supply, it is noticeable that assessments of degrees of risk are most frequently signalled through use of the lexeme *INCREASE*, or, less frequently, *HIGH*. Examples are *significantly increased risk of a number of problems, increases your risk of a primary liver cell tumour, a higher risk of dental caries, fivefold increase in the risk of heart attack.* Calculations of degrees of risk are often based on considerations of specific factors. This is signalled by the use of the expression *risk factors,* as in *risk factors for diabetes*. The assessment of the extent to which an individual is assessed as being vulnerable to a particular unwanted outcome is expressed by the expression *at risk of*. In the professional category, *high risk* was often used as a noun modifier in the expression *high risk items,* while in the transactional category, risk is regarded as an object to be managed: the technical expression *risk management* occurred repeatedly.

The meanings of the word *risk*, therefore, shift from one context to another. It is an over-simplification to talk about a dichotomy between technical and lay understandings. However, one can point to particular phraseologies which are more frequent in medical or professional contexts and others which are more frequent in informal interaction. Furthermore, it has been found that participants in focus group discussions on risk do indeed often talk about risk in technical terms, using 'the language of risk assessment, comparison, costs and benefits' (Myers, 2007: 290). The corpus-assisted study by Hamilton, Adolphs and Nerlich (2007), in particular, has brought to light interesting features related to the lexico-grammatical contexts in which *RISK* usually occurs. Corpus linguistic techniques are well suited to illuminating such features.

### 2.2.7 Expressing causation

The MMR controversy centred on the perceived risks of the vaccine, but the perception of risk emerged from the hypothesis that there was a causal connection between the vaccine and a syndrome characterised by regressive autism and pervasive bowel disorder. Examination of expressions of causation is therefore crucial to the current study. Of course, cause-effect relations are not always explicitly signalled in texts (Carter, 1998: 92) and corpus methods are best suited to examining explicit signals of meaning. However, the strength of corpus analysis is that it can account for frequent patterns in language use. Analysis of repeated instances of explicit signalling of causation is therefore suitable for the aims of this study. A salient question, though, is precisely what items constitute explicit signals of causation. An obvious starting point is to look for those Vocabulary 1 and 2 items (Winter, 1977) which express

causal relations, including subordinators such as *by*, and sentence connectors such as *because, therefore, as a result of*, and so on. Vocabulary 3 items (ibid.) are much more diverse but might involve obviously explicit signals such as *cause* or less obvious items such as *instrumental*, as in *x was instrumental in …* (Hoey, 1983: 23-24). Xuelan and Kennedy (1992) identified 130 causative devices in the LOB corpus. These included unambiguous examples of 'explicit causatives', such as *cause, effect, because (of), result from, as a result (of), lead to, therefore* and *outcome*; ambiguous examples (i.e. those which express causation in only certain contexts) such as *produce, through, as,* and *accordingly*; and 'non-explicit causatives', comprising verbs which entail the meaning of causing something to happen or causing someone/something to do something, such as *make, destroy, damage*. A cause-effect relation may be implied through other means, though. A mother who asserts that her child developed autism *after* vaccination with MMR might do so in order to imply a causal connection.

Nominalizations may act as a signal of causality. The effects of nominalization and the related phenomena of passivization and ergativity in scientific texts have been much discussed and are well known. The preference for passive and ergative forms and for nominalization serves to give an appearance of objectivity. But use of these phenomena also helps to construct taxonomies, as well as creating textual cohesion and driving the argument forward (Halliday, 1998; Veel, 1998). Nominalization is a particularly powerful tool. By condensing clausal meaning into a nominal group, nominalization represents processes as entities which can then participate in processes themselves (Halliday, 1998: 199). But nominalization, like use of the passive and ergative voice, conceals agency and, in so doing, can obscure cause-effect relations (Veel, 1998: 115-116). This is of particular relevance to the MMR debate. In the article by Wakefield *et al.* (1998), which sparked the controversy, the authors use the terms

*association between* and *link between/with* when talking about the causal hypothesis. They state, 'We did not prove an association between measles, mumps, and rubella vaccine and the syndrome described' and, 'If there is a causal link between measles, mumps, and rubella vaccine and this syndrome …', and 'Published evidence is inadequate to show whether there is a change in incidence or a link with measles, mumps, and rubella vaccine'. Use of the nominalized forms *association between* and *link between/with* does not make clear the nature of the supposed causal relation. Furthermore, use of the modifier *causal* triggers the presupposition that associations or links of this kind may equally well not be causal in origin. Evidence from the *JABS* corpus data indicates that *link between* is the most common expression used to refer to the causal hypothesis.

### 2.2.8 Conclusion to Section 2.2

This section has highlighted the following:

- MMR discourse is characterised by a preoccupation with risk and uncertainty: parents reach vaccination decisions against a background of uncertainty and online newsgroup discussions of risk are typified by expressions of uncertainty;
- Parents have been found to hold a particularized notion of risk based on an individualized view of their child's immune system and vaccination decisions are based on their personal experience and related to the degree to which they trust the medical authorities;
- Media sources were influential in influencing public opinion and the framing of MMR news stories and 'false balance' contributed to amplifying perceptions of the risks of MMR;

- Vaccine-critical groups exploit notions of uncertainty, risk and trust: they reframe discussion of risk as uncertainty, claiming that scientific understanding of immune system function is insufficient; they challenge the expertise of scientist and they challenge the claims they make, suggesting that claims about the dangers of diseases are exaggerated and that knowledge is often concealed from the public;
- Vaccine-critical groups exploit medical-scientific discourse in order to counter mainstream arguments: they challenge the precepts of epidemiology (the germ theory of disease and antibody response); they claim that clinical, not epidemiological, science should inform vaccination policy;
- Vaccine-critical groups exploit the notion of the parent as expert;
- There are similarities and differences in the ways in which notions of risk are expressed in different contexts: in all contexts assessments of degrees of risk are evident; in informal contexts, risk may be assessed on a cline from no risk to high risk; in health communication, it is frequently expressed in terms of risk factors for negative outcomes or increases in degrees of risk; in some work-based contexts, the technical term *risk management* is frequently used;
- In the MMR debate, notions of causation are closely related to notions of risk – causal relations may be expressed using overtly explicit means or through use of specific material processes; when nominalized forms are used, as is often the case in talk about MMR risks, agency may be concealed.

The discussion now moves to considering the ways in which meanings shift as scientific claims travel in the public sphere.

## 2.3 Scientific discourse in the public sphere

### 2.3.1 The public sphere

This project aims to find out how claims about science, health and vaccination risks are expressed once they leave the scientific domain and enter the public sphere. Before going any further, we need to consider what we mean by *the public sphere*. Defined as 'the sphere of private people come together as a public' (Habermas, 1989: 27), the original conception of the public sphere was as a broad and unitary entity, standing in opposition to the state and acting as a forum for negotiation of matters of public policy. Nowadays the public sphere is seen as fragmented. The perceived fragmentation arises from a shift in the nature of public debate. For example, political struggles nowadays are more likely to be organised around group identities and shared interests than to be based on the class divisions which characterised 19th and early 20th century politics (Fairclough, 2003: 40). The debate about the risks of MMR is typical of this trend. Vaccine safety is a particular concern for certain stakeholders, whether parents, health professionals or policy makers, and organised campaign groups, such as JABS, played an active role in driving the debate. Notions of identity are thus bound up with public sphere debate. Myers (2004: 7) points to the ways in which focus group participants often refer to particular aspects of their identity to provide a warrant for their expertise, for example, prefacing a comment with 'as a mother', and so on. This is a common feature of interaction on the JABS website.

But the public sphere is not composed entirely of private individuals. Habermas (1989: 16) emphasizes the crucial role played by 'traffic in news' in facilitating the public sphere, with commercial media institutions at the heart of the process. These still play an important role in

transmitting information and influencing public opinion today, but the growth of online media has extended the public sphere, offering the potential for greater public participation in deliberation (Dahlberg, 2001; Dahlgren, 2005). A salient question is the extent to which media sources inform people's opinions and the extent to which people have opinions of their own already. Section 2.2 above discussed the fact that media messages, although influential in causing people to doubt the safety of the vaccine, were not the decisive factor in influencing parents' decisions because a range of other factors came into play. Pre-existing attitudes and beliefs may also have played a role, as findings from opinion research would seem to indicate.

### 2.3.2 The processes of scientific popularization

#### 2.3.2.1 Lexico-grammatical and rhetorical shifts in expression

As mentioned in Section 2.2 above, this study differs from previous studies of vaccine-critical discourse or MMR discourse online in that it is concerned with uncovering the ways in which scientific claims travel and how they are rearticulated as they do so. The complexity of the processes involved is increasingly recognised. The 'canonical' or 'culturally dominant' view of scientific popularization, whereby scientific knowledge is disseminated to the public in a simplified and, at worst, distorted, form (Hilgartner, 1990) has been roundly challenged. The dissemination of scientific knowledge to a wider audience does not take place by means of a one-way, linear process; instead, scientific and popularization genres interact with each other in complex ways (Myers, 2003). Scientific claims circulate in a complex network of intertextually related genres (Solin, 2004). As scientists make greater efforts to enhance public understanding of science, the boundaries between scientific and popularized accounts have become more porous, with some scientific journals making a concerted effort to present research in ways which are more accessible to a more general audience (Fahnestock, 2004).

Scientific knowledge is also transmitted through interaction between scientists and journalists (Ciapuscio, 2003) or doctors and patients (Gűlich, 2003). In all of these situations, particular accommodations are made. Reformulations are made, for example, by repetition or paraphrase (Ciapuscio, 2003; Gűlich, 2003). Illustration is offered, variously expressed through use of exemplification, scenarios, and metaphor (Gűlich, 2003). Meanings are negotiated between interactants (Ciapuscio, 2003; Gűlich, 2003) and new meanings emerge as scientific claims move between discourse communities (Beacco *et al.*, 2002; Moirand, 2003).

It has long been accepted that, when scientific claims move from their source domain, they undergo particular lexico-grammatical and rhetorical transformations which can imbue them with an authority often lacking in the original text (Myers, 1994). For example, hedges and qualifications tend to disappear, so that claims are expressed with a greater degree of certainty in a popularization than in the original scientific report (Fahnestock, 1986). The tentativeness which is a feature of scientific writing is thus diminished in popularized accounts (Myers, 1994). There may be a shift in the use of metaphors, especially where new concepts or new threats are an issue. For example, the metaphor of decoding a text was drawn on in stories about the decoding of the human genome in the Spanish press (Calsamiglia and van Dijk, 2004). More interestingly, while the use of war metaphors has been a characteristic of discourse about disease and immunity for a century or more, the emergence of the newly discovered SARS virus occasioned the use, in the British press, of new metaphors (Wallis and Nerlich, 2005). The war metaphor was replaced by the conceptual metaphor SARS IS A KILLER, so that the disease was not represented as an army, but as an individual acting alone (ibid.). Finally, the narrative of the scientific process is replaced by a narrative which emphasises the scientist as actor (Myers, 1994) and there is greater emphasis in

popularizations on expressing evaluative assessments about the value of the research (Fahnestock, 1986).

Framing is reflected in the structure of news articles, in which the nucleus (the headline, subheadings and opening paragraph) indicates how the writer intends the reader to interpret the overall message (White, 1997). This represents a particular challenge when using corpus techniques to uncover meaning. The way one interprets a quotation in a news article from, for example, a representative of the DoH saying that there is no evidence of a link between MMR and autism, is likely to vary depending not only on the reporting verb used but on the ideational content expressed in the headline. Corpus methods tend to divorce individual utterances from their wider context. This problem is discussed further in Chapter Three and the methodological considerations are addressed in Chapter Four. Analysing the framing of smaller, cited stretches of text in news articles in corpus data presents less of a challenge, since reporting verbs and expressions referring to sources generally occur relatively close to the node-word. An obvious way of analysing framing is to examine the evaluative meanings encoded in reporting verbs and in the ways in which a source is referred to. Meanings may be encoded in implicit ways, though. Taylor (2010), for example, notes that terms such as *the science*, *(the) research*, *(the) scientists* and *(the) experts* are increasingly used in science stories in the British press in appeals to authority, in place of more specific terms. The expressions *the science* and *the research* are frequently used in the role of Actor in Material processes (e.g. *the science shows*) or Sayer in a Verbal process (e.g. *the science suggests*). In this way they are represented as autonomous entities whose authority is to be trusted (ibid.: 233-238).

Not all features which are typical of scientific writing are transformed in popularizations. In scientific articles aimed at a more general audience, the core argument is generally maintained, even though certain accommodations are made which may occasion a shift in meaning, such as, for example, making causal relations appear more certain (Fahnestock, 2004). Jones (2013: 46-47) points out that nominalization is often carried over into journalistic accounts and comments on the use of the terms *uncertainty* and *links* in the following citation from a news story on MMR in the *Daily Mail*, 2 February, 2002:

> Although health chiefs insist that the MMR vaccine is safe, many parents have been put off by the uncertainty over possible links to autism and bowel disorders.
> (cited in Speer and Lewis, 2004: 174, reproduced in Jones, 2013: 46)

Jones (2013: 46-47) argues that the term *uncertainty* nominalizes the process of public debate which arose from Wakefield's claims and thus conceals agency, while the nominalized form *links*, even though hedged with the word *possible*, represents the causal hypothesis as more certain than it was.

The observation that certain features thought to be typical of scientific writing may be found in popularizations supports the observation that the boundaries between scientific and popularization discourse are fuzzier than once thought (Myers, 2003). There is some evidence that scientific writing might, at times, show features which are more typical of popularized texts. In their examination of four news articles on MMR and two scientific articles (one of them the paper by Wakefield *et al.*, 1998), Rundblad, Chilton and Hunter (2006) found that, although the article by Wakefield *et al.* (1998) conformed in some respects to what one would expect from an academic paper, in other respects it displayed features more typical of news articles. The low frequency of deontic modals used was as expected in a scientific article and

stood in contrast to the use of deontic modals in the news articles. However, it included a lower proportion of epistemic modals than is typical in scientific writing and the proportion of vague or generic (as opposed to specific) references and references to named authors was more in line with what was found in the news articles.

#### 2.3.2.2 New forms of public discourse about science

One can also approach the analysis of popularization from the point of view of discourse. It is posited that a new form of public discourse about science has emerged, characterised by the mixing of different voices (the voice of the scientist, government officials, the journalist or even the public), so that what results is a patchwork of intertextual chunks of language (expressed through direct and indirect quotation) and elements of interdiscursivity (language use characterised by different ways of representing the world). The phenomenon that Boyce (2007) described, whereby the MMR issue was frequently depicted as a political issue, is part of a general trend in media discourse about science. In media reporting of science issues, there is less concern with explaining scientific 'facts' and greater focus on examining the social issues at stake, especially where there is a political dimension to an issue (Moirand, 2003). The journalist typically makes analogies between the issue or event at hand and other issues and events, thus appealing to (and contributing to) an 'interdiscursive memory bank', which is drawn on to frame debates (Beacco *et al.*, 2002; Moirand, 2003). The idea of the interdiscursive memory bank chimes with Bellaby's (2003) suggestion that, to fully understand parents' thinking on the risks of MMR, one needs to consider broader social and historical factors. Analysing interdiscursive relations also offers a potentially fruitful way of approaching analysis of the *JABS* corpus. But the concepts of discourse and interdiscursivity are problematic and raise questions of how one defines *discourse(s)* and how one identifies a

specific discourse. These questions are addressed in Chapter Three. For now, the discussion moves on to the topic of expertise.

### 2.3.3 Expressing expertise and using intertextual references in online forums

Unlike the patients who post questions on healthcare websites set up by health professionals (see Adolphs *et al.*, 2004; Harvey *et al.*, 2007; Atkins and Harvey, 2010; Seale *et al.*, 2010; Harvey, 2012; Harvey, Locher and Mullany, 2013), the parents who post questions on the JABS message board and forum have not set out to gain information from medical experts. They are aware that they are consulting their peers. However, they do so on the implicit understanding that members of their peer group possess a kind of expertise which they value. This reflects a shift in the notion of expertise and a blurring of the boundaries between the expert and the lay-person which is increasingly recognised in popularization studies. The early view of the process of popularization as one in which scientific knowledge is communicated in a simplified form by the 'expert' to a 'lay' person, who possesses no technical expertise at all, has been widely challenged (Hilgartner, 1990; Grundman and Cavaillé, 2000; Myers, 2003). To begin with, 'experts become less expert as soon as they step outside their very limited specialism' (Myers, 2003: 268). Furthermore, 'lay' people may possess considerable technical expertise in a certain field, whether because they have undertaken training or research or simply through lived experience (ibid.: 268-269). They therefore have their own persuasive resources which they can draw on (ibid.). The writers on the JABS website possess a variety of types of expertise. Almost all are parents. Some have the experience of living with a disabled child. Some have medical or scientific expertise. But what must not be overlooked is that, as members of a vaccine-critical group, many JABS members actively seek to enhance their own knowledge and expertise through a variety of

means. In vaccine-critical discourse, a common persuasive strategy is to emphasise the expertise of the individual parent (Hobson-West, 2005).

When people engage in debate, whether face-to-face, for example in focus groups, or online, they adopt particular strategies for claiming an entitlement to speak and establishing credibility. Participants in focus groups often support their claims by referring to specific sources but have also been found to attribute claims to non-specific sources, using, for example, the generic pronouns *you* or *they*, or to refer to hypothetical sources (Myers, 2004: 135-142). In newsgroups and online forums, there is an obvious need to express a warrant for expertise, since the issues at stake are highly technical (Richardson, 2003). Richardson (ibid.) identifies five warranting strategies used in newsgroups. The most common strategy involves reference to sources (mass media; science publications; policy publications; industry statements). In MMR newsgroups, a range of sources were used to support claims, including web pages, publications by bodies such as the World Health Organization (WHO) or the Centers for Disease Control and Prevention (CDC), or lobby groups, but by far the greatest number of references were to mass media publications (Richardson, 2005: 154). References are frequently specific but sometimes vague, for example, 'I saw a report recently …', or generic, for example, references to 'the media' (ibid.: 153). While there is a general reliance on mass media sources, people often express negative evaluations of them (ibid.). Richardson (2005) does not correlate the use of reference to sources with the writer's stance, but Hodson-Champeon (2010) reports that newsgroup posters tend to use a direct quotation or make some other explicit reference to a source when they are rebutting another's claim. When they use reference to support their own claims, they tend not to refer to their sources explicitly. The participants on the JABS forum are not all of one mind, though, and disagreement is common.

Some of the interaction on the JABS forum exhibits some of the features uncovered by Vayreda and Antaki (2011) in their analysis of a forum thread discussing the necessity, or otherwise, to vaccinate against the H1N1 virus. Participants often disqualified the contributions of others by challenging their expertise, challenging the quality of their sources, or on the grounds that they were partial and had interests in the pharmaceutical industry.

Another frequently used warranting strategy is that of drawing upon personal experience (Richardson, 2003: 178-179). This strategy is found frequently in the *JABS* corpus where parents often supply a narrative to allow them a warrant for raising a topic or use their child's experience as evidence in support of an argument. Narratives are diverse in form and content, but keyword analysis can be used in a corpus-assisted study to highlight candidate terms (Sealey, 2010). Likely candidate terms in the JABS corpus are *OCCUR* and *HAPPEN*. Other warranting devices identified by Richardson (2003: 179-180) concern referring to one's status (for example, stating that one is a scientist) and using a technical register, for example, using specialized vocabulary or the sorts of grammatical features typical of scientific writing, such as nominalization, pre-modified noun groups, or higher than average lexical density. Where no explicit warrant is provided, a disclaimer of expertise is often given, such as, *I'm not a scientist or anything special, but …* (ibid.: 181). A disclaimer may indicate that the writer is aware of their own lack of expertise or it may function as a rhetorical device, indicating that technical expertise is not called for to be able to express a valid opinion on a subject. Writers also often challenge others on the lack of sources cited or the credibility of the sources they have used and occasionally respond to challenges by defending their sources (ibid.: 182-183). Strategies such as these are powerful persuasive resources and are evident in the *JABS* data.

### 2.3.4 Conclusion to Section 2.3

- Public debate today is structured more around group identities and shard issues than it was in the past and the MMR debate, and the activities of the JABS group, is typical of such a change;
- It is increasingly recognised that 'lay' people have their own forms of expertise (often rooted in personal experience) and their own discursive resources on which they can draw;
- Although the boundaries between scientific and popularized genres are increasingly recognized as fuzzy, it is still acknowledged that when scientific claims move from their source domain and are reproduced in other genres, they undergo shifts in meaning, expressed through particular lexico-grammatical and rhetorical changes: such shifts often result in making claims seem more certain than they are in their original form; however, nominalization often carries over from the source domain;
- Public discourse about science has undergone discursive changes in recent years so that issues are framed more as political or public policy issues and scientific detail is backgrounded; public discourse about science draws on and contributes to an interdiscursive memory bank;
- Forum interactants establish credibility by using particular warranting strategies, such as drawing on personal experience, drawing on technical expertise (for example, using features of scientific discourse) or referring to sources to support their claims; participants in online newsgroups were found to rely particularly on mass media sources to support claims.

## 2.4 Conclusion

This chapter has reviewed literature on the expression of attitudes towards risk and the MMR vaccine, as well as considering ways in which notions of risk and causation are expressed in discourse. It has looked at the strategies used by vaccine-critical groups. It has also looked at theories concerning scientific popularization. It has brought to light the fact that long-held distinctions between experts and non-experts do not always stand up to scrutiny. Lay people have a degree of expertise and particular persuasive resources on which to draw. In the case of discussion of risk in online discussion forums, these include the use of certain warranting devices and intertextual resources. Vaccine-critical groups, in particular, emphasize the expertise of the parent. It has also been found that one cannot easily categorize ways of talking about risk into a binary division between lay and expert, or technical, talk. Instead, the meanings of *risk* shift according to context. Similarly, there are no clear boundaries between scientific and popularized or lay discourse. While there are certain distinctions, some lexico-grammatical patterns typical of scientific discourse carry over into popularizations. However, it is suggested that vaccine-critical groups exploit scientific discourse and technical risk discourse. They also make strategic use of the connection between risk and trust. Having considered research into MMR, popularization discourse, the next chapter discusses the principles of and approaches to corpus linguistics and discourse analysis. It also discusses approaches to examining evaluation in discourse.

# CHAPTER THREE

# CORPUS LINGUISTICS AND DISCOURSE ANALYSIS: PRINCIPLES AND PRACTICES

## 3.1 Introduction

It was explained in Chapter One that the four main research questions guiding this project focus on discovering:

- how beliefs about health, the immune system and risk are expressed in the *JABS* corpus (and how this compares with similar expressions in the *NHSvax* corpus);
- how discursive objects and sources of authority are evaluated in both corpora;
- to what extent, and how, strategic use is made of particular discourses;
- and how the JABS discussion forum participants express their warrant for expertise.

To answer these questions, the objectives of the project are to identify the typical lexico-grammatical patterns used in each corpus to express claims and to realise attitudinal meaning, and to identify common discursive strategies and the functions they fulfil. As outlined in Chapter One, this project adopts a Corpus Assisted Discourse Studies (CADS) approach to textual analysis, an approach which marries discourse analysis with corpus linguistics methods (Partington, 2004a). *Discourse* is a contested term with a variety of definitions. There are also different approaches to discourse analysis. But there is not simply one approach to corpus linguistics either. We can distinguish two broad traditions. In one, corpus linguistics is viewed merely as a method of textual analysis. In the other, the *Sinclairean* or *neo-Firthian* tradition (e.g. Sinclair, Jones and Daley, 2004; Sinclair, 1991; 2004), the method

of analysis is closely tied to a specific theory of meaning, in fact, some scholars in this tradition go so far as to say that corpus linguistics is a theory (notably, Tognini-Bonelli, 2001). The theoretical stance one adopts understandably impacts on the methods one uses. Finally, there is a variety of ways of approaching the study of attitudinal meaning.

The aim of this chapter is to clarify the theories of language and discourse which underpin and inform this thesis. Section 3.2 discusses the reasons for adopting a CADS approach and weighs up its strengths and limitations. Section 3.3 problematizes the notion of discourse and evaluates different approaches to discourse analysis, putting forward an argument for adopting a realist rather than constructionist approach. Section 3.4 evaluates approaches to corpus linguistics. Section 3.5 evaluates approaches to the analysis of attitudinal meaning in language. Section 3.6 explains the approach to CADS adopted in this thesis. Section 3.7 concludes.

## 3.2 Corpus assisted discourse studies

### 3.2.1 The advantages of using a CADS approach

The decision to use a CADS approach to analyse the data in this study was based on the fact that, as Hardt-Mautner (1995) demonstrates, corpus linguistic methods provide a way for the discourse analyst to make reliable observations about language use on the basis of a large body of texts. I use the term *CADS* in its broad sense, that is, I take it to refer to any and all studies which combine the quantitative methods of corpus linguistics with the qualitative approach of discourse analysis. These include the pioneering studies of Stubbs and Gerbig (1993), Hardt-Mautner (1995) and Krishnamurthy (1996), subsequent work in a similar vein (e.g. Stubbs, 2002; Baker, 2004; 2012; Orpin, 2005; Baker and McEnery, 2005; Bastow,

2006; O'Halloran, 2007; Baker *et al.*, 2008; Caldas-Coulthard and Moon, 2010; Grundmann and Krishnamurthy, 2010; Jaworska and Krishnamurthy, 2012; Sealey, 2010; 2012; Sealey and Oakley, 2014), corpus-based studies into healthcare interaction in online settings (Adolphs *et al.*, 2004; Harvey *et al.*, 2007; Atkins and Harvey, 2010; Seale *et al.*, 2010; Harvey, 2012; Harvey, Locher and Mullany, 2013), and those studies by the group of academics who actively identify themselves as CADS scholars (Partington, 2003; 2004a; 2008; 2010; Clark, 2010; Duguid, 2010a; 2010b; Marchi, 2010; Taylor, 2010; Freake, Gentil and Sheyholislami, 2011; Freake, 2012; Partington, Duguid and Taylor, 2013).

CADS developed originally in response to perceived methodological weaknesses in Critical Discourse Analysis (CDA) (e.g. Fairclough, 1989; 1992; 1993; 1995a; 1995b; 2000; van Dijk, 1991; 1993; 2003; 1997a; 1997b; Wodak, 1996; Wodak and Riesigl, 1999; 2001; Wodak *et al.*, 1999/2009; Riesigl and Wodak, 2001; Fairclough and Fairclough, 2011), although many of the benefits it brings to CDA hold true for other versions of discourse analysis. Corpus linguistic methods help overcome the limitations inherent in the qualitative approach of CDA, which render it unsuitable for analysing all but one or a few texts (Hardt-Mautner, 1995). It is also argued that corpus methods can lend CDA 'a firmer descriptive and methodological basis' (Stubbs, 1996: 125-126). Firstly, the inductive approach of corpus linguistics, exemplified in particular in the Sinclairean tradition of corpus analysis (see Tognini-Bonelli, 2001; Teubert, 2005), means that CADS practitioners are less open than CD analysts to the criticism that they use analysis to reinforce their preconceptions (Sharrock and Anderson, 1981; Widdowson, 1995a; 1995b; 1996) and of producing a circular argument as a result (Stubbs, 1997).[21] CADS aims at uncovering '*non-obvious meaning* – that is, meaning which

[21] The contention here is that critical discourse analysts look in a text for a few selected features which they deem ideologically significant and, on finding them, claim that these examples provide empirical evidence to

might not be readily available to perusal with the naked eye' (Partington, 2010: 88) and one of the strengths of corpus analysis is that it makes apparent recurrent patterns of language use which cannot be uncovered through native–speaker intuition alone. Secondly, CADS is essentially comparative in its approach: the patterns observed in the discourse type under consideration are compared with a norm (Partington, 2008: 99-100). One of the criticisms of CDA is that claims of a quantitative and comparative nature are often made with insufficient empirical evidence to support them (Stubbs, 1997). The example Stubbs (ibid.) gives is Fairclough's (1995a) claim that an increasingly frequent feature of public discourse is the mixing of genres and discourses (for example, the use of conversational styles of speech in more formal contexts). Such a claim can only be adequately supported by means of a comparison of recurrent features found in a large sample of texts from two distinct time periods, Stubbs (1997) argues. The current study is interested in synchronic rather than diachronic comparison, but the point Stubbs (ibid.) makes is relevant. As was discussed in Chapter Two, one of the key features of vaccine-critical discourse concerns the mixing of scientific and other discourses (Hobson-West, 2005). The mixing of different voices and discourses has also been found to be a feature of media discourse about science (Beacco *et al.*, 2002; Moirand, 2003). Comparison of data sets is therefore central to the current study. This, and other methodological issues, is discussed in the next sub-section.

---

support their interpretation. An oft-quoted example is the frequent focus on the agentless passive, used as evidence that agency has been deliberately concealed in a text. Although CDA practitioners acknowledge that one cannot assume a one-to-one correlation between form and function (see Fairclough, 1995a: 71) and accept that, in some cases, a passive is used for other purposes, Widdowson (2004: 96) argues that they tend to ignore this fact. Stubbs (1997: 112-113) comments on the circularity of an analytical method predicated on the supposition that a feature has ideological significance except when it does not.

### 3.2.2 Theoretical and methodological approaches in CADS

CADS projects typically involve the compilation of a specialized corpus and, often, a general reference corpus or a second specialized corpus is used for comparison. Sometimes both are used. Comparison with a general reference corpus helps highlight the distinct features which characterise the discourse type under consideration. Partington (2003) and Bastow (2006), for example, use a general reference corpus in order to highlight statistically significant keywords in their corpora of, respectively, White House press briefings and US Department of Defence speeches. Baker and McEnery (2005) compare the representations of refugees and asylum seekers in their corpus of British news articles with those in a specially compiled corpus of United Nations High Commissioner for Refugees (UNHCR) texts on the grounds that a general corpus of English, such as the *BNC*, has such a diverse range of texts that it was deemed unlikely that it would enable the researchers to more easily distinguish between hegemonic and resistant discourses. They use the *BNC* in their study as a means of establishing language norms against which they compare their concordance data. The current study follows Baker and McEnery's (ibid.) example and uses the specially compiled *NHSvax* corpus to highlight the typical features of mainstream medical discourse related to vaccination. This will make it possible to distinguish mainstream medical discourse from discourses of resistance in the *JABS* corpus. Following Partington (2003) and Bastow (2006), the *BNC* is used to highlight statistically salient keywords in each.

Although CADS arose in response to CDA, not all CADS studies adopt a CDA framework. While many explicitly appeal to CDA (e.g. Hardt-Mautner, 1995; Sotillo and Wang-Gempp, 2004; Baker and McEnery, 2005; Mautner, 2005; 2007; 2009; Orpin, 2005; Coffin and O'Halloran, 2006; O'Halloran, 2007; 2009; Baker *et al.*, 2008; Alexander, 2010; Caldas-

Coulthard and Moon, 2010; Baker, 2012; Sealey, 2012; McDonald and Hunter, 2013; Wild *et al.*, 2013), a similar number do not (e.g. Krishnamurthy, 1996; Stubbs, 2002; Partington, 2003; Baker, 2004; Cook, Pieri and Robbins, 2004; Gales, 2009; Ensslin and Johnson, 2006; Duguid, 2010a; 2010b; Grundmann and Krishnamurthy, 2010; Marchi, 2010; Sealey, 2010; Taylor, 2010; Jaworska and Krishnamurthy, 2012; Prentice, Rayson and Taylor, 2012; Taylor, 2013; Sealey and Oakley, 2014). Many of those which do align themselves with CDA, however, do not draw extensively on the CDA framework. Baker and McEnery (2005), for example, explain that they see their work as making a contribution to CDA but focus entirely on the collocational patterns of specific lexical items and the conceptual metaphors such patterns realise. Sealey's (2012) aim is the same, but the writer appears to adopt a standpoint from outside of CDA, using the study in a certain respect as a critique of their theoretical approach. In any case, as (Baker, 2004: 89) points out, there is considerable overlap between CDA and critically-oriented discourse analysis in its broader sense.

All CADS studies, whether taking a CDA or a broader discourse analysis approach, are essentially concerned with uncovering ideological meaning. The current study does not employ a CDA framework, largely because the overtly political intent of CDA is not appropriate to the research aims. The avowed aims of CDA are to uncover 'the way social power abuse, dominance, and inequality are enacted, reproduced, and resisted by text and talk in the social and political context' (van Dijk, 2003: 352). It also aims to perform a kind of political action. As van Dijk (ibid.) explains, critical discourse analysts want to 'understand, expose, and ultimately resist social inequality'. The aim of this project is not to uncover and resist inequality, but to understand how people who hold a certain point of view articulate their arguments. However, the project is also concerned, to a certain extent, with the

expression of ideology. After all, vaccine-critical discourse represents a form of resistance to the hegemonic discourse of the state and the medical-scientific community. CADS studies which adopt a broad discourse analysis approach (and those which make limited use of CDA frameworks) demonstrate that it is possible to make insightful observations about the ideological meaning of texts without recourse to the highly complex models of context employed in CDA.

CADS studies which do not make extensive use of CDA (or which make no use of it) tend to inform their interpretation by appeal to whichever aspects of disciplinary, or critical and cultural theory they consider relevant. Some draw on findings from relevant cognate disciplines to inform their analyses, such as, for example, healthcare research (Adolphs *et al.*, 2004; Harvey *et al.*, 2007; Seale *et al.*, 2010; Harvey, 2012; Harvey, Locher and Mullany, 2013). Others draw on the Foucauldian notion of discourse (e.g. Baker, 2004; Ensslin and Johnson, 2006) and the Bakhtinian or Kristevan concepts of intertextuality and interdiscursivity (Baker and McEnery, 2005; Ensslin and Johnson, 2006). The Foucauldian notion of discourse is notoriously complex and is interpreted in different ways. Intertextuality and interdiscursivity is similarly complex. Since these concepts are relevant to the current study, they are discussed in full in Section 3.3.

### 3.2.3 CADS and considerations of context

By using corpus linguistic methods, the CADS approach makes it possible to uncover ideological meaning whilst avoiding some of the problems inherent in qualitative approaches to discourse analysis. However, corpus linguistic methods have attracted some criticism, mainly from Widdowson (2000; 2004). While he concedes that corpus linguistics enables the

analysis of large quantities of data, he states that a weakness is that it 'focuses on items in isolation from their co-textual dependencies' (Widdowson, 2004: 120). He also states that it studies text as a product rather than as the process of discourse production and that considerations of textual interpretation are overlooked (Widdowson, 2000). It is true that corpus linguistic methods effectively strip texts from the situational context in which they occurred. However, that does not mean that the corpus linguist has no grounds on which to base claims about interpretation. As Stubbs (2001) argues in response to Widdowson (2000), by highlighting recurrent patterns in language use, corpus linguistics methods help establish what is typical usage so that the researcher has a reliable means of identifying unusual or divergent uses (Stubbs, 2001: 151). Nor is it the case that context-dependent pragmatic meaning is hard to uncover using corpus methods. A lot of work in corpus analysis has revealed that pragmatic meaning is frequently encoded in recurrent lexico-grammatical patterns (ibid.: 153).

Widdowson's (2000; 2004) criticisms are perhaps more pertinent to traditional corpus linguistics than to CADS. The highly specific aspects of context that Widdowson (ibid.) has in mind are not particularly relevant to the needs of corpus linguistics conducted for the purposes of general language description. As (Partington, 2010: 89) explains, early corpus linguists tended to treat the corpus as a 'black box'. CADS scholars, on the other hand, aim to 'acquaint [themselves] as much as possible with the discourse-types under investigation' (ibid.: 89-90). Having compiled their own specialized corpora to suit the needs of their research question, CADS researchers frequently enhance their own understanding of the discourse type by, for example, reading all, or parts of, the data and by consulting other sources of information, or comparing one set of data with another (Partington, 2010: 90).

Importantly, a consideration of contextual factors plays a significant role in many CADS studies. For example, Partington's (2003) study of White House press briefings begins with a thorough account of the context of situation. The analysis then takes into account the impact on the text of the differing interactive goals of the two groups involved: the journalists and 'spin doctors'. Baker and McEnery (2005) similarly take into account the different communicative goals of the two text types under consideration (news articles and UNHCR documents) and the intertextual relations between each. However, they concede that the breadth of ground that a corpus approach enables the researcher to cover may mean that particular subtleties are overlooked (Baker and McEnery, 2005: 223). In the current study, the differing goals of the JABS and NHS websites are an important factor in informing analysis. Understanding was also enhanced because it has been possible throughout the project to visit the JABS website and, until the summer of 2010, the 'MMR Thefacts' and 'NHS immunisation' websites were still functional.

### 3.2.4 Conclusion to Section 3.2

This section has argued that using corpus methods of textual analysis have particular advantages over adopting the traditional qualitative approach of discourse analysis. It has also acknowledged that the fact that the concordance separates stretches of text from their wider contextual environment poses a challenge in the context of this study. The ways in which aspects of the theories and methods of discourse analysis and CDA are employed in CADS has also been considered. Since discourse is the focus of this study, the next section considers the different ways in which discourse is conceptualised in different branches of discourse analysis. In the subsequent section, the key principles underpinning the approach to corpus

linguistics adopted in this study are discussed, as are key concepts such as the lexical item, collocation, colligation, semantic preference and semantic prosody.

## 3.3 Discourse

### 3.3.1 Textlinguistic definitions of *discourse*

As mentioned in Chapter One, *discourse* is a much used term with a number of meanings. In this sub-section, I aim to examine the ways in which it is used in text linguistics and to explain the rationale for using the terms *discourse* and *text* in the way they are in this thesis. In the following sub-section, I will discuss theories from text linguistics of relevance to this study. *Discourse,* as conceptualised in the textlinguistic tradition, can be defined as 'language in use' (Brown and Yule, 1983: 1) or 'language above the sentence or above the clause' (Stubbs, 1983: 1). While these definitions appear adequate, they are problematic. The first can be criticised for being too broad, since it fails to give any indication of what aspects of language use are implied. The second, although neatly encapsulating the way discourse is envisaged in text linguistics, has been criticised for being too narrow. Widdowson (2004: 6-7) questions the necessity for *discourse* to be defined as dealing only with supra-segmental units of language, since sub-clausal units can be meaningful in certain contexts. He gives as examples the way in which we easily interpret the intended meanings of public notices bearing words such as *open* or *closed* or the letter *P* to indicate a place where parking is permitted (ibid.). This is a fair criticism in some respects and particularly pertinent to the current project, which is concerned with examining how meaning is negotiated between readers and writers. However, it overlooks the fact that Stubbs (1983) crafted this particular definition in the context of a work whose aim is to shed light on the macro-organisational features of texts. What he has in mind are the sorts of organisational features of extended

stretches of interaction, whether spoken interaction, as described by conversation analysts (e.g. Sacks, Schegloff and Jefferson, 1974) and discourse analysts (e.g. Sinclair and Coulthard, 1975), or written interaction, as described by discourse analysts concerned with the ways in which cohesion and coherence is achieved in texts (e.g. Halliday and Hasan, 1976; Winter, 1977; Hoey, 1983; 1991; 1994; Tadros, 1985; Francis, 1986). It is discourse in this last sense which is of particular relevance in this project, which aims to discover how coherent and persuasive arguments are constructed in the *JABS* corpus.

A further problem with the way in which *discourse* is used in text linguistic literature concerns an unfortunate overlap with use of the word *text*. In some works (e.g. Hoey, 1983; 1991; Stubbs, 1996; 2002), little or no distinction is made between *discourse* and *text*. In others (e.g. van Dijk, 1997a; 1997b), *discourse* is used to refer to spoken language and *text* to written language, although this practice has now generally fallen out of use (Partington, Duguid and Taylor, 2013: 2). Elsewhere, *text* refers to actual instances of language use and *discourse* to 'the situated use of text' (Georgakopoulou and Goutsos, 1997: 4; Verschueren, 1999: 50). Stubbs (1996: 4) claims that there is often no good reason for drawing a distinction between *discourse* and *text*. But Widdowson (2004: 8) argues that it is crucial to make a clear distinction between *text* and *discourse*. Meaning arises in a social context and is activated through the interplay of context and the linguistic code. Discourse, he suggests, is 'the pragmatic process of meaning negotiation', text its 'product' (ibid.). Widdowson's (ibid.) argument is persuasive and the distinction he draws between *discourse* and *text* is a useful one. *Discourse* in this thesis is therefore used to refer to the pragmatic aspects of communication in context and to language use above the level of the clause or sentence. *Text* is used to refer to instances of language use.

### 3.3.2 The Foucauldian notion of discourse

#### 3.3.2.1 Foucault's definition of *discourse*

One of the questions this project aims to answer is whether, and how, strategic use is made of particular 'discourses' in the *JABS*. It is important to clarify precisely how *discourse(s)*, in the Foucauldian sense, is understood for the purposes of this project, since one's interpretation of it, in particular the way one views the relationship between discourse and society, directly affects the approach to analysis and the interpretation of the data. Michel Foucault's notion of discourse has been particularly influential in cultural and critical theory in recent decades. Unfortunately, his theories are open to diverse readings, not least on account of the fact that he uses the term inconsistently and his concept of discourse (and discourses) changes over the course of his writings (Martin Rojo and Galibondo Pujol, 2000: 2). As Foucault himself observes:

> I wonder whether I have not changed direction on the way, […] whether, while analysing 'objects' or 'concepts', let alone 'strategies', I was in fact still speaking of 'statements' […] instead of gradually reducing the rather fluctuating meaning of the word 'discourse', I believe that I have in fact added to its meanings: treating it sometimes as the general domain of all statements, sometimes as an individualizable group of statements, and sometimes as a regulated practice that accounts for a certain number of statements.
>
> (Foucault, 2002: 90)

What Foucault refers to as 'statements' might more usefully be referred to as propositions. Foucault uses *discourse* as an uncountable noun to refer to the expression of all or any propositions, in other words, to language use in general. But it is the ways in which he uses *discourse* as a countable noun which has had a profound influence on discourse studies of a critical nature. For Foucault, discourses are systems of propositions which consist of and construct objects of knowledge, concepts, social identities and social relations (Fairclough,

1992: 39). He sees discourse as a form of social practice and discourses as ‘practices that systematically form the objects of which they speak’(Martin Rojo and Galibondo Pujol, 2000: 3-4). Discourses also construct areas of knowledge so that, to a certain extent, they correspond to disciplines (Foucault, 2000: 236), although disciplines and discourses are not reducible to each other since disciplines ‘constitute a system of control in the production of discourse, fixing its limits through the action of an identity taking the form of a permanent reactivation of the rules’ (Foucault, 2000: 237).

#### 3.3.2.2 Interpretations of the Foucauldian notion of discourse

The sheer complexity of Foucault’s theories and the plurality of meanings he gives to *discourse(s)* lead to a variety of interpretations of his ideas. His theories are often interpreted in radically post-modernist or contructionist ways (e.g. Potter, 1996; Wetherell, Taylor and Yates, 2001; Baxter, 2002), which privilege Foucault’s view that all aspects of social life are constructed by discourses, or constructivist-structuralist ways, ‘concerned with the constraining role of social structures as well as with the active process of the production of social practices which can transform social structures’ (Martin Rojo and Galibondo Pujol, 2000). The constructivist-structuralist aspects of Foucault’s theories have been criticised for failing to attend to the concept of agency and for fostering an impoverished view of human conduct (Wooffitt, 2005: 179) and leaving an overall impression of ‘people being helplessly subjected to immovable systems of power’ (Fairclough, 1992: 57).

The idea that discourses construct social life is more persuasive and deserves closer attention. There are indeed certain concepts which are socially constructed and which we encounter (and construct) through discourse. Teubert (2013) gives the example of a group of students

discussing *body image*. He shows how meaning is negotiated and the concept discursively constructed as the students engage in interaction, supplying definitions (sometimes intertextually linked) and explanations using terms they deem relevant (for example, *self-esteem, self-confidence, self respect, attractive, internal feelings*) and endorsing or rejecting each other's suggestions. One can agree that a concept such as body image is culturally determined and thus socially constructed. However, he goes on to comment:

> We do not know why participants said what they said, and why some items, but not others, were seconded. **The discourse has put itself in charge**[22] of what should be added in terms of knowledge to what was already known about 'body image'.
>
> (Teubert, 2013: 293)

Thus discourse is represented as an autonomous object independent of the people who utter the words. This is a common notion among writers who accept a constructionist view of language. We see this in particular in studies which talk about *competing* (e.g. Lee, 1992; Baxter, 2002) or *conflicting discourses* (e.g. Coupland and Williams, 2002), which construe conflict between groups or resistance to authority in terms of competition or conflict between discourses. But the habit of reifying discourses as 'autonomous collusive actors' (Wodak and Riesigl, 2001: 383) eliminates any real sense of social agency so that the social world 'is seen as constructed by authorless discourses which themselves become agents' (Sealey and Carter, 2004: 47). In some constructionist accounts, the idea that discourse constructs reality is even extended to physical objects. Teubert (2013: 276), for example, argues that concepts such as 'cats' or 'gardens' have no basis in reality. There is a strong argument for claiming that the categories we divide the world into are socially constructed, however, as Sealey (2014) points out, the taxonomies we construct often arise in response to our experience of the world and sometimes correspond to real-world phenomena. The claim that humans and chimpanzees are

[22] My emphasis.

'cousins' is metaphorical and says something about the way we categorise the world. But at the same time, in material terms,

> Humans and other primates are scientifically classified as 'related' in two main ways: having relatively recently shared a common ancestor and continuing to share a high proportion of DNA.
>
> (Sealey, 2014: 315)

There are phenomena, Sealey (ibid.: 311) argues, which 'our empirical senses cannot perceive, but that are nevertheless real … and not reducible to the labels we invent in discourse'. The adoption of a radical interpretation of Foucault's notion of discourse is therefore ultimately untenable. However, it is true to say that certain concepts are socially constructed, as Teubert's (2013) example of *self-esteem* illustrates. As Sealey and Carter (2004: 47) argue, interpreted in its less radical form, the Foucauldian notion of discourse has useful things to say about the way particular linguistic choices can contribute to different forms of expression.

The notion of discourse (in its Foucauldian sense) adopted in this study is realist rather than constructionist. Discourse emerges from our interaction with the world: we draw on linguistic resources, using them creatively to name and categorise the phenomena we experience and to effect action in the world (Sealey, 2014). The argument for adopting a realist perspective rather than a constructionist one may seem purely academic, but it has implications for the ways in which I interpret what I see in the data and how I express my interpretations. Let us take, for example, a mother's claim that her child has suffered vaccine damage (there are many such claims in the *JABS* corpus data). If one interprets such a claim from a

constructionist perspective, one might state that the mother has ‘constructed’ her child as vaccine-damaged. From a realist perspective, on the other hand, one infers that the child probably displays symptoms of neurological impairment which probably have a pathological cause. One also infers that the alleged impairment may, or may not, have been caused by a vaccine. One therefore states that the mother has ‘expressed’ the belief (or the proposition) that her child is vaccine-damaged. The constructionist view does not allow for the fact that the condition which the mother reports might have a basis in fact. The realist view allows for the idea that the proposition expressed by the mother might be aligned with reality.

From the point of view of the aims of this thesis, it is useful to adopt a weak version of Foucault’s notion of discourse and to use his idea of discourses as corresponding to particular areas of knowledge or disciplines, or, more usefully, of particular ways of talking about certain phenomena. If it is the case that vaccine-critical groups often exploit ‘scientific discourse’ or particular ‘discourses of risk’ in their arguments, such discourses should be recognisable from the language used. Furthermore, although the notion of discourses as autonomous actors is rejected in this thesis, the notion of dominant discourses is nonetheless useful. Dominant discourses reflect dominant ideologies. Dominant ideologies are those which are associated with powerful social groups and which are ‘mediated through powerful political and social institutions like the government, the law and the medical profession’ (Simpson, 1993: 9). Medical-scientific discourse can thus be viewed as a dominant discourse. We turn now to the question of how to define *discourse* so that we are clear how we might recognise a particular discourse.

### 3.3.2.3 Towards a working definition of *discourse(s)*

As a result of the relative lack of clarity in Foucault's writings as to the precise nature of discourse, it is not unusual for critical linguists who adopt Foucault's concept of discourse to elaborate on his definitions in ways which draw on their understandings of how language works and how ideology and language intersect. It is important to this study to define *discourse(s)* in a way which makes clear how discourses are realised linguistically. Some scholars define *discourse* in a highly abstract way, focusing on language as a semiotic activity but giving little idea how discrete discourses might be identified.[23] The following two definitions, by contrast, are clearer in this respect:

> …recurrent phrases and conventional ways of talking which circulate in the social world, and which form a constellation of repeated meanings.
>
> (Stubbs, 1996: 158)

> a set of meanings, metaphors, representations, images, stories, statements and so on that in some way together produce a particular version of events.
>
> (Burr, 1995, cited in Baker and McEnery, 2005: 198)

We can therefore define *discourses,* for the purposes of this project, as sets of meanings which are realised through recurrent lexico-grammatical patterns, metaphors and other forms of representations (such as narratives), and which circulate in the discursive world. Discourses may be recognised through the repeated use of certain lexico-grammatical patterns or certain metaphor systems (Baker and McEnery, 2005).

[23] Examples are definitions such as the following: 'a general mode of semiosis, i.e. meaningful symbolic behaviour … language-in-action' (Blommaert, 2005: 2); '… a complex bundle of simultaneous and sequential interrelated linguistic acts which manifest themselves within and across the social fields of action as thematically interrelated semiotic (oral or written) tokens that belong to specific semiotic types (genres)' (Wodak and Riesigl, 2001: 383).

### 3.3.3 Intertextuality and interdiscursivity

One of the objectives of the thesis is to discover how intertextual and interdiscursive resources are exploited in the *JABS* corpus for rhetorical effect. As explained in Chapter One, *manifest intertextuality* (or *explicit intertextuality*) is the phenomenon whereby a stretch of one text is incorporated in another (reported speech, for example), while *interdiscursivity* is the phenomenon whereby a text, or stretch of text, is made up of a mix of discourses or discourse conventions (Fairclough, 1992: 104). The sorts of manifest intertextuality characteristic of internet-based newsgroup and discussion forum interaction were discussed in Chapter Two. To reiterate, these involve the inclusion of whole texts in a post, the inclusion of hypertextual links, and the repetition of parts of prior posts (Richardson, 2001). A reliance on mass media sources is another characteristic (ibid.).

Intertextual references of this kind can pose particular challenges when one adopts a corpus linguistic approach to analysis. The main challenge concerns the reliance in the *JABS* corpus data on mass media sources and the inclusion of whole texts in forum posts. A concordance line effectively divorces a stretch of text from its wider context. For example, when looking at a concordance line which represents reported speech and originates in a news article, it may be easy to see which reporting verb has been used but it is harder to access the nucleus of the article (the headline, sub-headings and opening paragraph) which frame the news story (White, 1997). With almost 2,000 forum threads represented in the *JABS* corpus data, however, checking each thread manually would be labour-intensive and unfeasible, given the time constraints of the project. The repetition of parts of previous messages represents a challenge insofar as the concordance may obscure whether instances of repeated patterns are due to repetition of prior utterances in the thread or whether they represent instances of

interdiscursivity. This can be overcome through close inspection of concordance lines. One could argue, of course, that it is immaterial as to why a particular stretch of text is repeated. The very act of repetition indicates that the topic is considered particularly salient by the participants on the forum. Methods for addressing these particular challenges are discussed in Chapter Four.

Interdiscursivity is often approached in discourse analysis from the point of view of *discoursal hybridity* or *hybrid genres* (Bakhtin, 1981). It is claimed that hybridity is a growing feature of discursive life in late modern society and is connected to the weakening of social boundaries (Fairclough, 1992; Chouliaraki and Fairclough, 1999). Hybridity, it is claimed, can be used strategically to 'construct' expertise (Candlin, 2005) or to make an argument more persuasive or appealing to an audience (Chouliaraki and Fairclough, 1999). The problem with the notion of hybrid genres or discourses is that it presupposes the prior existence of 'pure' genres or discourses. The notion of discoursal hybridity also presupposes the reification of discourses (Hasan, 2000). A more convincing argument is that speakers do not consciously merge one discourse with another, instead they move seamlessly between registers (ibid.). Either way, interdiscursivity involves the mixing of patterns of language typical of different domains of use. Comparison of two specialized corpora, by making it possible to identify the patterns of language typical of a particular domain, can help identify instances of interdiscursivity. However, it is important to compare any perceived similarities with norms in the language as a whole, in order to overcome the weakness of many approaches to studying intertextuality and interdiscursivity. Widdowson (2004: 147-148), for example, questions the efficacy of the methods used in CDA for analysing intertextuality and interdiscursivity. He cites Chouliaraki and Fairclough (1999), who define *intertextuality* as,

> the property texts have of being full of snatches of other texts, which may be explicitly demarcated and merged in and which the text may assimilate, contradict, ironically echo and so forth.
>
> (Chouliaraki and Fairclough, 1999:199, cited in Widdowson, 2004: 147)

Widdowson (2004: 148) argues that, defined in this way, intertextuality is hard to trace. Sometimes, he says, it is clear when the writer of one text draws on the words of another, but sometimes it is not clear. Since all texts are composed of regularly occurring patterns, as corpus analysis reveals, it can be difficult to distinguish between a 'snatch' of another text and an example of a pattern which occurs frequently in other texts simply because of the ways in which language is patterned.

### 3.3.4 Summary of Section 3.3

This section has highlighted the following points:

- It is important to distinguish between *discourse* and *text*. Following Widdowson (2004: 8), *discourse* refers to 'the pragmatic process of meaning negotiation', *text* to its 'product';
- *Discourse* is also usefully understood as referring to the ways in which cohesion and coherence is achieved in texts;
- Interpreted in its 'weak' form, and viewed in terms of a realist understanding of the relationship between language and society, the Foucauldian notion of *discourse* offers a useful way of describing some of the ways in which people exploit discursive resources;
- Discourses can be identified by the recurrent use of lexico-grammatical patterns which together express a particular ideological viewpoint;

- Certain features of intertextuality specific to online interaction pose particular challenges when using a corpus-assisted approach.

Having addressed issues related to the notion of discourse, the discussion moves on to the principles of corpus linguistics which guide the analysis in this thesis.

## 3.4 Principles of Corpus Linguistics

### 3.4.1 Corpus as theory or method

Having surveyed CADS literature and looked at the way in which *discourse* is conceptualized, we now turn to the principles and theories of corpus linguistics which guide the analysis undertaken in this project. The theory of language espoused by a researcher influences the methodological approach he or she adopts and this, in turn, influences the kinds of observations which can be made and the inferences drawn from them. The current study is sympathetic to the *Sinclairean* tradition (e.g. Sinclair, 1991; 2004; Francis, 1993; Louw, 1993; Stubbs, 1996; 2002; Hunston and Francis, 1999; Tognini-Bonelli, 2001; Hunston, 2002; 2011; Sinclair, Jones and Daley, 2004; Teubert, 2005; Barnbrook, Mason and Krishnamurthy, 2013; Hanks, 2013), which sees corpus linguistics as firmly rooted in the Firthian theory of language, if not a theory in its own right. This contrasts with the tradition which views corpus linguistics simply as a method of analysis against which pre-existing theories of language are tested (e.g. McEnery and Wilson, 2001; Bowker and Pearson, 2002; Meyer, 2002; Baker, 2006; Scott and Tribble, 2006; Gries, 2010). It is important to examine the differences between the two approaches because they have theoretical and methodological implications for this study.

The argument for claiming theoretical status is that corpus linguistics is best conducted as an inductive process whereby pre-existing theories of language are eschewed and the analyst forms theories on the basis of empirical observation (Tognini-Bonelli, 2001; Teubert, 2005). By applying a pre-existing grammatical model to corpus analysis, one risks overlooking potentially significant findings. Of course, one need not go as far as Tognini-Bonelli (2001: 84) does when she claims that the theory does not exist independently of the evidence. This is a somewhat extreme position which is ultimately untenable since it collapses the distinction between theory and data (McEnery and Hardie, 2012: 148). A less radical interpretation is that theory emerges from observation. However, even this moderate view is challenged by McEnery and Hardie (2012: 148-149). Data, they say, is not self-explanatory and requires the analyst to theorize; nor is it possible or desirable to approach analysis without pre-existing theories of language. To support their position, they refer to studies which have successfully combined corpus methods with other theoretical frameworks, such as Baker *et al*'s (2008) study which draws extensively on CDA and Deignan's (2005) work which uses corpus evidence to test aspects of conceptual metaphor theory. McEnery and Hardie's (2012) criticism is valid in many respects, but their description of the Sinclairean approach is not a wholly accurate reflection of it. Teubert (2005: 4), for example, explains that, in Sinclairean corpus linguistics (or 'his version' of corpus linguistics), traditional categories of linguistic description (word class, for example) are made use of but are open to question. We see the advantages of this approach in Sinclair's (1991: 81-98) analysis of the word *of*. *Of*, it transpires, does not behave at all like a preposition, in that its primary use is not as the head of prepositional phrases. Instead, it is used to create elaborations of nominal groups (e.g. *this kind of problem*; *one of my oldest friends*); it occurs in fixed expressions (e.g. *of course*; *in spite of*); and follows particular adjectives (e.g. *capable of*) or verb forms (e.g. *made up of*)

(ibid.: 85-86). On the basis of these observations, Sinclair (1991) theorizes that *of* rightly belongs in a grammatical category of its own.

### 3.4.2 Corpus-driven versus corpus-based approaches

The argument as to whether or not corpus linguistics constitutes a theory is largely academic, although it does have methodological implications. Those who espouse a corpus-as-theory stance adopt a *corpus-driven* approach to analysis (Francis, 1993; Tognini-Bonelli, 2001) as opposed to the deductive *corpus-based* approach. Thus linguistic description emerges from observation of data and the analyst does not risk overlooking potentially important features. But it is argued that an important consideration, especially in CADS studies, is the nature of the research question (McEnery and Wilson, 2001). In order to avoid overlooking potentially salient features, the current study employs keyword analysis, a proven method for enabling corpus-driven analysis (Groom, 2007). The keywords, though, are used as a rough guide for indicating where one might usefully look for answers to the research questions. Insights from the literature reviewed in Chapter Two are drawn on to narrow down candidate keywords.

### 3.4.3 The idiom principle and the process of co-selection

In many ways, the Sinclairean approach to corpus linguistics does not differ much from approaches which see corpus linguistics as a method. Both approaches recognise the key role played by collocation in creating meaning. Researchers in this tradition are not unlike many in the corpus-as-method tradition in accepting the fundamental premise that the basic unit of meaning is the extended phraseological unit. However, the Sinclairean approach is based on a more complex theory of how meaning is created. It gives prominence not only to collocation and multi-word units, but to the view, articulated by Firth (1957; 1968a; 1968b) and embraced

by Halliday (1985), that the semantic and grammatical systems are interdependent. Language is seen as a system of choices where choice is constrained not only by the grammatical system but by the phraseological tendency of language. The phraseological tendency as an organising principle of language is summed up in Sinclair's description of the *idiom principle* (which contrasts with the *open choice* or *slot-and-filler principle*):

> The principle of idiom is that a language user has available to him or her a large number of semi-preconstructed phrases that constitute single choices, even though they might appear to be analysable into segments.
>
> (Sinclair, 1991: 110)

Of course, once we define the extended phraseological unit as the basic unit of meaning we run into problems of definition and of terminology. Defining the precise extent of a phraseological unit is problematic in itself, as Sinclair (1991: 111) himself observes. He gives the example of *set eyes on*. Remarking that it is often preceded by a pronoun subject and either the adverb *never* or a temporal expression, such as *the moment* or *the first time*, he poses the problem of how to decide how much is part of the phrase and how much is simply the result of collocation. But defining the extended unit is also problematic because of inconsistencies in the terminology used. Although the ambiguity of the term *word* is widely acknowledged in the literature as problematic and *lexical item* is preferred[24], the term *word* is nonetheless commonly used in the literature. Sometimes it refers to extended units of meaning and sometimes to units of meaning which correspond to the orthographic word. The problem arises because corpus linguistic analysis relies on computational tools which can only recognise the orthographic word. The following definition of unit of meaning attempts to

[24] Hanks (2013: 23), however, comments on the ambiguity of the term *lexical item*, pointing out that it is sometimes used both to refer to the forms which make up a lexeme (or lemma) and to refer to phrasemes, or multi-word expressions.

make clear how unambiguous meaning arises from the occurrence of a word (corresponding to a single orthographic unit) in its collocational environment:

> A unit of meaning is a word (often called the node or keyword) plus all those words within its textual context that are needed to disambiguate this word, to make it monosemous.
>
> (Teubert, 2004: 83)

Sinclairean corpus linguists typically hold the view, often expressed by some lexicographers, that words do not have intrinsic meaning and that meaning is construed from the lexico-grammatical environment in which a word typically occurs (see Hanks, 2013). There is, of course, a potential circularity in this idea, a problem to which Barnbrook, Mason and Krishnamurthy (2013: 43) allude. How can a word, empty of meaning, acquire meaning from other words, presumably also empty? This question is answered if we assume that 'words in isolation are neither completely meaningful nor completely meaningless; rather, they consist of clusters of semantic components constituting a meaning potential' (Hanks, 2013: 81-82).

On the subject of the interdependence of lexis and grammar, Sinclair's (2004) theory of choice aims at addressing the shortcomings of traditional linguistic accounts, which, he argues, have wrongly separated grammar and semantics and have not given collocation the prominence it deserves, focusing instead on explanations of language which favour the paradigmatic dimension. It is worthwhile looking briefly, at this point, at the model he proposes, since it is crucial to understanding how, according to Sinclair, phenomena such as *collocation, colligation, semantic preference* and *semantic prosody* interact to create meaning. His model proposes five categories of co-selection, two of which are obligatory and three of which are optional. The obligatory categories are the *core*, 'which is invariable, and constitutes the evidence of the occurrence of the item as a whole' (ibid.: 141) and *semantic*

*prosody*, which he describes as 'the determiner of meaning as a whole' (ibid.) and as 'a subtle element of attitudinal, often pragmatic meaning' (ibid.: 145). The optional categories are *collocation, colligation,* and *semantic preference*. *Collocation* is defined as 'the co-occurrence of words with no more than four intervening words' (ibid.: 141), *colligation* as the 'co-occurrence of a member of a grammatical class … with a word or phrase' (ibid.: 142) and *semantic preference* as 'the restriction of regular co-occurrence to items which share a semantic feature, for example, that they all relate to, say, sport or suffering (ibid.). A useful summary of the theory is supplied byCheng (2012: 114):

> First the speaker/writer selects a **semantic prosody** of ***x*** [applied to a **semantic preference *y***].
> The **semantic preference** in turn controls the **collocational** and **colligational patterns**.
> The final component of the **lexical item** is the **invariable core**.

Sinclair (2004: 142-145) illustrates his argument with the example of the verb *budge*. The reason a speaker chooses the word *budge* in preference to its near-synonym *move,* is that its attitudinal meaning allows the speaker to express frustration at their inability to make something or someone move, despite repeated attempts, or at someone's refusal to move (Sinclair, 2004: 144). The semantic prosody of *budge* encodes the idea of refusal and inability. This is evidenced in the semantic preference of *budge* for items indicating these notions. The core, *budge*, tends to colligate, in its immediate left-hand environment, with the negative particles *not* or *n't* and with modal verbs expressing ability or willingness, such as *wouldn't, didn't, couldn't, won't* and *can't* (Sinclair, 2004: 143-144). It also collocates with the verb *REFUSE*[25] (ibid.). Sometimes, it co-occurs with prepositional phrases which begin with prepositions such as *from* and *on* and which indicate a position from which a person or

[25] In accordance with common usage in lexicological studies, lemmas are written in upper case, word forms in lower case.

thing will not budge (ibid.: 144). These observations inform subsequent work in the Sinclairean tradition on the nature of phraseological patterns and the relationship between lexis and grammar, and on the nature of the meaning expressed in collocational patterns. Approaches to the analysis of phraseology are discussed in the next sub-section. The subsequent sub-section addresses the challenges involved in reconciling a phraseological view of language with an approach to analysis which uses as its starting point a keyword analysis, an approach which uses the single orthographic unit as its basic unit of meaning. The vexed question of the various ways in which semantic prosody is interpreted by writers other than Sinclair is addressed in Section 3.5.2, which looks at approaches to analysing attitudinal meaning.

### 3.4.4 Phraseology

As mentioned above, it is commonly accepted in both traditions of corpus linguistics that language use is characterised by the recurrence of particular phraseological patterns. Corpus methods for examining such recurrent patterns typically look for repeated consecutive strings of orthographic units, variously referred to as *n-grams*, or *lexical bundles* (Biber, Conrad and Reppen, 1998; Biber *et al.*, 1999; Biber, Conrad and Cortes, 2004), *clusters* (Scott and Tribble, 2006) or *chains* (Stubbs, 2003). The advantages of using a fully automated method to uncover repeated patterns are clear and work in this field has yielded useful insights. Biber, Conrad and Cortes (2004), for example, find that most multi-word items fulfil a cohesive function, typically used in spoken conversation to bridge two clauses (for example, *I want to know or well that's what I*) and in academic English to bridge two phrases (for example, *in the case of* or *the base of the*) (ibid.: 377). They also express attitudinal meaning. For example, *I don't know what* expresses uncertainty, *I don't know if* and *I don't think (so)*

express uncertainty and possibility (ibid., 2004: 389). Stubbs (2002), meanwhile, uncovers interesting facts related to the colligational environment of the word *cases*, which, it transpires, frequently occurs in phrases which act as quantifiers, such as *in some cases*, *in many cases* (ibid.: 65).

While studies of this kind have been valuable in providing insight into the meanings and functions of multi-word units, this approach has the drawback of overlooking instances where there is variation in phraseological patterns. Varying degrees of regularity and variability is an inherent property of phraseology. For example, while some expressions and strong collocational pairings are relatively fixed (for example, *of course, hard work, hard luck, hard evidence, hard facts*), there is internal lexical and/or syntactical variation in many, for example *in some cases/ in some instances* and *set x on fire/ set fire to x* (Sinclair, 1991: 111-112). In the expression *it's not in his nature to …, nature* is fixed but *his* can be replaced with any possessive form and *not* can be replaced by any broad negative (ibid.). Recent work which addresses questions of regularity and variability in phraseology are Hunston and Francis' (1999) pattern grammar, Hoey's (2005) work on lexical priming and Hunston's (2008) concept of the semantic sequence.

Hoey's (2005) work is ground-breaking in that is provides a way of describing the semantic associations which a given word may have, even though the notions in question may be expressed through different words. For example, *hour* has an association with words from the semantic sets NUMBER and JOURNEY,[26] realised through phrases such as *half-hour drive; four-hour flight; two-hour trip; three-hour journey; two-hour hop; three-hour slog* (ibid.: 16-

[26] The names Hoey (2005) gives to the semantic sets he refers to are written in upper case.

17). Hunston's (2008) concept of the semantic sequence is similar but, unlike Hoey (2005), Hunston (2008) is not concerned with describing cognitive processes but with the product alone. Importantly, her model draws on pattern grammar and her description takes into account the evaluative and cohesive functions of semantic sequences. She shows, for example, that the phrase *to make sure* has a positive semantic association, since among its frequent left-hand collocates are modal expressions of volition, obligation, necessity, importance, possibility or achievement, for example, *want, try, go out of one's way, need, had, BE important, the best way,* and so on (ibid.: 275-276). It therefore participates, among others, in the semantic sequence 'possibility + *to make sure* + *that*-clause' (ibid.: 277). It is cohesive in its function in that it marks a transition between two grammar patterns: it follows a verb or noun phrase which governs a clause beginning with a *to*-infinitive and is itself followed by a *that*-clause, for example, *the one way to make sure (that)* …. (ibid.: 274).

Meanwhile, nouns which occur in the **N-*that*** pattern, that is, a noun followed by an appositive *that*-clause, are particularly important in the construction of disciplinary epistemology and indicate the ideological view of the world adopted in a given discipline (Hunston, 2008: 278). These nouns typically fall into one of six meaning groups:

1. The 'suggestion' group, in which the noun refers to something written or spoken and the *that*-clause supplies information about it: *allegation; assertion; claim; suggestion*; and so on.
2. The 'belief' group, in which the noun refers to beliefs, ideas, and thought processes: *belief; idea; notion*; and so on.

3. The ‘happiness’ group, in which the noun refers to an emotion: *concern; fear*; *worry*; and so on.
4. The ‘sign’ group, in which the noun refers to signs or evidence of something: *evidence; proof*; and so on.
5. The ‘possibility’ group, in which the noun refers to an assessment of how likely something is: *chance; danger, likelihood; possibility; probability; risk*.
6. A final group in which the proposition expressed in the *that*-clause defines or expands upon the meaning expressed by the noun, for example, *consequence, basis, fact, problem*, *truth*, and so on.

(Francis, Hunston and Manning, 1998: 108-113)

Nouns which are used with the N-*that* pattern are a frequent feature of the discourse of the *JABS* corpus and are examined in Chapter Seven. The six-way sub-categorization of these nouns illustrated above is also used to inform the division into semantic sets of some of the keywords in the *JABS* and *NHSvax* corpora (see Chapter Six).

### 3.4.5 Phraseology and keyword analysis: some implications and challenges

As Hunston (2002: 3) remarks, a corpus on its own simply provides raw data. Software tools are necessary to manipulate the data so that the corpus linguist can make observations about language use and the features of a given software program necessarily influence the sorts of observations the analyst is able to make (McEnery and Hardie, 2012: 123). Several software tools, among them *WordSmith Tools* (Scott, 2004), which is employed in this study, include a keyword function. As discussed in Sub-section 3.4.2 above, keyword analysis provides a reliable starting point in a corpus-driven study (Groom, 2007). Keyword analysis has been

used to good effect in corpus-assisted studies of healthcare interaction online, where analysis of lexical keywords has been used to identify key themes in the data (Adolphs *et al.*, 2004; Harvey *et al.*, 2007; Seale *et al.*, 2010; Harvey, 2012; Harvey, Locher and Mullany, 2013). However, keyword tools function by identifying individual orthographic units and comparing their relative frequency with that in a reference corpus. If one assumes that the basic unit of meaning is the extended phraseological unit, there is a tension here. One way of overcoming this is to focus on grammatical keywords rather than lexical keywords. Grammatical keywords help identify salient semantic sequences, which provide a view of the salient epistemological features of a discourse type (Groom, 2007). This thesis is interested in discovering the ways in which epistemological claims are made, but it is also interested in discovering, to a certain extent, what sort of claims are made. For example, it is interested in discovering how claims about risk are expressed. There are 110 instances of *at risk of* in the *JABS* corpus. This is estimated to represent a relative frequency of 27 instances per million words, in stark contrast to the 6 instances per million words in ukWaC.[27] However, neither *at* nor *of* are keywords in the *JABS* corpus, either when cross-compared with the *BNC* or when compared with *NHSvax*. *Risk* is a keyword in the *JABS* corpus, though, when compared with the BNC. This study therefore uses lexical keywords to identify central themes but also uses grammatical keywords where appropriate to analyse, for example, use of pronouns or particular cohesive features. Since one of the central features of phraseology is that phrases tend to have a fixed core with collocates which show a degree of variation, keyword analysis can lead the researcher to salient repeated phraseological patterns. As is standard practice in CADS work, keyword analysis in this thesis acts as a prelude to qualitative analysis of concordances.

[27] Relative frequencies calculated by the Skylight program, http://www.skylight-to-english.co.uk/skylight.

### 3.4.6 Summary of Section 3.4

We can summarize the main points of the literature reviewed in this section as follows:

- A word does not have meaning in itself but meaning potential which is realized when the word is used in a particular lexico-grammatical context;
- The basic unit of meaning is the extended unit of meaning, composed of an invariable core with variable elements in its environment;
- The core of a lexical item should be defined as a word-form, not a lemma, since the various word-forms of a given lemma differ from each other in terms of their collocational environments – in corpus analysis, the individual forms of a lemma should therefore be examined separately;
- The concept of semantic sequences best accounts for the way in which meaning is mapped onto the lexico-grammatical system;
- Semantic sequences are central to the construction of epistemology and ideology in disciplines.

## 3.5 Attitudinal meaning

### 3.5.1 Introduction to section 3.5

One of the most powerful discursive resources concerns the expression of attitudinal meaning. The terms most commonly used in discourse studies to refer to attitudinal meaning are *evaluation* (e.g. Labov, 1972; Hoey, 1983; Carter, 1987; 1998; Hunston, 1994; 2000; 2011; Channell, 2000; Hunston and Thompson, 2000; Bondi and Mauranen, 2003), *intensity* (Labov, 1984), *evidentiality* (Chafe and Nichols, 1986), *affect* (Ochs, 1989), *stance* (e.g. Biber

and Finegan, 1988; 1989; Beach and Anson, 1992; Barton, 1993; Conrad and Biber, 1999; Charles, 2006a; Biber, Connor and Upton, 2007; Myers, 2010), *stance* and *engagement* (e.g. Hyland, 1999; 2004; 2005; 2007; 2010),[28] *APPRAISAL*[29] (Martin, 2000; Martin and White, 2005), or *sentiment* (e.g. Pang and Lee, 2008; Thelwall, Buckley and Paltoglou, 2012). In some works, the study of attitudinal meaning is limited to examination of *modality* (Halliday, 1985; 1994) or *hedging* (Holmes, 1988; Hyland, 1996a; 1996b). The term a writer chooses to use normally reflects the theoretical approach which is adopted or signals a different emphasis in approach, but sometimes it indicates that a different phenomenon is under consideration. Labov (1972) and Hoey (1983), for example, use the term *evaluation* to refer to the function of a section of text. Carter (1987; 1998) and Channell (2000) use it to refer to the meaning associations a word acquires through its collocates. Used this way, it equates to semantic preference and semantic prosody. Generally speaking, *evaluation*, *stance*, *intensity*, *evidentiality*, *affect*, *APPRAISAL* and *sentiment*, refer to the various lexico-grammatical resources available to a user to express opinions and judgements. I choose to use the term *evaluation*, not only on the grounds that it is the most commonly used term and therefore covers a wide range of phenomena, but also because, although all approaches have useful things to say about the ways in which discursive resources are used by writers (or speakers) to align readers (or listeners) with their point of view, particular prominence is given in this study to Hunston's (2000; 2011) model. Semantic preference and semantic prosody are obviously central to the expression of evaluative meaning. They are discussed in the next subsection. But there are aspects of the signalling of attitudinal meaning which operate at a macro-textual level. Models of evaluation which attempt to provide a more comprehensive model of the phenomenon are discussed in subsequent sub-sections.

[28] *Stance* and *engagement* are sometimes dealt with as aspects of *metadiscourse* (e.g. Hyland and Tse, 2004).
[29] I follow Martin and White's (2005) convention of writing the term in upper case script.

### 3.5.2 Semantic preference and semantic prosody

Semantic preference and semantic prosody are central to the creation of meaning, so their analysis forms a central element of much work in corpus linguistics, as well as forming the basis of some approaches to studying evaluation in language (e.g. Channell, 2000). Unfortunately, there is some confusion in the use of these terms and the concept of semantic prosody is contested. Starting with the distinction between the two concepts, we have already established that Sinclair (2004) conceives of semantic prosody as something which determines meaning as a whole and governs choice, whereas semantic preference is simply a facet of collocation. Few writers, however, distinguish between semantic preference and semantic prosody and most writers on semantic prosody conceive of it in simpler terms than we see in Sinclair's (ibid., 2004) writing. Before examining the way in which semantic prosody is understood by most writers, it is useful to consider the findings of scholars who have looked at semantic preference so that we might better distinguish between the two. Two such writers are Stubbs (2002) and Partington (2004b). Stubbs (2002: 65) demonstrates that *large* shows a preference for collocates denoting quantities and sizes, such as *number(s), scale, part, amounts, quantities* and *area(s).* Partington (2004b: 145-146), commenting on his own earlier work (Partington, 1998: 34-39), explains that the intensifying adjective *sheer* shows a semantic preference for five sets of semantically related words: (i) denoting magnitude, weight or volume; (ii) expressing the idea of force, strength or energy; (iii) expressing persistence; (iv) expressing strong emotion; (v) denoting a physical quality. The boundary between semantic preference and semantic prosody is somewhat fuzzy, as Stubbs

(2002: 66) concedes, but semantic preference is purely a semantic phenomenon whereas semantic prosody is a pragmatic phenomenon (ibid.).[30]

Turning now to semantic prosody, few writers conceive of it in the way in which Sinclair (2004) does. The notion of semantic prosody commonly embraced is that described by Louw (1993). Louw (ibid.) argues that the collocational environment in which a word habitually occurs can imbue that word with a particular connotative meaning. Having taken on this meaning, the connotation remains even when the word is used in a context free of its typical collocates. When a word is used in a context which runs counter to its prosody, it is either the result of intentional irony on the part of the speaker or writer or, if done unconsciously, it indicates insincerity. Louw's (ibid.) concept of semantic prosody as akin to connotation has been highly influential, with the result that most writers see it in terms of a binary opposition between negative and positive connotational meaning (Hunston, 2007). It is also seen as the property of a word rather than the property of an extended unit of meaning. *CAUSE*, for example, is said to have a negative connotation since it collocates with words such as *problems, death, damage, concern, trouble, cancer* and *disease*, while *PROVIDE* has a positive prosody, collocating, as it does, with *information, services, support, help, money, protection, food*, and *care* (Stubbs, 2002: 65). Even Sinclair, in some of his work, describes prosodies in these terms. He states, for example, that the verb *HAPPEN* typically collocates with words referring to unpleasant events (Sinclair, 1991: 112) and the subjects of the phrasal verb *SET in* overwhelmingly denote unpleasant states of affairs, for example *rot, decay, malaise, despair, ill-will*, and so on (ibid.: 74-75). In some work (e.g. Partington, 2004b)

[30] Because semantic prosody is a pragmatic rather than semantic phenomenon, Stubbs (2002) prefers the term *discourse prosody*. Although this is a sound point, most writers continue to use the term *semantic prosody*, and so this is the term used in this project.

semantic prosody is conceived of as gradable. For example, *SET in* is described as having 'an extremely unfavourable prosody' (ibid.: 135).

The notion of semantic prosody has come in for substantial criticism, notably from Whitsitt (2005), although others, even writers supportive of the concept, such as Hunston (2007), find fault with particular theoretical aspects or with the way in which the concept is operationalised in the literature. Two aspects of Louw's (1993) argument attract particular criticism. The first point of contention concerns Louw's (1993) claims that semantic prosody provides evidence for intentional irony or brings to light insincerity. Hunston (2007: 260-261) concurs with this argument to a certain extent, agreeing that semantic prosody can help explain a reader's response to a text. But she takes issue with the idea that semantic prosody has predictive power, since this implies greater uniformity among language users than is the case. The second point of contention concerns the fact that Louw's (1993) theory relies on the argument, criticised earlier for its potential circularity, that words have no intrinsic meaning but acquire their meaning from other words around them. Whitsitt (2005: 291-292) argues that it is impossible to prove that such a one-way flow of meaning from content-full to content-free words occurs and, if it occurs, is inevitable. He also cites examples of words which collocate with words denoting unpleasant things but which do not have a negative prosody, such as *alleviate, heal, relieve* and *soothe* (Whitsitt, 2005: 296-297). Hunston (2007: 266) concedes that this criticism has some validity. It highlights the inherent contradiction in claiming, on the one hand, that meaning exists only in context, and, on the other, that meanings can be transferred between contexts. But she argues that it is undeniable that intertextual resonances do occur and, while one cannot say that meanings always transfer between contexts, one cannot argue that they never do. Of course, if, like Hanks (2013), one

accepts that words have meaning potential, the paradox is resolved: the meaning potential of *alleviate, heal, relieve,* and *soothe* entails the idea of making an unpleasant condition less severe.

It is has been pointed out that conceiving of semantic prosody in terms of negative or positive connotation is simplistic, as is ascribing prosody to individual words (Hunston, 2007: 256). Prosody, Hunston (ibid.) argues, is best conceived of in Sinclair's (2004) terms as a determiner of meaning which is expressed in extended phraseological units. Furthermore, as Partington's (2004b) work shows, semantic prosody is context dependent in complex ways. Even though he (ibid.) defines *semantic prosody* in terms of positive and negative connotation, his discussion covers aspects of semantic preference and colligation as well as connotational meaning. He also shows that semantic prosody is not uniformly present. Its occurrence varies between the different forms of a lemma, and the same word form may be found in negative, positive or neutral contexts in different contexts. The evidence emerges from his comparison of the prosodies, in an academic corpus, of the semantically related set of words *HAPPEN, OCCUR, COME about* and *TAKE place*. He finds that the forms *happen* and *happens* occur twice as often in negative contexts as in positive, but *happened* occurs almost five times as often in negative contexts than in positive. All occur frequently in contexts which are neither positive nor negative. Grammatically, *happen* often occurs with modals or conditionals plus *if*. *Happens* and *happened* occur frequently in questions or relative clauses with *what* or *why*. Uncertainty or 'non-factuality' is therefore part of the meaning of *HAPPEN* (ibid.: 140). *OCCUR* also expresses non-factuality but is used in academic texts as a predicator following nominalizations of processes or actions, as in, for example '*no violent behaviour*

**had occurred**'[31] (ibid.: 141). *COME about* appears to be prosodically neutral and 'seems to be used when writers wish to emphasize the concept of *process* rather than inherent unpleasantness' (ibid.: 142) and *TAKE place* connotes factuality (ibid.: 143). Finally, all of the words in this semantic set serve a cohesive function in texts, linking one clause with another, for example, '*ask/look/consider/don't care what happens, What happens depends on / is that, What happens...? […] When/If/Whether this happens...'* (ibid.: 137).

Partington's (2004b) discussion highlights the complex nature of the lexico-grammatical system as a means for making meaning. Meaning involves semantic and pragmatic elements but it is not always easy, and may not be desirable, to fully distinguish between expressions of semantic and pragmatic meaning. The associations that a word-form has are highly context dependent. They depend on the lexical and grammatical environment in which they occur. It is also important to examine different forms of a lemma separately since they each build up different associations. It is important too to consider the discursive functions phraseological units fulfil. Perhaps, as Hunston (2007) argues, *semantic prosody* is best understood in Sinclair's terms (1991; 2004) as referring to 'the consistent discourse function of the unit formed by a series of co-occurrences' (Hunston, 2007: 257). The words *cause*, *happen* and *occur* are indeed keywords in the *JABS* corpus, as are other terms used to express causation and events.

[31] Use of italics and bold type face are Partington's (2004b).

### 3.5.3 Corpus-based and discourse approaches to evaluation

#### 3.5.3.1 Sentiment analysis

It is clear that an important resource for expressing evaluative meaning is represented by the choice of words or phrases with particular semantic prosody. But the expression of evaluation may go beyond semantic prosody, as much work in the growing field of the study of evaluation has shown. In a corpus-assisted study, such as the current one, it is important to find an approach which offers as comprehensive a model of evaluation as possible but lends itself to corpus analysis. As a computationally based approach to analysing evaluative language, sentiment analysis (e.g. Pang and Lee, 2008; Thelwall, Buckley and Paltoglou, 2012) might seem a useful approach to adopt in a corpus-assisted project. However, it relies on complex algorithms to highlight evaluative instances of language in texts, assigning each occurrence a value on the basis of pre-determined assessments of sentiment. It is suggested that current algorithms are not suitable for the analysis of topic specific texts (Thelwall and Buckley, 2013).

#### 3.5.3.2 Stance and metadiscourse models

The *stance* or *metadiscourse* models (e.g. Chafe and Nichols, 1986; Biber and Finegan, 1988; 1989; Barton, 1993; Conrad and Biber, 1999; Biber, Connor and Upton, 2007; Hyland, 1999; 2004; 2005; Hyland and Tse, 2004)[32], although originally developed on the basis of the qualitative analysis of individual texts, have been extensively employed using a corpus approach. They lend themselves well to corpus-based analysis because, while the complexity of systems of evaluation in language is recognised, these approaches involve the identification and analysis of sets of linguistic markers associated with evaluation. Such markers are usually referred to as *evidentials* (Chafe and Nichols, 1986; Barton, 1993) or *stance markers* (Biber

[32] I include here writers who refer to the phenomenon as *evidentiality*.

and Finegan, 1989; Hyland, 1999; 2005; Biber, Connor and Upton, 2007). These include modals and modal-like expressions (including hedges and boosters), sentence adverbs, sentence-initial conjunctions, and verbs, nouns, adjectives or adverbs which express attitude. Hyland and Tse (2004) include such markers in the wider category of *interactional resources*. Their model represents an elaboration on Hyland's (Hyland, 2005; 2007) model of *stance and engagement*, in which the category of *engagement* includes the resources with which the author interacts with the reader. These resources include the use of directives, second person pronouns, self-mentions, questions, appeal to shared knowledge and personal asides.

The approach of Biber and his associates differs from the others in this group in that it recognises the importance of multi-word units in the construction of meaning. An approach based on the analysis of *stance bundles* has been developed (e.g. Biber, Conrad and Cortes, 2003). Stance bundles are useful indicators of how an argument is structured, since they form a bridge between sections of the text. In conversation, stance bundles typically bridge two clauses, although in academic prose, they more frequently bridge two phrases. The functional taxonomy of stance bundles is illustrated in Table 3.1 below.

| **Function** | **Dimension** | **Examples** |
|---|---|---|
| Epistemic modality: certain | personal | *do you know what; know what I mean* |
| Epistemic modality: certain | impersonal | *the fact that the; it's going to take* |
| Epistemic modality: uncertain | personal | *I don't know if; I don't know what; I don't know how/whether/why; I don't think so* |
| Epistemic modality: probable/ possible | personal | *I thought it was; I think it was* |
| Epistemic modality: probable/ possible | impersonal | *it is possible to* |
| Desire | | *if you want to; do you want to; do you want me to; I don't want to; I would like to* |
| Obligation | | *we're going to have; it's necessary to* |
| Intention | | *are we going to; are you going to; I was going to; going to have to; let's have a look* |

**Table 3.1** Functional taxonomy of lexical bundles (4 orthographic words), from (Biber, Conrad and Cortes, 2003: 80)

The drawback of the stance and metadiscourse approaches, though, is that they are corpus-based rather than corpus-driven. They rely on the a-priori identification of specific lexico-grammatical features. It has been persuasively argued that there is not a straightforward one-to-one relationship between a certain lexical or grammatical feature and evaluation. Carter (1987: 78) discusses the way the meanings of a lexical item are negotiated in naturally occurring discourse so that the collocational environment of a given word can signal a certain evaluation. In other words, semantic prosodies are central to evaluative meaning. By adopting a corpus-based approach, the researcher risks overlooking words where evaluative meaning is not immediately obvious even to a competent native speaker. Furthermore, not only is there no closed set of lexico-grammatical items which express evaluation, some instances of evaluation are realised without the use of explicitly evaluative terms An otherwise neutral term may accrue an evaluative meaning, especially if there is an accumulation of explicitly evaluative language in the context (Hunston, 2011: 3). By way of example, Hunston (ibid.) cites following passage:

> As I write this, Professor Smith, now a distinguished scholar, has her job under threat from the ghastly, grey accountants who run the University of Biggin-on-Sea. We are now in an epoch of production-line universities with celebrities paid fortunes to teach eight hours a week and genuine scholars dumped in the bin.
>
> (Ali, 2008: cited in Hunston, 2011: 3)

Hunston (2011: 3) argues that, because of the cumulative effect of phrases such as *ghastly, grey accountants* and *production-line universities* and *genuine scholars dumped in the bin*, we interpret *celebrities paid fortunes to teach eight hours a week* as a negative evaluation.

#### 3.5.3.3 APPRAISAL

The APPRAISAL model, on the other hand, attempts to describe the use of APPRAISAL resources across a range of contexts. As with the stance and engagement model, Martin and White (2005) distinguish between two main types of discursive resource. Their category of *attitude* equates to *stance*. Their category of *engagement* is similar to Hyland's concept, but it is conceptualised in terms of Bakhtin's theory of dialogicality. Texts, or sections of texts, are either monoglossic, representing only one voice, or heteroglossic, that is, they entertain the possibility that other voices exist. A writer who acknowledges the existence of other voices, though, may express propositions in such a way as to close down the discursive space or may expand it. Dialogic contraction is realised through expressions such as *I concur that…, the facts of the matter are that…, this is not the case*, and so on. Dialogic expansion is realised through expressions such as *it's probable that …, it seems to me that …, x argues that …*, and so on. The model also takes into account the resources available for intensifying or hedging the force of a proposition.

The APPRAISAL model presents a complex but comprehensive way of describing evaluation in interaction. Unlike the stance and metadiscourse models, which are based largely on academic discourse, it is based on the analysis of texts from more than one register. It is feasible that it might provide a better fit for analysing the texts in the *JABS* corpus. However, Hunston (2011: 54) reports that attempts at automatic evaluation recognition using APPRAISAL have met with limited success. She attributes this to the fact that this model is designed for qualitative analyses of texts. Its shortcomings regarding the application to corpus analysis might also lie elsewhere. Martin's (2000) early explanation of the APPRAISAL model was criticised on the grounds that it is unclear from his account how frequently the

expressions he lists are used in naturally occurring contexts (Hyland, 2005: 175). A similar criticism might be levelled at Martin and White's (2005) later work. Although the description is also based on examination of a large body of texts, one has to ask how often words such as *scold* and *castigate,* for example, are used in real life contexts to express dissatisfaction or phrases such as *this is not the case* are used to deny the truth value of a proposition. It may be that, because the work is not corpus based, several of the examples they use to illustrate their model appear unrepresentative.

#### 3.5.3.4 Hunston's (2000; 2011) 'status and value' approach

Like the stance and metadiscourse models, Hunston's (2000) model was designed for the qualitative analysis of individual texts. It is also based on the analysis of academic texts. However, in a later work Hunston (2011) successfully applies her analysis to texts from other genres and she demonstrates how the model can be adapted for the purposes of corpus anlaysis. The model draws on Sinclair's (1981) idea that sentences in texts simultaneously operate on the *autonomous plane,* that is, they make statements about the world, and the *interactive plane,* through which the writer informs the reader about the text. On the autonomous plane, evaluation is achieved through labelling entities, while on the interactive plane, evaluation is connected to whether a statement is (following Sinclair, 1986) averred or attributed, that is, whether a writer states something as a fact or delegates responsibility for the validity of the statement. A complicating factor is that a statement can become an object of evaluation. In becoming a discursive object, a proposition is labelled and the label in itself expresses the writer's evaluation of the status of the proposition.

The notion of the 'status' noun is similar to Francis' (1986) interpretation of the anaphoric noun. Like general nouns, such as *problem* and *issue*, anaphoric nouns, such as, *fact*, *hypothesis* and *truth* encapsulate the preceding stretch of discourse and signal the writer's interpretation of its epistemic status. An anaphoric noun points forward in the argument, since it is presented as 'the *given* information in terms of which the *new* propositional content of the clause or sentence in which it occurs is formulated (ibid.: 4). Anaphoric nouns can also signal whether or not the speaker ascribes factuality to the proposition referred to. Nouns such as *fact* and *truth* and so on, belong to the class of 'factive' nouns , that is nouns which encode a presupposition that what they refer to is true (ibid.: 25). *Hypothesis* clearly does not. The status that a writer affords a proposition tells the reader how he or she should respond to it: whether to agree or disagree, for example. It is also connected to whether the statement is averred or attributed. The status of one statement constrains the *value* that can be afforded it, that is, how the reader should evaluate it and how it should subsequently be evaluated in the text. Status works on both the interactive and autonomous planes, although on the interactive plane it is equivalent, to use Halliday's (1994) terms, to *modality* and on the autonomous plane to *affect*. The notion of value, meanwhile, equates to notions of *stance* (Hyland, 2005) or *attitude* (Martin and White, 2005).

Hunston's (2000) model provides a powerful method for analysing evaluation in text. It was, of course, originally developed with the aim of analysing individual texts. In her exploration of the ways in which the model can be adapted for corpus analysis, Hunston (2011) acknowledges, as Biber, Conrad and Cortes (2003) do, that multi-word units are a fundamental means for creating meaning. She examines the ways in which modal meaning is encoded not only in modal auxiliaries or modal phrases (such as *I think, kind of, sort of*, and

so on), but in less obvious expressions. The collocational environments of verbs such as *decide* and *distinguish*, for example, indicate that the phraseology in which those words occur encode modal meaning, since they are frequently preceded by words and phrases which express possibility or obligation, and so on (Hunston, 2011: 70-77). The expression *to make sure*, similarly encodes modal meaning. Expressions involving status nouns, such as *the assumption that, the fact that, evidence that,* and so on, form a bridge between one clause and the next and drive the argument forwards at the same time as expressing an evaluation of the proposition to which they refer.

The meaning encoded by such nouns in certain lexico-grammatical environments may shift, however. The noun *fact*, when used in the expression *the fact that*, may not necessarily signal factivity. Firstly, as Francis observes, *fact* used with an appositive *that*-clause has become largely delexicalised since,

> the fixed phrase *the fact that* has taken on a general role as all-purpose nominalisation device in cases where a noun or nominal group is required by the grammar of the preceding elements.
>
> (Francis, 1986: 154)

Secondly, while most contexts of use of *fact* indicate that the status of the 'fact' is non-negotiable, in some contexts, for example when preceded by certain adjectives, co-ordinated with *and* and *or*, or occurring in the pattern V n *as* n, its status is negotiable (Hunston, 2011). It also encodes other meanings. The environment in which *the fact that* occurs reveals that the phrase participates in a number of semantic sequences. Hunston (ibid.: 115-116) identifies three broad 'motifs' expressed by phraseologies which include *the fact that*: the 'cause' motif (i.e. 'facts' are the basis of an outcome or reasoning, they explain something, or cause a

problem or solution); the 'orientation' motif (i.e. 'facts' imply something, are acknowledged or ignored); and the 'human response' motif (people are (un)aware of a fact, talk about a fact, react to a fact). Although the expression *the fact that* is often delexicalised, it nonetheless serves the function of reifying the proposition in the *that*-clause and carries with it the presupposition that the said object exists as a fact (Francis, 1986: 154). It implies that the speaker or writer is in possession of facts that exist in the world, and, since it implies given knowledge, it can represent an appeal to shared knowledge (ibid.).

Hunston (2011) notes that expressions involving factive status nouns are worth investigating because they shed light on the ways in which propositions become 'facts' and how they 'travel' (Morgan, 2007, cited in Hunston, 2011: 116-117). They are vehicles by means of which propositions advanced in one domain travel to another (Hunston, 2011: 118). *Fact* is the most frequent of the several status nouns which appear as keywords in the *JABS* corpus, so that part of the analysis focuses on *the fact that* and other similar expressions. By acknowledging the role of status nouns in the expression of evaluation, Hunston's (2000; 2011) model is suited to the needs of the current project, which is interested in how propositions which originate in the domain of academia are rearticulated as they travel. Hunston's model is suited to the Sinclairean approach to corpus linguistics in that it accounts for the primacy of phraseology, the role of semantic prosody in contributing to meaning, and the interdependence of lexis and grammar.

### 3.5.4 Summary of Section 3.5

- The lexico-grammatical patterns in which a lexical item typically occurs colour the meaning of that item, contributing to both semantic and pragmatic meaning – we can refer to this phenomenon as *semantic prosody*;
- Multi-word items may encode evaluative meaning and often act as a cohesive link between clauses;
- The examination of status nouns is a guide to the ways in which epistemological propositions are expressed;
- Particular phraseological expressions, for example, the fact that, may participate in a variety of semantic sequences, each of which expresses a different form of evaluative meaning.

## 3.6 The approach to CADS adopted in this thesis

The beliefs about linguistic analysis expressed in this chapter have been largely sympathetic to the Neo-Firthian or Sinclairean view of language. As discussed in Section 3.4, the view of language espoused puts phraseology at the heart of linguistic description. However, the Sinclairean tradition of corpus linguistics also advocates taking a corpus-driven approach to language analysis. The extent to which any method of analysis can be wholly corpus-driven is a moot point since, as already discussed, even Sinclaireans, to a certain extent, draw on pre-existing descriptions of language to inform their analyses. There is no denying that a corpus-driven approach has its merits when the aim of the exercise is to produce a description of a language or a language variety, but it has its limitations in CADS, where the aim is to reach conclusions about the social implications of language use in a given situation. It is because of this difference in aims that CADS researchers typically deviate from a strict Sinclairean

approach. Often, they do not rely solely on corpus data but combine corpus analysis with insights from other sources of information (Partington, 2010: 90), whether, for example, dictionary definitions (e.g. Krishnamurthy, 1996; Duguid, 2010b) or selected publications (Taylor, 2010), and so on. In some studies, information gleaned from external sources is used to determine exactly which terms are to be analysed. Grundmann and Krishnamurthy (2010), for example, draw on previous studies on the discourse of climate change to select the search terms they investigate. More importantly, many CADS researchers read all, or parts of, their corpus data rather than treating the corpus as a 'black box' (Partington, 2010: 89).

In many respects, there are ultimately few material differences between corpus-driven studies and corpus-based studies, especially where corpus studies of discourse are concerned. CADS researchers use an inductive approach, even in those works which they label corpus-based. For example, in the best Sinclairean tradition, Grundmann and Krishnamurthy's (2010) 'corpus-based' study of the discourse of climate change starts with an analysis of word frequencies before moving on to investigate collocational patterns. Baker and McEnery (2005), in their 'corpus-based' study of the terms *refugee(s)* and *asylum seeker(s)*, base their conclusions about the semantic prosodies of the words entirely on collocational information gleaned by examination of concordances. Whatever way a researcher may choose to label his or her work, the basic approach in CADS is essentially inductive. CADS work also has a tendency to be somewhat eclectic in its approach. A good example is Partington's (2003) investigation of a corpus of White House press briefings. He starts with a keyword analysis and includes concordance analysis in his study, but he also draws on theories and methods from other branches of applied linguistics (such as conversation analysis) and much of his study relies on the use of extended concordance lines for exemplification. The approach

adopted in the current study is in line with the general CADS approach. It starts in a corpus-driven way by investigating corpus wordlists and keyword lists. However, insights from the literature reviewed in this and the previous chapter, together with consideration of the research questions driving the thesis, guide the selection of keywords which are singled out for closer scrutiny. Extensive use is also made of extended concordance lines.

## 3.7 Conclusion

This chapter has set out the theories about language which inform this thesis. It has argued against adopting a radically Foucauldian approach to discourse analysis but concedes that, in its weaker form, Foucault's notion of discourses can offer useful insights into discursive practices. It has also argued in favour of accepting the basic tenets of the Sinclairean, or neo-Firthian, theory of language, namely that the basic unit of language is the extended phraseological unit, a unit which encodes evaluative as well as semantic meaning. Hunston's (2011) approach to evaluation lends itself well to a corpus assisted approach to corpus linguistics. The chapter which follows considers some of the implications of neo-Firthian principles on aspects of corpus compilation, manipulation and analysis. It discusses basic principles which guide corpus compilation and discusses issues such as the use of statistics in corpus linguistics. It describes the process of compilation of the *JABS* and *NHSvax* corpora and the process of analysing the corpora.

# CHAPTER FOUR

# CORPUS COMPILATION AND METHODS OF ANALYSIS

## 4.1 Introduction

The aims of this chapter are to justify and describe the methods used for compiling, handling, and analysing the *JABS* and *NHSvax* corpora, and to describe the corpus contents and the sources from which the data was gathered. Section 4.2 discusses the factors to be considered when compiling and handling a specialist corpus and reference corpus, with particular attention paid to the challenges posed when using web-based data. It addresses issues pertinent to corpus construction, such as balance, representativeness and comparability of corpora, and considers the challenges related to rendering web-based texts suitable for use with corpus tools. It also considers arguments for and against part-of-speech tagging. Section 4.3 addresses the ethical issues related to using data from an online discussion forum. Section 4.4 starts by describing the data sources, before continuing to a description of the processes of compilation of the *JABS* and *NHSvax* corpora. The *JABS Cleanup* program, which was designed especially for the purpose of automating the compilation of the JABS discussion forum data, is described in this section.[33] A description of the size and contents of the two corpora is then presented.

Issues related to the processes of corpus analysis are addressed in Section 4.5. This section starts by discussing the advantages and disadvantages of using currently available statistical

[33] The *JABS Cleanup* program was constructed by Andrew Dickinson.

measures of significance in corpus linguistics. It then discusses keyword analysis, starting by distinguishing between the idea of the keyword as a discursive phenomenon and the keyword as an expression of a measurement of statistical significance, and going on to discuss the advantages and drawbacks of using keyword analysis. The discussion moves on to approaches to concordance analysis and the analysis of multi-word units and concludes by addressing the challenges of reconciling the qualitative approach of discourse analysis with the constraints and possibilities of corpus methods. Section 4.6 describes the method of analysis followed in this project and Section 4.7 concludes the chapter.

## 4.2 Corpus compilation and handling: considerations

### 4.2.1 Gathering corpus data

The decision to use web-based data was informed, in part, by the observation that, on the internet, 'emergent issues and current usage are generally well represented in online texts' (Fletcher, 2004: 191). The internet is also less dominated by traditional elites and represents a multiplicity of voices (Mautner, 2005: 815-816) and so a useful place to search for vaccine-critical discourse. Finally, web-based texts are easily processed (ibid.). Since this project is interested in the ways in which discursive resources are exploited in vaccine-critical discourse, it was decided to compile a specialist corpus from a specific source, following the example set by Richardson (2001; 2003; 2005) and Sotillo and Wang-Gempp (2004). A specialist reference corpus was then required to offer sufficient examples of vaccine-promotional discourse.

The decision to gather data from specific sources meant that, in the early stages of construction at least, compilation was carried out manually, since the available automated means were not suitable. The early years of the project coincided with the development of web-based corpus inquiry and corpus-building tools. The so-called *web-as-corpus* approach involves using a program which piggy-backs onto a commercial search engine and displays returns in concordance format (e.g. Fletcher, 2004; 2007; Renouf, Kehoe and Banerjee, 2005; 2007). Alternatively, a web-crawler can be used to compile a corpus. A web-crawler downloads, 'cleans', and stores web-based texts, which are identified on the basis of 'seed' terms, before loading the texts into a corpus query tool (e.g. Baroni *et al.*, 2009; Kilgarriff, 2007). Even though tools such as these became publicly available in the early years of this project, the drawback of using such tools is that one would be unable to distinguish between texts from vaccine-critical sources and others. A successful method for automating corpus compilation involves selecting texts from an online archive using specific corpus query terms (e.g. Baker and McEnery, 2005; Baker *et al.*, 2008). However, this method is best suited for topic-specific corpora limited to texts of a particular genre or domain. Manual compilation seemed the only option for constructing the *JABS* and *NHSvax* corpora.

With relatively small corpora, it is feasible to carry out the task of compilation manually. Examples are Partington's (2003) 450,000 word corpus of White House press briefings and Sotillo and Wang-Gempp's (2004) 274,639 word corpus of bulletin board discussions. In the case of the *NHSvax* corpus, most texts were relatively short, so compilation did not pose much of a challenge. Similarly, gathering the articles and such-like which were uploaded onto JABS website by the website administrators was relatively straightforward. But gathering the JABS message board and discussion forum data proved more of a challenge. A notable

problem with compiling web-based corpora lies in the ephemeral nature of the web: it changes constantly and pages are often amended or removed (Mautner, 2005: 817). Also, frequent additions to the existing data are a typical feature of websites. The NHS vaccination sites' data and the non-interactive texts on the JABS site were relatively stable. However, the content of the message board and discussion forum grew daily. Not only that, but the message board was taken down from the JABS site in September 2006, before compilation of that part of the corpus had started. Following Mautner's (ibid.) suggestion that a Wayback machine (an online archive of texts) can be used to fill gaps in a corpus, a Wayback machine was used to attempt to retrieve data from the JABS message board from 2001 and 2006. This was only partially successful, though, since, as Mautner (ibid.) observes, online text archives themselves have gaps. As for the discussion forum, as it grew in size, manual compilation became less feasible in the time available. The sheer quantity of the data (the discussion forum section of the JABS corpus totals over 3.5 million running words) meant that there was greater risk of introducing human error. The discussion forum data was therefore compiled automatically, using a specially designed program, the *JABS Cleanup* program. The program is described more fully in Section 4.6.

Whatever method of compilation one chooses, a problem with web-based texts is *noise*. Noise includes, among other things, boilerplate text (the headers and footers which identify the document) and general formulaic elements (Fletcher, 2004: 3). Boilerplate text does not contribute to the message of the text and so needs to be stripped out before web-based text is suitable for corpus analysis. This may be done manually, although it is better if an automatic means can be found. Manual stripping out was carried out on the *NHSvax* corpus and on the parts of the *JABS* corpus which were compiled manually. The *JABS Cleanup* program

stripped out boilerplate text. Other aspects of cleaning, such as converting html codes into plain text, were carried out once the corpora had been compiled and the wordlists had been drawn up. This process is described in Section 4.6 below.

### 4.2.2 Balance and representativeness

It is accepted that a corpus should be representative of the language, or variety of language, the description of which it is designed to enable (McEnery, Xiao and Tono, 2006: 5). To be adequately represented, a corpus must be *balanced*. The notion of balance is explained in the following way:

> A corpus is 'balanced' when the size of its subcorpora (representing particular genres or registers) is proportional to the relative frequency of occurrence of those genres in the language's textual universe as a whole.
>
> (Leech, 2007: 136)

However, the extent to which a corpus can ever be truly balanced is debatable. As Sinclair (2005a: 81) observes, corpus building 'is an inexact science and no one knows what the ideal corpus would look like'. In any case, the notion of balance that Leech (2007: 136) has in mind applies more to general corpora, that is, those which represent a language or language variety. Like many CADS studies, the current study uses a specialised corpus, the *JABS* corpus, with a specially-constructed reference corpus, *NHSvax*. Specialised corpora are, by their very nature, unlikely to be balanced (McEnery, Xiao and Tono, 2006: 5). The aim in constructing the *JABS* corpus was to capture, as far as was feasible, the entire contents of the JABS site in a given time frame. The corpus, therefore, consists of varying proportions of different text types. The aim in compiling the *NHSvax* corpus was similar, although more than

one website was chosen, specifically MMR The facts,[34] NHS immunisation,[35] and the vaccination pages from the Health Protection Agency (HPA) website.[36] There is greater homogeneity in the texts, since they are all produced by the same institution, the DoH, but there are, nonetheless, different text types included in the corpus, from DoH reports to FAQ sections. However, since the *NHSvax* corpus is used to provide a norm against which the *JABS* corpus content can be compared, questions regarding the extent to which they are comparable are relevant. These are addressed in the next sub-section.

### 4.2.3 Using a reference corpus: questions of comparability

CADS research is essentially comparative. It is accepted that the linguistic features a researcher uncovers can only be effectively interpreted if they are compared against a norm (Partington, 2008: 99-100). In this study, the *NHSvax* corpus has been compiled as a reference corpus against which the lexico-grammatical features of the *JABS* corpus can be interpreted. Core issues related to the use of reference corpora concern the representativeness of each corpus, the homogeneity of each, their comparability and the reliability of the statistical measures used (Rayson and Garside, 2000). As discussed above, the issue of representativeness, that is, 'the extent to which a sample includes the full range of variability in a population' (Biber, 1993) is of greater relevance in studies concerned with comparing two or more language varieties than it is to the present study. The issues of homogeneity and comparability need closer examination, however, since these have a bearing on the extent to which one can accurately assess the similarity between two corpora (Kilgarriff, 1997). If one text type is over-represented in one corpus, the results may be skewed. One must also

[34] http://www.mmrthefacts.nhs.uk.
[35] htttp://www.immunisation.nhs.uk.
[36] http://www.hpa.org.uk.

compare like with like. In the case of the *JABS* and *NHSvax* corpora, it is evident that we are dealing with corpora which contain varying proportions of different text types. As mentioned in Chapter One, both corpora contain documents on vaccination. But a very large proportion of the *JABS* corpus data comes from the JABS website discussion forum. The *NHSvax* corpus has nothing similar. However, this study is concerned with the use of discursive resources in the *JABS* corpus and uses the *NHSvax* corpus as a means of establishing the typical patterns of mainstream medical-scientific discourse about vaccination. Cross-corpus comparison provides a starting point for the analysis, but does not comprise the entire analytical approach adopted here.

The issue of the comparability of the two corpora also involves consideration of their relative sizes. There are two ways in which two (or more) corpora can be compared: either each corpus is compared with a large, normative corpus (for example, a corpus of general English such as the *BNC*); or two (or more) roughly equally sized corpora are compared with each other (Rayson and Garside, 2000). When two corpora are compared with each other, it does not matter how large a reference corpus is, so long as it is not too small (Berber-Sardinha, 2000). Results remain more or less constant if the reference corpus is larger than the target corpus, but they are less reliable if the reference corpus is less than five times the size of the target (ibid.). It is not unusual in CADS studies to compare differently sized corpora. The corpora that Baker and McEnery (2005) compare consist of 138,505 and 265,224 running words respectively, while Seale *et al.* (2010) compare a 2,536,799 word corpus of online discussion forum postings with a 1,119,106 word corpus of interview data. However, the discrepancy in size between the *JABS* and *NHSvax* corpora is greater. At 4,502,813 running words, the *JABS* corpus is approximately seven times the size of the *NHSvax* corpus, which

has 610,727 running words. Furthermore, *WordSmith Tools* works on the assumption that the reference corpus is larger than the target corpus (Berber-Sardinha, 2000). The problem of comparability is overcome by comparing each corpus with the *BNC* corpus. The *BNC* corpus is therefore used as the reference corpus to produce keyword lists for the *JABS* and *NHSvax* corpora. The results are then cross-compared.

The decision to use the *BNC* corpus as a general reference corpus was based on the fact that the software program chosen for manipulating the data was *WordSmith Tools* (Scott, 2004; 2012). The *BNC (World* Version*)* wordlist is available from the *WordSmith Tools* website[37] and configured for use with the *WordSmith* Keywords function. The *BNC* is a balanced corpus, as far as any corpus may be said to be balanced, but the drawback with using it is that its construction ceased in the mid-1990s. Its contents are somewhat dated. There are, for example, very few terms in the *BNC* relating to electronic media. However, it still represents a rich resource, particularly for the description of the most frequent items in English (Kennedy, 2007: 152). Its use in this study for compiling keyword lists means that statistically salient terms related to the MMR vaccine are brought to light in both corpora. Of course, the keyword lists are likely to include items which are statistically significant largely on account of their relative rarity (or absence) in the *BNC*. Items which relate to the field of electronic communication, such as *www, http, email,* and so on, are likely candidates. But this project is interested in examining how concepts such as risk, causation, and so on, are expressed. As Kennedy (ibid.) argues, the representation of core vocabulary items in the *BNC* is reliable today. *WordSmith Tools* version 4 was used at first, but in 2013, an upgrade to *WordSmith*

[37] www.lexically.net/downloads/version4/downloading%20**BNC**.htm

*Tools* version 6 was necessary for technical reasons. Wordlist and keyword calculations conducted using *WordSmith Tools* version 4 were repeated using version 6.

### 4.2.4 Corpus mark-up

#### 4.2.4.1 Storing contextual information

Ideally, relevant contextual information is stored separately from the text files which make up the corpus, so that the integrity of the corpus texts is not compromised (Sinclair, 2005b). Each corpus file in the *JABS* and *NHSvax* corpora was given an identification title and this information was stored in a spreadsheet. The discussion forum data was gathered using automatic means, though, and potentially useful contextual information, which was included as boilerplate text, was stored by *JABS Cleanup* in file headers, as suggested by Sinclair (2005a: 82). As regards the *JABS* discussion forum data, the term 'header' might be a little misleading. Each file in the discussion forum sections of the *JABS* corpus corresponds to one thread and each thread is composed of at least one post. The beginning of each file was marked with a file header and the title of the thread was recorded in it. The end of each file was marked up in a footer. It was useful to mark the boundaries of each post in order to retain a sense of cohesion when examining longer stretches of text or entire files. To this end, headers and footers were included at the beginning and end of each post. Contextual information stored in the headers of each post included the name of the person posting, their location, the total number of posts they had made on the forum, and the date and time of the post. As is usual practice, contextual information was marked up using angle-bracketed tags. *WordSmith* settings can be set to ignore information which is marked up with tags, so that contextual information can be omitted from the process of building a wordlist or calculating keywords or collocational significance, for example.

#### 4.2.4.2 Part-of-speech (POS) tagging

A contentious issue in corpus linguistics relates to the question of whether a corpus should be parsed. It is argued that the inclusion of part-of-speech (POS) tags enriches the corpus and, importantly, aids sense disambiguation (Leech, 2005: 17). Unfortunately, POS tagging software programs are somewhat inconsistent. Even the most accurate, it is said, 'struggles to achieve much more than 79 percent accuracy on unrestricted text' (McEnery and Hardie, 2012: 156). But a more serious problem with POS tags is that their use implies the application of a pre-existing theory of language. Using POS tags in corpus analysis means that the 'the description which produces the tags in the first place is not challenged – it is protected' (Sinclair, 2004: 191). It is for this reason that POS tagging is not used in corpus-driven corpus linguistics. The *JABS* and *NHSvax* corpora have not been tagged for the same reason. Having considered the principles of corpus compilation, we now discuss ethical considerations.

## 4.3 Ethical considerations

The public-private nature of much online interaction means that, for researchers using data gathered from internet sources, taking decisions as to whether or not permission needs to be sought is not as clear-cut a procedure as with, for example, telephone conversations. The common assumption that all internet texts are created in the public domain is challenged by the increasing prevalence of secure network domains (Berry, 2004: 323). Indeed, the very concept of a clear distinction between 'public' and 'private' spheres is contested (ibid.: 324). People posting on publicly available online discussion forums 'cannot always be assumed to be "seeking public visibility"' (Eysenbach and Till, 2001: 1104). Furthermore, some participants post sensitive information about themselves, thus raising the question as to

whether it is ethical to use data from online discussion forums without consent (Herring, 1997: 19). At the same time, insisting on obtaining consent in all instances would undermine academics' ability to undertake online research (Herring, 1997: 22). In order to determine whether or not consent is required, Eysenbach and Till (2001: 1104) suggest that the researcher should decide the extent to which the participants on a given site are likely to regard it as a private or public space. For example, users are more likely to consider a site a private space if registration is needed to access the site or if the number of users is relatively small (1000 as opposed to ten subscribers, for example). Most importantly, they suggest, 'the perception of privacy depends on an individual group's norms and codes, target audience and aim, often laid down in the "frequently asked questions" or information files of an internet community' (ibid.). The FAQ section on the JABS forum states that 'registration is not required to view current topics on the Forum; however, if you wish to post a new topic or reply to an existing topic registration is required' (http://www.jabs.org.uk/forum/faq.asp, accessed 7/05/2015). One can therefore reasonably assume that members posting on the forum are likely to be aware that the material is accessible to a vast number of people. Furthermore, the forum has a large number of members. Membership currently stands at 948[38] and, at the time at which the *JABS* corpus was compiled, there were 935 members, of which 920 had posted on the forum. Following these considerations, when compilation of the *JABS* corpus started, it was decided that there was no need to seek consent.

Informed consent is just one aspect of the ethical considerations which need to be taken into account when using data from online discussion forums. Another area for consideration concerns the use of biographical details which might help identify a user. It is common

[38] http://www.jabs.org.uk/forum/, accessed 7/05/2015.

practice on internet discussion forums for forum moderators to advise users to adopt a 'chat nickname' and to avoid posting information which might help identify family members. For example, Skea *et al.* (2008), in their analysis of interaction on the Mumsnet site, found that all users in their data set had adopted a chat nickname and had used codes when referring to family members. They therefore did not need to carry out an anonymization procedure. This was not the case with Seale's (2006) study of interaction on cancer support group forums and message boards. Nonetheless, it was not deemed necessary to anonymize the corpus of texts used in this study, instead, Seale (ibid.) anonymizes only those potentially identifying features which occur in the examples he cites in his article. He explains:

> Because these are open access public forums, postings were considered to be in the public domain for ethical purposes although prominently biographical details have been concealed in any quotations used in this report.
>
> (ibid.: 348)

The majority of JABS forum members use a nickname, however some use their full name and surname. In the posts themselves, there are instances where forum participants refer to their own children by name or include biographical details about other JABS members (for example, they address them or refer to them by their real names). It is not possible to carry out full and accurate anonymization of such occurrences using automated means. A process of manual compilation was trialled but found to be inefficient, given the time constraints of the project. Furthermore, the replacing in the text of names with codes was found to compromise textual cohesion on account of the sheer length of some of the forum threads, the large number of participants in long threads, and the fact that the asynchronous nature of discussion forum interaction disrupts normal conversational coherence. It was therefore decided to use non-anonymized data for the purposes of this thesis. However, following Seale (2006),

biographical details will be anonymized in any of the quotations used in any future publications which draw on *JABS* corpus data. The discussion now describes the process of compiling the *JABS* and *NHSvax* corpora.

## 4.4 Compiling the *JABS* and *NHSvax* corpora

### 4.4.1 The *JABS* corpus

4.4.1.1 Description of the data source: the JABS website

As explained in Chapter One, the data for the *JABS* corpus was gathered from the JABS website. Figure 4.1 below shows a screenshot of the JABS website homepage as it appeared in 2008, during the early stages of data gathering.[39] Then as now, the top left-hand corner of the homepage bears the name of the organization *JABS: Justice Awareness and Basic Support* with the banner *Welcome to the support group for vaccine-damaged children* in the top centre and right-hand area. The central and right-hand area of the screen displays an article, either one which has been specially authored for JABS by 'guest' writers, or, more usually, one which is copied and pasted from a news publication.[40] The menu on the left of the screen links to other parts of the site. What is not clear from the homepage is that the discussion forum is divided into three sub-fora: Experience and Support (for seeking advice and exchanging stories); News and Comment (for discussing topical issues); and the Parlour for 'chit-chat'). Since the screenshot shown in Figure 4.1 was taken in 2008, there is no link to the message board, which was discontinued in September 2006.

[39] The layout of the site was updated in early 2013.
[40] Preferred sources were the *Daily Mail* and the *Express*. Examples of titles are *MMR Killed My Daughter* and *Doctor's MMR Fears.*


**Figure 4.1** The JABS homepage, 2008

Figure 4.2 below gives a schematic overview of the organization of the website, as it was in 2006, with a brief explanation of the contents of each section of the site.

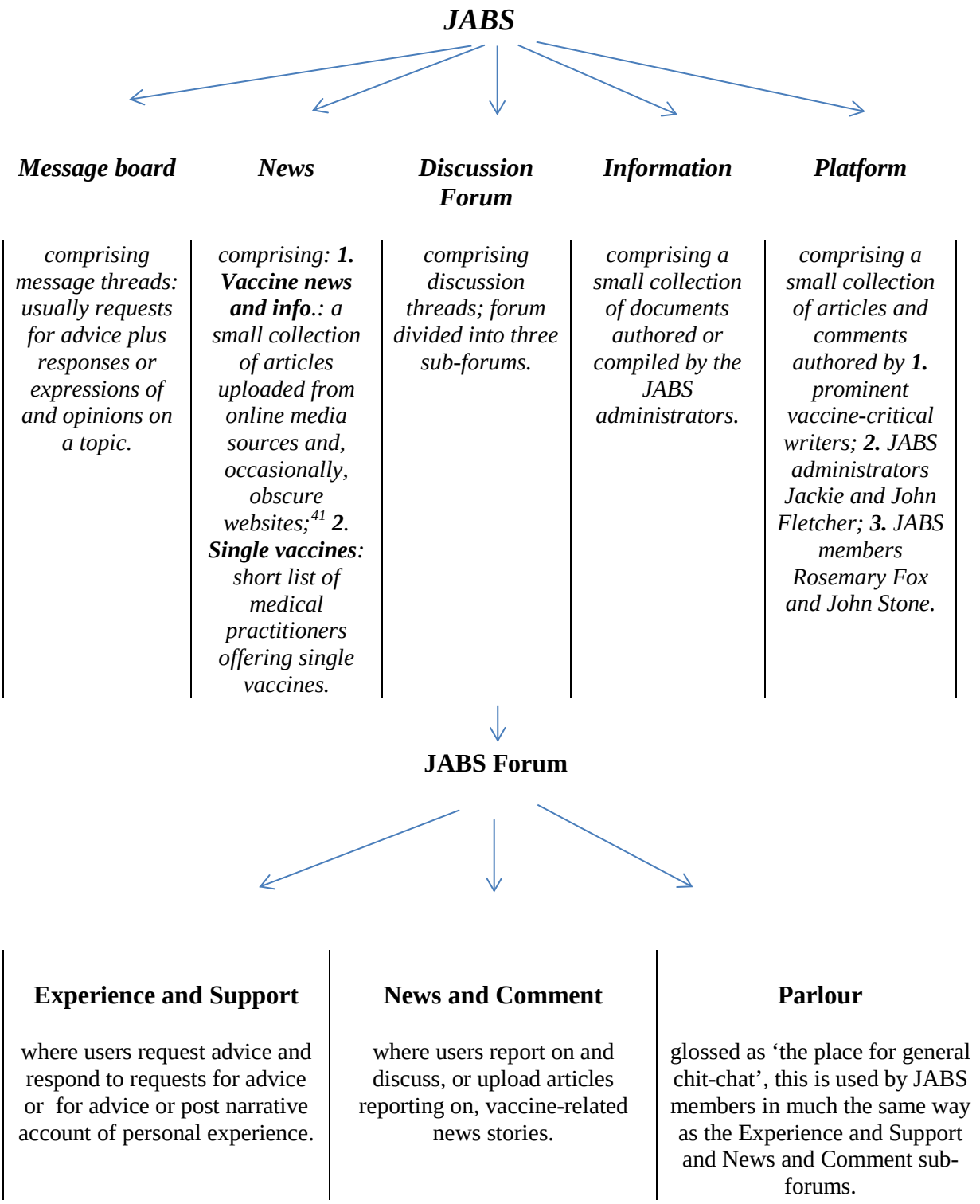


**Figure 4.2** Schematic representation of organization of contents of JABS website, as it was in 2006.

[41] Preferred sources mainly represent the right wing of the British press: *The Daily Mail*, *The Sunday Express*, *The Times*, *The Telegraph*, and *The Scotsman*.

#### 4.4.1.2 Compiling the *JABS* corpus

As explained already, in the early stages, compilation was conducted manually. Manual compilation began in March 2006 and continued until the end of December 2008. The contents of the Information, Platform and News sections were copied and pasted into text files Notepad + +. Some lost data from the Message Board was retrieved using the Internet Archive Wayback Machine. [42] Unfortunately, the data returned was incomplete. Data was available for the years 2001 to 2004 and for 2006, but not for 2005. Message Board interaction typically consisted of requests for advice and responses. As is typical of Message Board interaction, more than one conversation takes place at any one time and 1st and 2nd part adjacency pairs are often separated by several unrelated messages. An example of a request for advice and an adjacent (but unrelated) response to a previous request is shown in Figure 4.3 below.

[4.1] Message received on 9/26/2006 from: :
Where I can find out more about the US stats and reported side effects for the newly introduced vaccine for children (Pneumococcal conjugate vaccine). Many thanks, Beth

[4.2] Message received on 9/26/2006 from: :
Dear Slugabedz my son had the MMR in the Febuary and in the April he came out in a rash which the GP suspected was Measles hope this helps

**Figure 4.3** A request for advice and a response to a request on the JABS Message Board

Data retrieved from the Message Board was copied and pasted into Notepad + + text files. The number of words of data per year captured by the Internet Archive Wayback Machine from the JABS Message Board is shown in Table 4.1 below.

[42] http://www.archive.org/web.

| JABS Message Board data | |
|---|---|
| **Year** | **Number of words** |
| 2001 | 12,601 |
| 2002 | 152,380 |
| 2003 | 33,814 |
| 2004 | 60,464 |
| 2005 | - |
| 2006 | 3,532 |
| **Total** | **262,791** |

**Table 4.1** Number of running words of text per year from JABS message board collected for inclusion in **JABS** corpus

As is clear from Table 4.1 above, only a small amount of data was retrieved from the message board. A greater proportion of the data relates to 2002 to 2004 than to the other years. It is impossible to say to what extent this is an accurate representation of the relative amount of traffic the message board received in each of these years, although it is clear that the lack of data for 2005 is unrepresentative. Little cleaning of texts from the Information, Platform, News and Message Board sections of the JABS website was necessary at this point. Where it was necessary to strip out boilerplate text, the 'Find and Replace' function on Notepad + + was used.

In order to render more reliable the process of compiling the Forum sections of the corpus, the *JABS Cleanup* program was used, which was available for use in its final version in 2010. All threads on the forum, from its inception in 2005 up until 31 December 2008, were downloaded. When run, the *JABS Cleanup* application goes directly to the *JABS* forum home-screen and displays the forum content within the *JABS Cleanup* interface (see Figure 4.4 below).

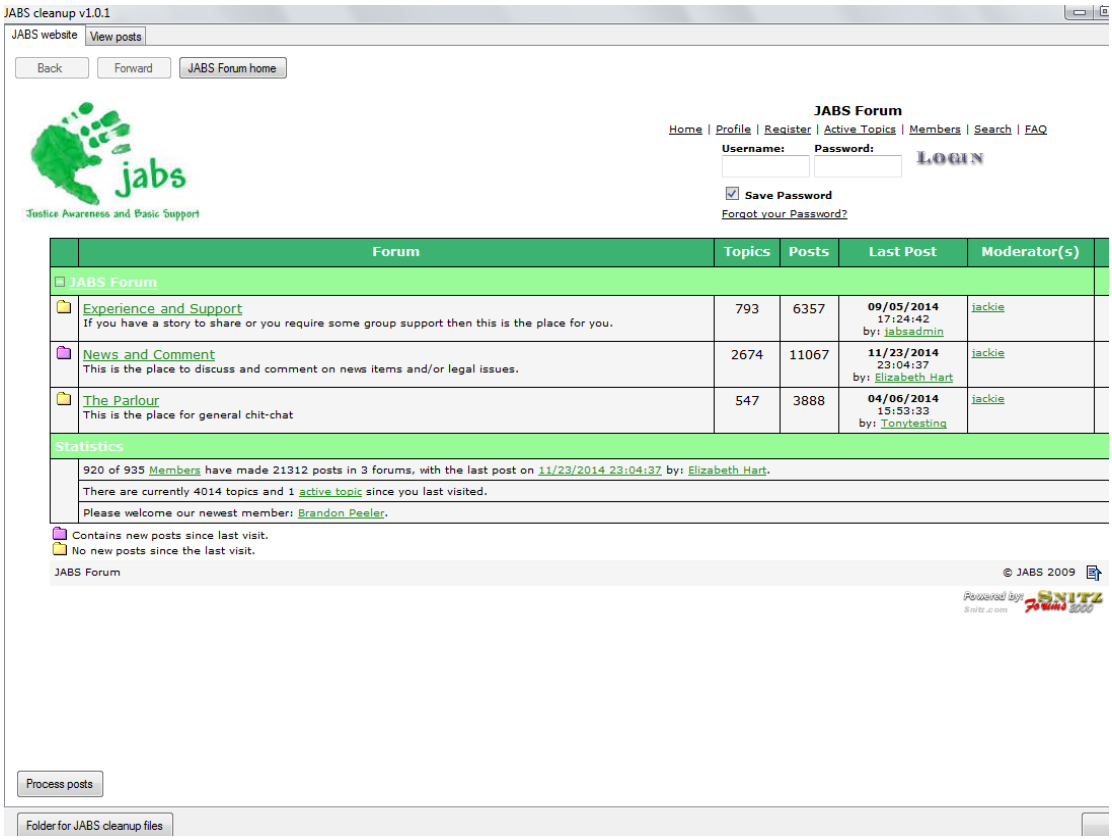


**Figure 4.4** JABS Forum homepage

The user is able to navigate around the forum in the usual way, by clicking on the appropriate links. Once a thread is open, the user clicks on the “Process posts” button and the contents of the entire thread are extracted and saved in a text file (see Figure 4.5 below).

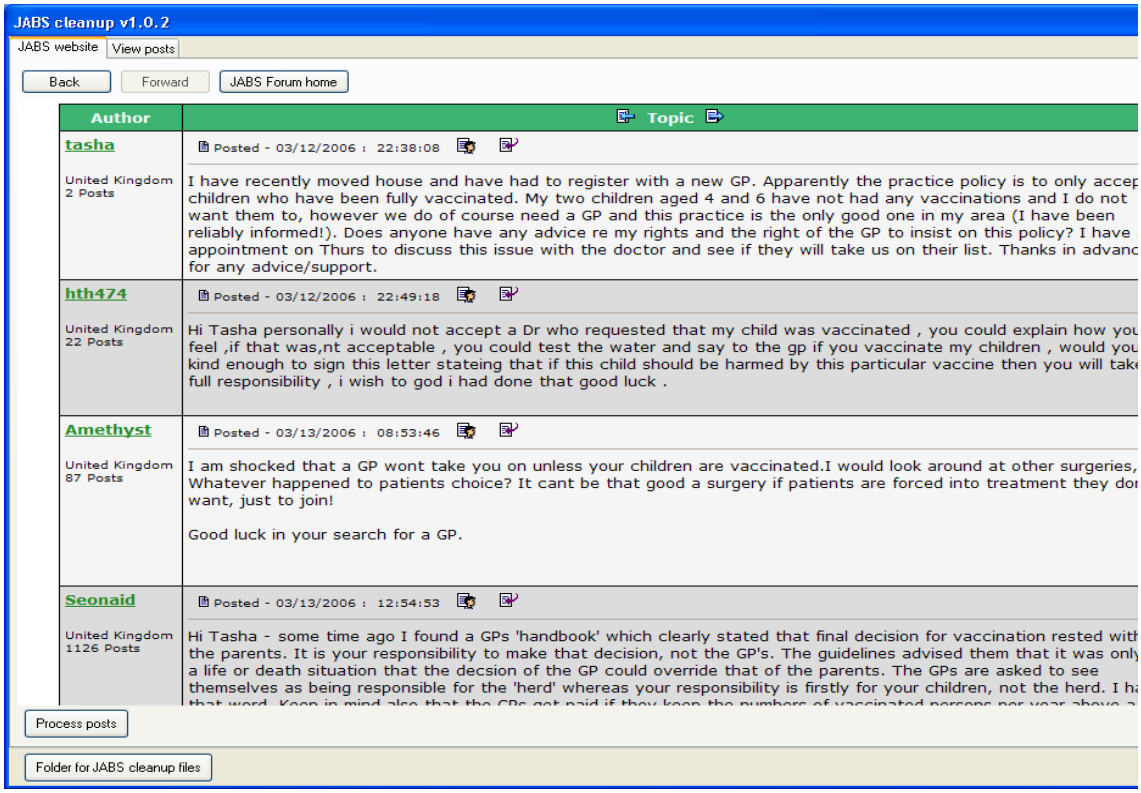


**Figure 4.5** An example of part of a JABS discussion forum thread displayed in the *JABS Cleanup* interface

The *JABS Cleanup* program strips out boilerplate text and converts the remaining content into plain text format, displaying contextual information (poster's name, date and time of post, and so on) in headers (see Figure 4.6 below).

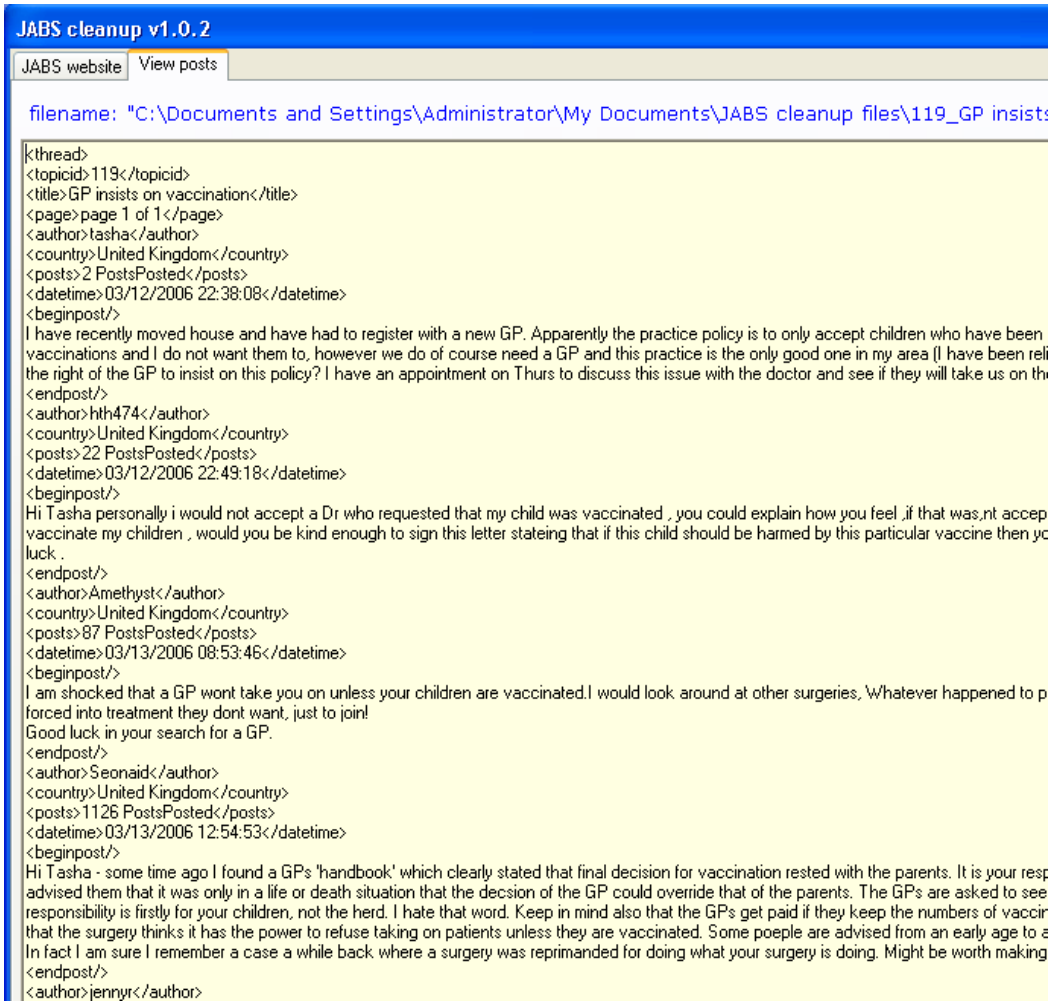


**Figure 4.6** An example of part of a thread following conversion by JABS Cleanup program into text file format

The content is saved in a designated folder and given an ID number and name. The name assigned to the file corresponds to the title of the thread, for example, ‘GP insists on vaccination’ or ‘Adverse Event to the triple’. The data is saved as a text file and is therefore ready for use in a concordancer. As Figure 4.7 below shows, contextual information is recorded between angle-bracketed tags. The tags in the headers mark up the beginning and end of the thread, the number of pages of the thread and the current page number, the ID number of the file, and the title of the thread. Header information is also used to record the

name of each poster, their location (where this information is supplied), the number of posts they have posted in total on the website, the date and time of each post, and the beginning and end of each post.

<thread>
<topicid>533</topicid>
<title>Adverse Event to the triple</title>
<page>page 1 of 1</page>
<author>Tara r</author>
<country>United Kingdom</country>
<posts>3 PostsPosted</posts>
<datetime>01/09/2007 20:16:58</datetime>
<beginpost/>
If anyone out there has had, or heard of similar experiences I'd be very pleased to hear about them. My twin sons were delivered pre-term at 36 weeks by elective cesaer. The hospital insisted on giving them T.B jabs before I left the hospital. I was VERY concerned, having reacted to vaccines myself (reaction to the typhoid was pretty bad, went blind, passed out, vomited etc). I was told it was regulations as we live in an area where sadly TB cases are happening. At 2 months and 4 months they had parts one and two of the triple comprising Diptheria/Tetanus/Whooping Cough, polio, Hib and Men C. 10 days after the 2nd part one of the twins developed encephalitis and nearly died, he was in Great Ormond Street ICU for two weeks. They never got to the bottom of it, conclusively. But the doctors did say that perhaps it would be safer not to continue the vaccination programme with my children, given family history and the fact that they couldn't say DEFINITELY that it wasn't the vaccine that caused his illness. Did once read on the World Health Organisation Website that at least three of the triple elements can cause encephalitis and DEATH in 1 in 100,000 cases, and that reactions can be upto 15 days after the jab. Anyone else had a similar experience?
<endpost/>
<author>Hannah</author>
<country>United Kingdom</country>
<posts>84 PostsPosted</posts>
<datetime>01/09/2007 20:30:29</datetime>
<beginpost/>
Hi Tara, sorry to hear one of your boys had such a bad reaction. My two eldest sons both reacted to this vaccine although this was back in 1997 and 1999 so it was a different one then, but it was not this sort of reaction. It was an eczema/asthma/allergies reaction. As a result, they haven't had any more vaccines and my youngest son is completely un-vaccinated. At least the doctors advised you not to continue vaccinating your sons, as they usually try to cover up any vaccine damage link - if anything bad happens after vaccination, it is just cooincidence. However I think it was very bad that they co-erced you into accepting the TB jab - no vaccine is compulsory in this country and you would have been fully within your rights to refuse this vaccine.
Make sure that your son's reaction is noted in the yellow card reporting system. If enough cases are reported, it might help save another child's life.
Hannah
<endpost/>
<author>Tara </author>
<country>United Kingdom</country>
<posts>3 PostsPosted</posts>
<datetime>01/09/2007 20:47:15</datetime>
<beginpost/>
Hannah, thanks for your reply and comments - its nice to make contact with another parent who has had bad experiences and concerns about these particular vaccines. I think you are right that the doctors have previously tried to hide any evidence, rather than report it, because my reaction to the whooping cough as a baby, and EXTREME reaction to typhoid jab were not reported, even though I lost consciousness and had to spend an hour recovering at my doctors surgery and then driven home - they didn't report it. I was 18 at the time and didn't follow it up. Could kick myself.
<endpost/>
<endpage/>
</thread>

**Figure 4.7** Example of text file from *JABS* discussion forum sub-corpus

*WordSmith Tools* enables the user to use the tags to select, if desired, only specific parts of the files for inclusion in calculations and display in concordances. In this study, it was decided to include only information which users had written in their posts. Therefore, only the text between the tags marked <title> and </title>, and <beginpost/> and <endpost/> was selected for compilation of the *JABS* wordlist and to be displayed in concordances.

A noted feature of discussion forum interaction is that a post may include sections of prior text (Richardson, 2001). On the *JABS* forum, intertextual references of this kind are usually marked up. In order to clearly signal where a poster's words were reproduced by a subsequent poster, *JABS Cleanup* converted the tags used on the JABS website interface to mark the beginnings and ends of quoted sections into the tags <q> and </q> (see Figure 4.8 below).

```
<endpost/>
<author>Cybertiger</author>
<country>United Kingdom</country>
<posts>976 PostsPosted</posts>
<datetime>06/16/2008 07:59:41</datetime>
<beginpost/>
<q>Originally posted by Rosemary
â¬ÜThe balance of benefits to risks remains overwhelmingly positive.â¬"</q>
Except for the ones who die, presumably. But then they died from 'coincidence', an
overwhelmingly negative cause of death.
<endpost/>
```

**Figure 4.8** A forum post with quoted section of prior poster's text marked up

This was only partially successful. Sections of text were often repeated in a series of posts, becoming embedded in other sections of text which themselves got repeated. There were frequently quoted sections within quoted sections. Although *JABS Cleanup* was successful at marking up simple instances of intratextual repetition, it failed to cope with the more complex instances of embedded citations.

The *JABS* corpus data was divided into seven sub-corpora, each sub-corpus corresponding to the different sections of the website. There was so much discussion forum data, that the three forum sub-corpora were each divided into three further sub-corpora. These contained, respectively, data for 2005-06, 2007 and 2008. The size of each sub-corpus in the *JABS* corpus is shown in Table 4.2 below. Table 4.2 shows the number of files in each sub-corpus, the number of running words identified by *WordSmith*, and the number of tokens used by *WordSmith* to compile the wordlists.[43] Contextual information contained in the headers was not counted. As can be seen in Table 4.2, the majority of the *JABS* corpus data originates from the discussion forum. Of the 2,101 files in the corpus, 1,988 represent discussion forum data. In terms of the number of running words, the discussion forum data accounts for over 86% of the corpus: 3,648,368 out of the 4,277,663 running words in the corpus and 3,439,561 of the 4,010,213 tokens used to compile the wordlist. The number of threads posted on the News and Comment section of the website is greater than that in the other two sections. The News and Comment sub-corpora together contain 1,279 files and comprise 2,007,367 running words (with 1,895,749 used for the wordlist) in contrast to the Experience and Support and Parlour

[43] A calculation of the number of running words in a corpus is a calculation of the number of tokens in the corpus. From this point of view, the terms *running words* and *tokens* are synonymous. When compiling a word list, *WordSmith* is able to use most, but not all, of the tokens it counts. Therefore, a distinction is made between the total number of running words in the corpus and the number of tokens used to compile the wordlist.

sub-corpora which contain, respectively, 351 files (with 836,468 running words and 784,348 tokens used for the wordlist) and 358 files (with 804,533 running words and 759,464 tokens used for the wordlist). Posts on the News and Comment sub-forum frequently contain uploaded texts, most often news articles, in the body of the text.

| ***JABS* corpus contents** | | | | |
|---|---|---|---|---|
| **Sub-Corpus** | | **Number of files** | **Number of running words** | **Number of tokens used to compile wordlist** |
| Information | | 26 | 62,276 | 60,503 |
| Platform | | 49 | 269,679 | 235,240 |
| News | | 33 | 36,851 | 35,703 |
| Message Board | | 5 | 250,489 | 239,224 |
| Discussion Forum 2005-2006 | Experience and Support | 93 | 140, 868 | 132,347 |
| | News and Comment | 129 | 212,060 | 200,906 |
| | Parlour | 34 | 49,550 | 46,409 |
| Discussion Forum 2007 | Experience and Support | 101 | 112,967 | 105,145 |
| | News and Comment | 265 | 342,173 | 322,738 |
| | Parlour | 52 | 112,325 | 105,651 |
| Discussion Forum 2008 | Experience and Support | 157 | 582,633 | 546,856 |
| | News and Comment | 886 | 1,543,134 | 1,327,105 |
| | Parlour | 272 | 642,658 | 607,404 |
| **TOTAL** | | **2,101** | **4,77,633** | **4,010,231** |

**Table 4.2** Total number of files, running words, and tokens used for the wordlist in each sub-corpus of the *JABS* corpus

The high proportion of text in the News and Comment sub-corpora gives an indication of how high a proportion of news texts make up the *JABS* corpus discussion forum data. Possible solutions to the challenges of distinguishing between the voices of JABS members and the many voices represented in uploaded news texts are discussed in Section 4.4 below.

### 4.4.2 The *NHSvax* corpus

#### 4.4.2.1 The data sources

##### *4.4.2.1.1 The 'MMR The facts' website*

The *NHSvax* corpus contents were gathered from the 'MMR The facts' and 'NHS immunisation' websites, and from the vaccination related pages of the HPA website.[44] A screenshot of the 'MMR The facts' homepage, taken in 2008, is shown in Figure 4.9 below.

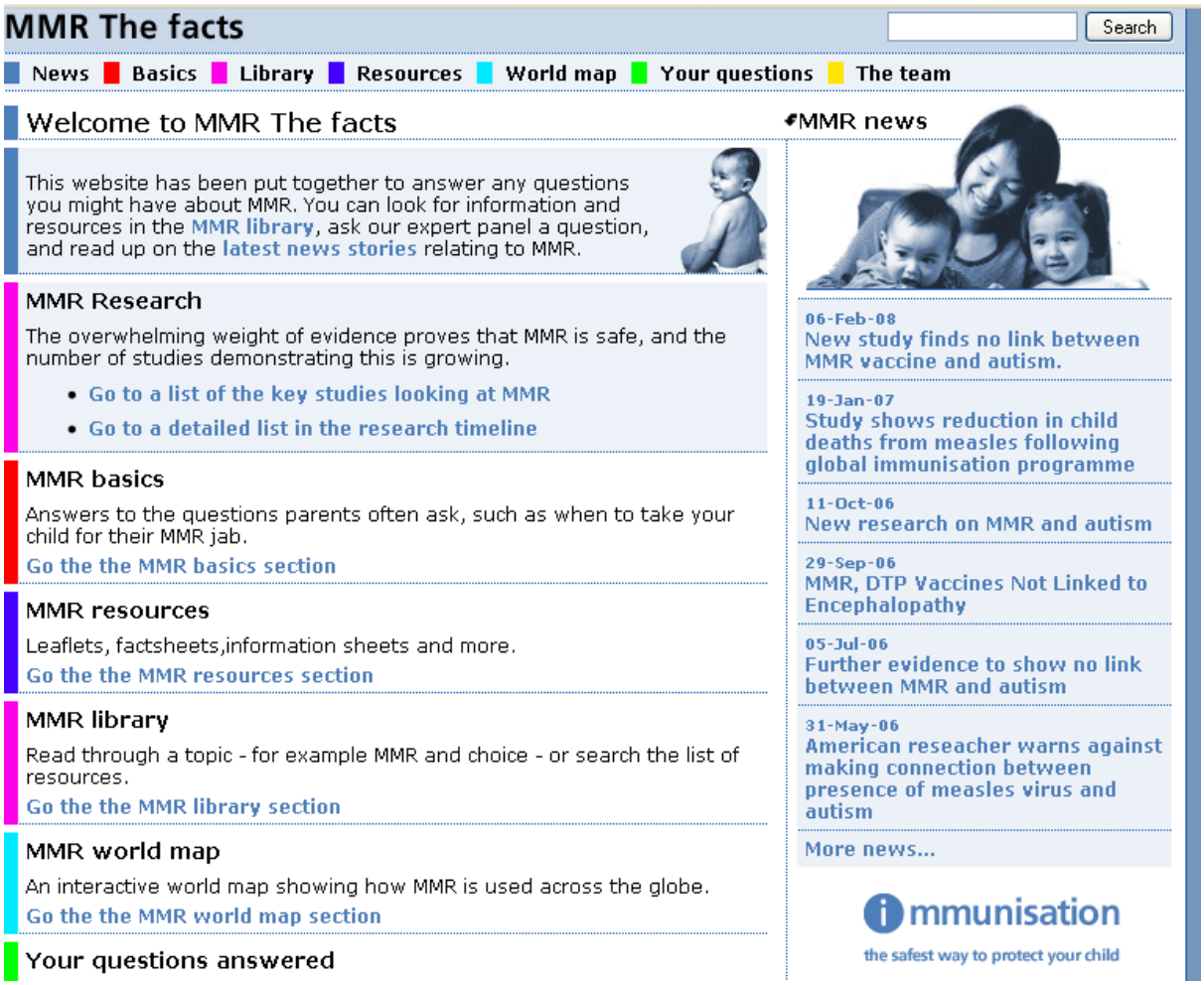


**Figure 4.9** Screenshot of 'MMR The facts' homepage, 2008.

[44] http://www.mmrthefacts.nhs.uk; http://www.immunisation.nhs.uk; http://www.hpa.org.uk.

The ‘MMR The facts’ website was aimed purely at parents. Users of the site were able to access the various sections of the site either via the links on the menu bar at the top of the page or via the links on the left-hand menu. The right-hand menu contained links to summaries, written specifically for the website, of recent news articles and academic articles. A brief summary of the type of content in each section on the website is given in Figure 4.10.

| **MMR research** | **MMR basics** | **MMR resources** | **MMR library** | **MMR news** | **MMR world map** |
|---|---|---|---|---|---|
| containing two, differently organized, bibliographic lists of studies into MMR | in essence an FAQ section, in which questions from the public are reformulated and standardized (Harvey, 2012). | containing electronic versions of the information leaflets routinely made available in clinics and general practice surgeries. | containing short discussions of some of the main topics of debate connected with MMR, for example, the link between MMR and autism or the issue of patient choice. | containing summaries and discussions of the findings of relevant scientific research papers with bibliographic references supplied; also summaries of and/or links to reports by public bodies, such as the WHO or the DoH. | a map showing use of MMR around the world, with brief description. |

**Figure 4.10** Description of contents of sections of ‘MMR The facts’ website

*4.4.2.1.2 The ‘NHS immunisation’ website*

The ‘NHS immunisation’ site differed from ‘MMR The facts’ in that it offered information on all vaccines provided under NHS provision, whether childhood vaccines or vaccines intended mainly for adults (for example, the influenza vaccine for people over 65). It also offered information for health professionals. The sections headed ‘About immunisation’, ‘The

vaccines' and 'FAQs' (see Figure 4.11 below) contained information about vaccines, the diseases against which they offer protection, and childhood vaccination schedules, as well as documents in which issues about the safety and risks of vaccination were discussed. These sections were aimed at lay people. The 'Professional Information' section was, as the name suggests, aimed at health professionals and contained DoH policy documents, reports and briefings. These texts were mainly addressed at heath visitors and nurses and, to a lesser extent, at general practitioners. The 'Library' section contained factsheets and was aimed both at professionals and lay people.

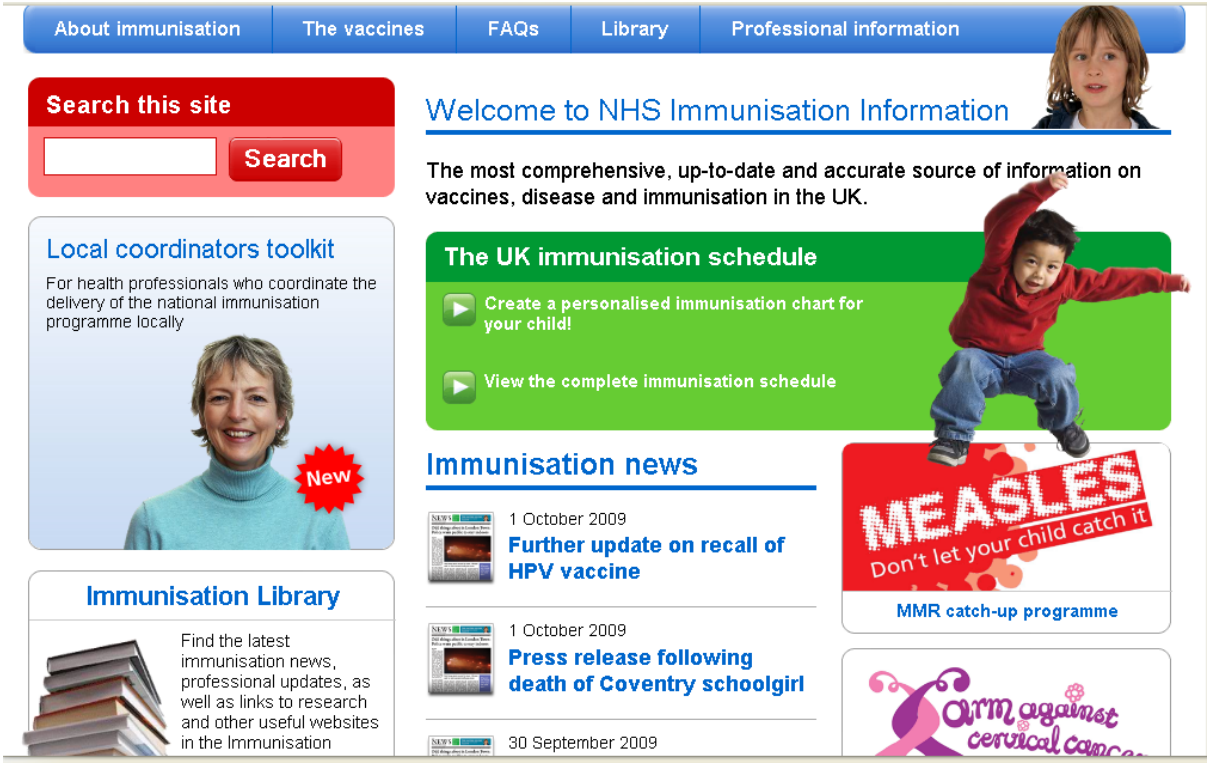


**Figure 4.11** Screenshot of 'NHS immunisation' homepage, 2008.

*4.4.2.1.3 The HPA immunisation pages*

The immunisation section of the HPA website consisted of one introductory page (see Figure 4.12 below) with links to fourteen pages, each with a brief description of a different disease and its vaccine. The site is aimed at health professionals and lay people.

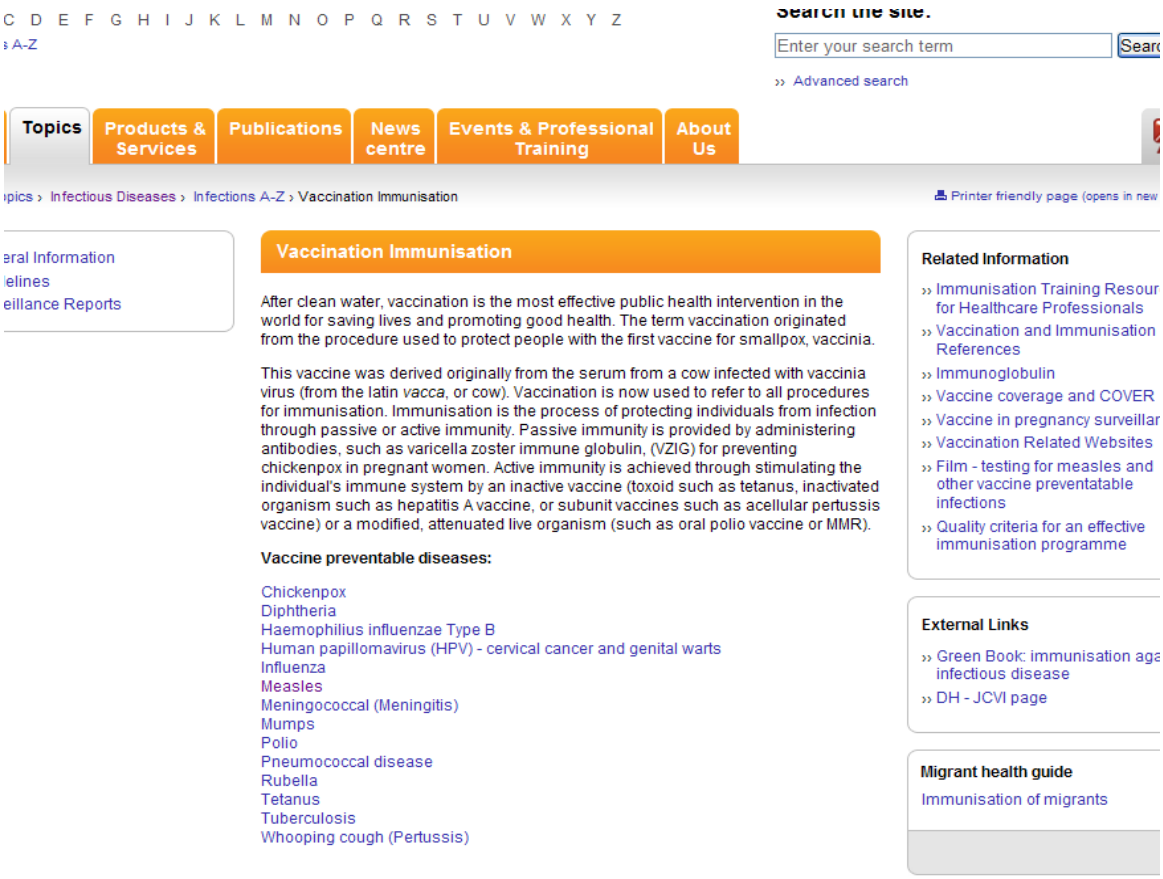


**Figure 4.12** The introductory page of the HPA website immunisation section

4.4.2.2 Compiling the *NHSvax* corpus

As with the JABS corpus, compilation took place between 2006 and the end of 2008.Textual content was copied and pasted into text files Notepad + +. What little cleaning was necessary was done using the Search and Replace function of Notepad + +. Wherever it was practicable,

organization of the *NHSvax* data into sub-corpora was mapped onto the layout of the websites. Some of the sections on the sites yielded little data. The data from the 'MMR The facts research' section is included in the *MMR The facts* library sub-corpus while the 'MMR The facts world map' and the file downloaded from the homepage was included in the *MMR The facts basics* sub-corpus. The NHS immunisation data was organized into three sub-corpora. The *NHS immunisation general information* sub-corpus contains all the data downloaded from the homepage, the 'About immunisation', the 'Vaccines' and the FAQs sections. This is the information aimed at lay people. The data downloaded from the 'Library' section, which contained documents aimed at a lay and professional audience, is stored in the *NHS immunisation library* sub-corpus, while the *NHS immunisation professional information* section contains the data from the 'Professional information' section. Only 15 files were downloaded from the HPA website. They are contained in the *HPA* sub-corpus. The size of each sub-corpus in the *NHSvax* corpus is shown in Table 4.3 below.

As Table 4.3 shows, out of the three sources, the 'NHS immunisation' site yielded the greatest quantity of data: 146 files and 479,474 running words (with 457,631 tokens used for the wordlist); as opposed to 126 files and 127,766 running words (with 121,820 words used for the wordlist) in the *MMR The facts* sub-corpus; and 15 files with 3,487 running words in the *HPA* sub-corpus. If we look at the corpus contents in terms of the proportion of texts aimed at lay people, health professionals or both, we find that 152 corpus files, consisting of 197,878 running words (190,267 of which were used to compile the wordlist) represent texts aimed at a lay audience alone. Meanwhile 62 corpus files, comprising 271,687 running words (with 258,212 tokens used in the wordlist) represent texts written for health professionals and 78 corpus files, comprising 141,162 running words (and 134,334 tokens used in the wordlist)

represent texts for a mixed audience. A greater proportion of corpus context is therefore aimed at health professionals than at lay people alone. However, since some texts are aimed at both lay and professional people, roughly half the corpus content represents texts written with a lay audience in mind. However, whether specific texts are intended for a lay or professional audience, none of them are academic articles. The science represented in these texts can clearly be considered to be represented in a popularized form. What is important, though, is that they reflect the voice of the mainstream medical-scientific community.

| ***NHSvax* corpus contents** | | | |
|---|---|---|---|
| **Sub-corpus** | **Number of files** | **Number of running words** | **Number of tokens used for the wordlist** |
| MMR The facts: news | 88 | 40,827 | 38,655 |
| MMR The facts: library | 16 | 32,534 | 30,989 |
| MMR The facts: resources | 22 | 45,020 | 42,743 |
| MMR The facts: basics | 5 | 9,385 | 9,433 |
| NHS immunisation: general information | 21 | 70,112 | 68,447 |
| NHS immunisation: library | 63 | 137,675 | 130,972 |
| NHS immunisation: professional information | 62 | 271,687 | 258,212 |
| HPA | 15 | 3,487 | 3,362 |
| **Total** | **292** | **610,727** | **582,813** |

**Table 4.3** Total number of files, running words, and tokens used for the wordlist in each sub-corpus of the *NHSVax* corpus

Having described the process of compiling the *JABS* and *NHSvax* corpora, we now move on to discuss the principles and practices of corpus interrogation, in particular, the use of statistics in corpus linguistics, and procedures such as keyword and concordance analysis.

## 4.5 Corpus Analysis: Methodological considerations

### 4.5.1 Using statistics in corpus linguistics

Statistical measures of significance are standardly used in corpus linguistics to make statements about the significance of collocational pairings (Church and Hanks, 1990) or, more recently, to identify words that are "key" in a text or corpus (Scott, 1997). The argument for using statistics is that any statement about the frequency of a given item in a corpus is only meaningful if it is normalized, in other words, if it is compared against a norm. The measures of significance which are used in corpus linguistics have been borrowed from the discipline of statistics and rely on calculations of probability. Unfortunately, most probability calculations, for example, Mutual Information (MI), z-score and chi-square, rely on the assumption that events are distributed randomly and, when plotted on a graph, form a standard bell curve. Because of the constraints of the grammatical system, language is never random (Kilgarriff, 2005). This raises serious questions about the appropriacy of using statistical measures of probability which presuppose randomness to calculate the significance of events in a non-random system such as language (ibid.). However, the fact remains that some form of statistical measurement is desirable in corpus linguistics if one is to make judgements about the saliency of particular features. The statistical measures currently available are not wholly suitable for language analysis, but, given that none are available which do not presuppose random distribution, they are the best we have. Provided sufficient data is examined, such measurements can bring to light significant co-occurrences (Kilgarriff, 2005).

The statistical measures used in corpus linguistics fall into two main types. The type which relies on the assumption that events are random and, when plotted on a graph, will form a bell-curve, such as MI, z-score and chi-square, are best for identifying compound nouns or other strong collocational pairings. The other type is exemplified by t-score and log-likelihood. These test a one-way collocational relationship and works by taking standard deviation into consideration. These measures have the effect of suppressing high frequency items, such as grammatical words, and bringing to the fore those items which occur with greater than random probability. Unlike statistical measures such as MI and z-score, which have been found to be particularly unreliable when word frequencies are low (less than five occurrences), statistical measures such as log-likelihood have been found to perform well whether word counts are high or low (Dunning, 1993). To calculate keywords, *WordSmith Tools* offers log-likelihood and chi-square as means for calculating statistical significance. For calculating the strength of collocational relations, it offers two variations of MI (MI3 and Specific Mutual Information), z-score, and log-likelihood. Following Dunning's (ibid.) observations on the reliability of log-likelihood, this measure was chosen as the statistical measure with which to calculate the keywords. Log-likelihood, MI3, Specific Mutual Information and z-score were all tested on the *NHSvax* and *JABS* data in order to see which performed best when calculating collocational relationships. It was found that, in most instances, MI3 performed best: it was more successful than log-likelihood at suppressing grammatical items but less prone than Specific Mutual Information and z-score to privileging very rare items. MI3 was chosen as the measure with which to calculate collocational significance. However, z-score proved to be the most efficient means for highlighting the

significant collocates of very high frequency items. In order to identify the reporting verbs and status nouns which have a strong collocational relationship with *that*, z-score was used.

### 4.5.2 Wordlists and keywords

Comparison of the relative frequencies of lexical items or recurrent patterns in a corpus is central to the quantitative approach of corpus linguistics (McEnery and Hardie, 2012: 2). Comparison of the frequency lists of two comparable corpora is useful in bringing to light significant differences between the texts comprising each (Hunston, 2002: 67). Keyword analysis programs, such as that included in *WordSmith Tools*, offer a faster and potentially more reliable way of identifying the significant features of a specialised corpus than reading through wordlists. However, the information produced by a keyword program is only as reliable as the statistical measure used to calculate the results and, as discussed above, the statistical measures used in corpus linguistics are not ideally suited to the peculiarities of language and are of variable reliability. It is for these reasons that corpus analysis should ideally start with comparison of raw frequency data before keyword analysis is undertaken. In this way the reliability of results can be assessed. Of course, a complicating factor when using keywords in CADS is that the discourse analyst is primarily concerned with the encoding of social meaning, in other words, with *keywords* as understood by Williams (1976). Corpus keyword tools highlight words which are statistically significant. It is therefore up to the CADS analyst to identify among the statistically significant keywords, candidate terms which are likely to yield culturally significant meaning.

A further problem with using statistical keywords as a point of departure was discussed in Chapter Three. It was highlighted that keyword calculation is based on the single orthographic unit rather than the extended phraseological unit. An alternative is to draw up a frequency list based on clusters, or n-grams, with a view to bringing to light recurrent phraseological units. There are some drawbacks with using clusters in this way, though. Firstly, a software program cannot distinguish meaningful phraseological units from other strings. Secondly, a cluster program will find thousands of clusters in all but the smallest of corpora. In the preliminary stages of corpus analysis for this project, the *AntConc* concordance program was used to draw up lists of three- to six-word clusters in the *JABS* corpus. There were 93,100 three-word clusters with a frequency of five or more, and 14,456 six-word clusters with a frequency of five or more. Many of the top two hundred three-word clusters and most of the top eight to nine hundred six-word clusters corresponded to elements encoded in the headers of the *JABS* forum files. Finally, programs which provide lists of clusters do not calculate the degree of significance of what they find. It is therefore difficult for the analysts to assess the extent to which a given phraseological pattern occurs recurrently in the data because it is a marked feature of the corpus under examination or because it is a frequent feature of language use in general. For these reasons, it was decided to use keyword analysis in this study to identify salient areas of meaning, but to bear in mind that subsequent analysis should aim at uncovering recurrent phraseological patterns. Stubbs (2010) argues that a corpus-driven approach to linguistic analysis with a focus on the encoding of semantic and pragmatic meaning in extended lexical units, as described by Francis (1993) and others, is key to uncovering culturally salient meanings. As discussed in Chapter Three, analysis of lexical keywords is useful for highlighting key themes in a specialised corpus (Hunston, 2002: 68; 2008), while analysis of grammatical keywords can bring to light repeated phraseological

patterns which form the building blocks of epistemological knowledge (Groom, 2007). Keyword calculations can throw up an extremely large number of keywords, though. The keyword calculation of the *JABS* corpus, using the *BNC World* version as reference corpus, identified over 1,600 positive keywords, for example.[45] Following Millar and Budgell (2008), it was decided to ignore any keyword which occurred in fewer than 10% of the texts in its respective corpus. The *WordSmith WordList* tool not only counts the frequency with which each type occurs in a corpus, it also counts the number of corpus files a type occurs in and then calculates the percentage of corpus files this represents. *JABS* and *NHSvax WordList* statistics were used to identify those words in the keyword list which occur in fewer than 10% of corpus files. This limited the keyword lists to a more manageable size and ensured that keywords of central importance were identified.

### 4.5.3 Concordance, collocation and clusters

The basic tool in corpus linguistics is, of course, the concordance. Examination of a concordance alone, especially once it is sorted, may be sufficient to bring to light frequently occurring patterns. When dealing with a lot of data, though, computational means can be used to aid analysis. *WordSmith Tools* includes functions for calculating the statistical significance of the collocates of a node word within a window of up to five words either side of the node. It also has a 'clusters' function. Clusters are calculated using concordance lines. The clusters function displays, in order of frequency, clusters of between two to six words found within a span of between five and twenty-five (depending on the settings) words either side of the node. Uncovering significant collocates or frequently occurring clusters can help narrow

[45] *WordSmith* also calculates *negative keywords*, that is, words which occur in the corpus with significantly less frequency than random probability would predict. Since the *JABS* keyword list was to be compared with the *NHSvax* keyword list, it was considered unnecessary to calculate negative keywords.

down a search. As Hunston (2002: 52) points out, a researcher will have difficulty in coping with more than 100 concordance lines if he or she is interested in general patterns in language. When examining detailed patterns, thirty concordance lines will suffice. Information regarding significant collocates or clusters can help the analyst to drill down into the data and choose specific phraseological patterns for more detailed analysis. Alternatively, one can use a small selection of lines to test hypotheses about particular linguistic patterns (ibid.).

### 4.5.4 Analysing contextual information

#### 4.5.4.1 Forum posters: frequency and ideological stance

As Partington (2008) observes, one way in which CADS methods deviate from the standard Sinclairean approach to corpus analysis is that the CADS analyst acquaints him- or herself as much as possible with the data. The elements of the *JABS* corpus data which demanded extra attention were the discussion forum data. A particularly salient feature concerned variations in the frequency with which different individuals posted on the site. Many only post once or twice but some are highly frequent posters. It was important to discover if there was a difference in the motivations between low-frequency and moderate- or high-frequency posters. There was also evidence that certain high-frequency posters were not vaccine-critical. In 2007 and 2008, the JABS forum attracted a small but vociferous number of members who expressed an overtly pro-vaccination and pro-science stance.[46] The pro-science posters were banned from the site at the end of 2008. While they were active on the site, their contributions attracted a lot of response, so the threads in which they participated tended to become very long. The arguments between vaccine-critical and pro-science posters on the forum were

[46] A number of these individuals were readers of Ben Goldacre's *Bad Science* blog http://www.badscience.net/ and one, who calls herself Becky Fisseux, had her own blog: 'JABS Loonies - Justice, Awareness, Basic Support and Mind Blowing Stupidity', http://jabsloonies.blogspot.co.uk/.

interesting from the point of view of discovering how people on the site constructed their arguments. It was therefore necessary to read sections of the data online in order, firstly, to find out how posts were distributed among posters, and, secondly, to discover whether low- and moderate- to high-frequency posters had different motivations for posting, and, thirdly, to distinguish the pro-science posters from the vaccine-critical ones.

There are 517 posters represented in the *JABS* forum sub-corpora. Most of them, 278, to be precise, are based in the UK, eighteen are located in the USA, five in Ireland, four in Canada, and two in New Zealand.[47] As tends to be the case with online forums, the majority of posts are made by a minority of posters, with a large number of posters making only one post. The 1,989 threads represented in the JABS forum sub-corpora comprise 20,250 posts. As Table 4.4 below shows, just over half the posters (279) make only one or two posts and the vast majority (406) make ten or fewer.

| **No. of posts** | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | **Total** |
|---|---|---|---|---|---|---|---|---|---|---|---|
| **No. of posters** | 200 | 79 | 52 | 18 | 19 | 12 | 4 | 11 | 3 | 8 | **406** |

**Table 4.4** Number of posters posting one to ten posts

Individuals who post only once typically ask for advice or information. These posters tend not to be overtly vaccine-critical, although the fact that they are posting on the forum indicates that they are certainly sceptical about the safety of the vaccines and resistant to accepting the

[47] In 210 cases, location information is not given.

advice or directives of health professionals uncritically. The example shown in Figure 4.13 is typical of this kind of request. The poster expresses uncertainty: *I am unsure what to do for the best; I really don't know what to do; I feel so confused*. She also expresses a lack of trust in her GP, since she questions what the GP has told her: *Does anyone know if this is true?* Instead she directs her request at anyone on the JABS forum who might read her post: *Can anyone help?*

<thread>
<topicid>97</topicid>
<title>5 IN 1 JAB.</title>
<page>page 1 of 1</page>
<author>1415victoria</author>
<country>United Kingdom</country>
<posts>1 PostsPosted</posts>
<datetime>02/26/2006 22:32:25</datetime>
<beginpost/>
Hi,
My Daughter is due to have the 5 in 1 this week, and i am unsure what to do for the best. I have been putting it off for so long(she is now 5 months), that my Doctor has told me if i dont hurry up and make my mind up then it will be too late to give it to her, as you can only give it up to 11 months. Does anyone know if this is true? Also my son who is now 3 yrs, had the old baby jab minus the whooping cough, but ive told that he can't have it that way now when he has his booster. I really dont know what to do. I was so sure when it came to the MMR(he did not have it). I feel so confused.
Can anyone help?
Victoria.
<endpost/>

**Figure 4.13** Example of a request for advice from a low-frequency poster

By contrast, the ten most frequent posters collectively make over 10,000 posts (see Table 4.5 below). Most posts made by moderate- to high-frequency posters on the forum (those who make more than three posts) are not requests for advice but observations and expressions of opinion or responses to the observations and arguments of others. Most moderate- to high-frequency posters are overtly vaccine-critical.

| | Name | Number of posts | | Name | Number of posts |
|---|---|---|---|---|---|
| 1 | rosemary | 2,019 | 6 | cybertiger | 976 |
| 2 | gus the fuss | 1,447 | 7 | truth seeker | 957 |
| 3 | john stone | 1,254 | 8 | aasa | 724 |
| 4 | seonaid | 1,126 | 9 | occam48 | 691 |
| 5 | jabsadmin | 987 | 10 | minorityview | 611 |
| **Total** | | | | | **10,792** |

**Table 4.5** Number of posts by the ten most frequent posters on the JABS forum.

As is typical of reformist vaccine-critical groups, many of the moderate- to high-frequency posters claim that they are not opposed to vaccination per se, but believe vaccines can cause neurological problems. Such a stance is illustrated in the example shown in Figure 4.14 below.

<thread>
<topicid>60</topicid>
<title>a thought</title>
<page>page 1 of 1</page>
<author>Andrea</author>
<country>United Kingdom</country>
<posts>28 PostsPosted</posts>
<datetime>01/10/2006 11:23:03</datetime>
<beginpost/>
We believe our daughter was made severely deaf by MMR. However with regards to autism I have come across several parents whose children have been diagnosed/being diagnosed with a disorder on the autistic spectrum and the parents are adamant that it was not vaccine related as they saw certain signs prior to the MMR vaccine being administered at about 13 months. This I can understand but it seems to me that many people seem to think MMR is the only vaccine to cause problems. I'm not sure that they have even considered the earlier/new born jabs to be a concern and that if these caused a problem then the side effects will be seen before 13 months. I am not anti-vaccine and I believe that vaccines have a place in society if used with great caution and with fully informed choice.
We can only do our best
<endpost/>

**Figure 4.14** An example of a JABS forum post from a moderate frequency poster

As we can see in Figure 4.14, the author states that she is 'not anti-vaccine', but believes MMR is responsible for her daughter's deafness and suggests that other vaccines administered in early life may cause autism. A large number of threads, especially those posted on the News and Comment sub-forum, are initiated by a poster drawing attention to a story in the news and, very often, uploading the story into the post. An example of the beginning and end of such an initial post and a response to the initial post is given in Figure 4.15 below.

<thread>
<topicid>3105</topicid>
<title>Baby vaccine could increase the risk of asthma</title>
<page>page 1 of 1</page>
<author>jabsadmin</author>
<country>not given</country>
<posts>987 PostsPosted</posts>
<datetime>10/21/2008 11:50:49</datetime>
<beginpost/>
http://www.dailymail.co.uk/health/article-1079259/A-vaccine-given-babies-increase-risk-childhood-asthma.html
Daily Mail 21 October 2008
A vaccine given to babies could increase the risk of childhood asthma
By Beezy Marsh
Last updated at 12:12 AM on 21st October 2008
A vaccination given to babies has been linked to asthma.
Experts believe the diphtheria, tetanus and whooping cough jabs might provoke an immune system response which predisposes the body to the lung condition. […] A Department of Health spokesman said 'Several large studies have looked at whether childhood vaccines can cause asthma or allergies - they have found no evidence for this.'
<endpost/>
<author>Truth Seeker</author>
<country>United Kingdom</country>
<posts>957 PostsPosted</posts>
<datetime>10/21/2008 12:49:45</datetime>
<beginpost/>
This is why Department of health Spokesmen are desperate to cover up any report of vaccine damage at all costs by blanket denials and even refuse to look at any evidence like the above.
Any study you like as long as we are the ones doing it, what a joker.
Vaccinations cause chronic diseases is as fully an accurate things to say as "smoking kills" warnings on cigarette packs.
<endpost/>
<endpage/>
</thread>

**Figure 4.15** Part of a thread from the News and Comment sub-forum

Most high-frequency posters are vaccine-critical, but, interestingly, the pro-science posters are among the most frequent posters. The names, number of posts, and ideological stances of all individuals with 40 or more posts are shown in Table 4.5 below.

| | **Name** | **No of posts** | **Vaccine-critical** | **Pro-science** | | **Name** | **No of posts** | **Vaccine-critical** | **Pro-science** |
|---|---|---|---|---|---|---|---|---|---|
| 1 | rosemary | 2019 | √ | | 32 | maxwell | 105 | | √ |
| 2 | gus the fuss | 1447 | √ | | 33 | scotmum | 103 | √ | |
| 3 | john stone | 1254 | √ | | 34 | jackie | 99 | √ | |
| 4 | seonaid | 1126 | √ | | 35 | jacquia | 97 | √ | |
| 5 | jabsadmin | 987 | √ | | 36 | amethyst | 87 | √ | |
| 6 | cybertiger | 976 | √ | | 37 | hannah | 84 | √ | |
| 7 | truth seeker | 957 | √ | | 38 | lee77c | 71 | | √ |
| 8 | aasa | 724 | √ | | 39 | pariah | 66 | √ | |
| 9 | occam48 | 691 | | √ | 40 | steve | 65 | | √ |
| 10 | minorityview | 611 | √ | | 41 | jabba | 64 | √ | |
| 11 | john | 590 | √ | | 42 | elga | 61 | √ | |
| 12 | joan | 534 | √ | | 43 | louise | 61 | √ | |
| 13 | aobbard | 430 | √ | | 44 | jenson | 59 | | √ |
| 14 | suba | 393 | √ | | 45 | jack hep | 57 | √ | |
| 15 | janet | 348 | √ | | 46 | prurient | 56 | | √ |
| 16 | jennyr | 381 | √ | | 47 | quail | 56 | | √ |
| 17 | laura_c_a | 337 | √ | | 48 | an ot | 55 | | √ |
| 18 | thomas p | 314 | | √ | 49 | informed | 53 | √ | |
| 19 | becky fisseux | 287 | | √ | 50 | veeg | 53 | | √ |
| 20 | whatif | 286 | √ | | 51 | phoenix | 52 | | √ |
| 21 | elizabeth | 241 | √ | | 52 | do not vaccinate | 48 | √ | |
| 22 | squib | 213 | √ | | 53 | bill2b | 46 | √ | |
| 23 | emerald | 201 | √ | | 54 | govna | 43 | √ | |
| 24 | barefoot1 | 187 | √ | | 55 | sam | 43 | √ | |
| 25 | commonsense | 178 | | √ | 56 | deej | 42 | | √ |
| 26 | lola | 173 | √ | | 57 | mrscat | 42 | √ | |
| 27 | allison | 170 | √ | | 58 | cfish | 41 | √ | |
| 28 | angladrion | 152 | √ | | 59 | deepika | 41 | | √ |
| 29 | fionas | 137 | √ | | 60 | aquamarine | 40 | √ | |
| 30 | justsayno | 133 | √ | | 61 | Missyc1 | 40 | √ | |
| 31 | wanda | 127 | √ | | 62 | Sarahwhale | 40 | √ | |

**Table 4.5** Posters with 40 or more posts, number of posts, and stance towards vaccination

As mentioned above, contributions from pro-science posters provoke a lot of response from vaccine-critical members. Threads to which pro-science posters contribute tend to run to

several pages. Figure 4.16 below shows a post from a pro-science poster. The degree of animosity she attracts from one of the high-frequency vaccine-critical posters is evident in the quoted section of text in her post.

<author>Becky [redacted]</author>
<country>United Kingdom</country>
<posts>287 PostsPosted</posts>
<datetime>07/09/2008 16:37:44</datetime>
<beginpost/>
<q>Originally posted by GUS THE FUSS
Dear Becks one observation of you you defend the Lucifer himself Sir(to you) Roy Meadows hardly balanced are you ,wheres your people evidence Becky anon the fishy fish...to refute Dr Martin??(wont hold my breath) or in your case your breath has to be worse than your bite...</q>
MMR RIP
Gus,
I've never defended Cybertiger's bÃªte noire Prof Sir Roy Meadows - and frankly I see no reason why it would be relevant here, even if I had.
I'm not just suggesting that there aren't "stealth viruses" in vaccines - I'm suggesting that it's rather suspicious that such an important development in our understanding of viruses has only ever been discovered by one researcher, and that no-one has managed to replicate his work. Neither has he put such important work up for peer-review. You have to ask yourself why. This isn't just a vaccine matter, it's potentially a massive step forward in the study of viruses. You can't simply use the old "oh, he knows it wouldn't be worth putting it forward for peer review because every other scientist in the world has been bought off by big pharma" argument, as there would be someone prepared to do the work, and, if Dr Martin's right, to back him up. "Oh, I'm so scared of the conspiracy..." just doesn't wash.
Follow the money.
Kind regards,
Becky
<endpost/>

**Figure 4.16** Example of a post from a pro-science poster

#### 4.5.4.2 Using expanded context

As discussed already, it is sometimes necessary to read extended context around a particular concordance line. This is especially the case where a section of text is quoted and one needs to see how it is framed or when a news text is included and one needs to scroll to the top of the article to discover how it is framed or how the person posting it has framed it. In

*WordSmith Tools*, by clicking on a concordance line, the researcher can access the entire text file in which the line occurs. This was done frequently throughout analysis.

## 4.6 Procedure

Once the *JABS* and *NHSvax* corpora had been compiled and sections of the data had been read in order to gain a general idea of the nature of the contents, the following steps were taken.

- Frequency lists of the *JABS* and *NHSvax* corpora were drawn up. *WordSmith Tools* settings were adjusted to exclude header information from the word count.
- In order to complete the cleaning of corpus data, instances of html codes in the word list were noted. These were converted to plain text using *WordSmith Tools Utilities Text Converter*.[48]
- Following conversion of html codes, frequency lists of the *JABS* and *NHSvax* corpora were drawn up afresh. The number of tokens in the corpora and their various sub-corpora were recorded from the *WordSmith Tools* statistics.
- Keyword lists of the *JABS* and *NHSvax* corpora were compiled using the *BNC World* version wordlist as a reference corpus. Log-likelihood was used as the statistical measure of probability. The parameters were set as follows: Log-likelihood threshold 24; maximum p value 0.000001; maximum number of keywords, 6,000; negative keywords were excluded.[49] In order to limit the number of keywords to a manageable quantity but still capture the most central ones, the minimum number of occurrences

[48] The symbol #, which is used by *WordSmith* to represent numerals, proved impossible to remove but, fortunately, it does not corrupt any lexical items in the corpus.
[49] *WordSmith* allows the option to draw up a list of negative keywords, that is, words which occur markedly infrequently when compared with a reference corpus. Since the *JABS* corpus keyword list was to be compared with the *NHSvax* corpus keyword list, it was not considered necessary to draw up a list of negative keywords.

for keywords was set at 15 for the *NHSvax* corpus, and, since the *JABS* corpus is 6.8 times the size of *NHSvax*, the minimum frequency for *JABS* was set at 100 occurrences.

- Analysis began with an examination of the most frequent words and statistically most significant keywords in the *NHSvax* corpus, in order to gain an idea of the key features of mainstream medical-scientific discourse about health, risk, immunity, and the MMR-autism causal hypothesis.
- The 200 most frequent words in the *NHSvax* corpus were scanned in order to identify key themes and to provide a means for gauging the reliability of the keyword list.
- The keyword list was scanned and cross-compared with the *NHSvax* word list. In order to limit the number of items selected for closer investigation to a manageable quantity, all keywords which, according to the *NHSvax* wordlist statistics supplied by *WordSmith Tools*, occurred in fewer than 10% of *NHSvax* text files were excluded.
- The remaining keywords were grouped into semantic and functional categories and these categories were further arranged into sub-categories. Some of the semantic categories contain both lexical and grammatical items on the grounds that there is, to a certain extent, overlap between semantics and grammar (for example, word *following* can be used as a verb, an adjective and a preposition; conjunctions are used to indicate meaning relations; prepositions such as *following, after,* and so on, suggest temporal relations). A number of semantic categories were identified and named on an ad-hoc basis.[50] Where it was not immediately apparent how a word should be categorized, concordance lines were examined in order to see the typical context of use. Words

[50] For example, several nouns were found which denoted diseases, types of vaccines, medical conditions, and so on, and a number of nouns, verbs and adjectives were found which related vaccination and healthcare in general or to risk and causation.

which were found to be used in more than one way were categorized according to the more frequent use.[51] The following reference works were consulted to aid the categorization of some items: Halliday's (1994) *Introduction to Functional Grammar*; the *Collins COBUILD English Dictionary* (Sinclair et al., 1995); *the Collins COBUILD Grammar Patterns 1: Verbs* (Francis, Hunston and Manning, 1996); the *Collins COBUILD Grammar Patterns 2: Nouns and Adjectives* (Francis, Hunston and Manning, 1998).[52]

- In order to render the task of examining the corpus even more manageable, a short-list of keywords was drawn up for close examination. The suitability of the members of each sub-category of words as candidate terms for closer examination was assessed on the basis of the likelihood that analysis of a given term might usefully help fulfil the research objectives of the thesis. Concordances were consulted to aid assessment and a list of candidate terms was drawn up.
- Concordances for each of the candidate keywords were examined in detail. Where there was a lot of data, the *WordSmith Tools* collocation and cluster programs were used to identify the most frequent patterns. MI3 was used as the statistical measure of significance with which to calculate significant collocates. Concordances were examined relating to the collocational patterns and strings identified.
- Analysis continued with an examination of the most frequent words and statistically significant keywords in the *JABS* corpus.

[51] For example, the word *may* was used as an epistemic modal in 86% of occurrences and as a reference to the calendar month in 14% of citations. It was therefore categorized as an epistemic modal.
[52] The semantic categories of those nouns which occur in the N-*that* pattern were identified according to the categorization scheme described by Francis, Hunston and Manning (1998: 108-113) and discussed in Section 3.4.4 of this thesis. Some of the labelling was adapted to reflect more accurately the uses of particular items in the contexts of the *JABS* and *NHSvax* corpora. For example, the group of words labelled by Francis, Hunston and Manning (ibid.) the 'happiness' group (which contains words such as *concerns*) was labelled 'affect' in this thesis.

- The 200 most frequent words in the *JABS* corpus were scanned in order to identify key themes and to provide a way of gauging the reliability of the keyword list.
- To limit the number of candidate terms to a manageable size, keywords with less than 10% distribution across the *JABS* corpus texts were excluded, using the same procedure as was followed during analysis of the *NHSvax* corpus.
- The remaining keywords were then grouped into semantic and functional categories and sub-categories. Categories and sub-categories were identified using the procedure with which the *NHSvax* keywords were categorized.
- The similarities and differences between the *JABS* and *NHSvax* corpus keywords were noted.
- A short-list of terms for close examination was drawn up. As was the case when drawing up the *NHSvax* keyword short-list, the suitability of the members of each sub-category of words as candidate terms for closer examination was assessed on the basis of the likelihood that analysis of a given term might usefully help fulfil the research objectives of the thesis. Concordances were consulted to aid assessment and a list of candidate terms was drawn up.
- Concordances of each of the candidate keywords were examined in detail. Where there was a lot of data, the *WordSmith Tools* collocation and cluster programs were used to identify the most frequent patterns. MI3 was used as the statistical measure of significance with which to calculate significant collocates. Concordances were examined relating to the collocational patterns and strings identified.
- Analysis of concordances and collocational patterns started with the keywords which were identified which related to the causal hypothesis, risk, health and immunity. MI3 was used to calculate the strength of collocational relationships. Typical lexical and

grammatical patterns were analysed, and comparisons were drawn with the patterns observed in the *NHSvax* corpus. Where necessary, expanded context was used to shed light on broader contextual features, such as framing.

- Analysis of the *JABS* data continued with an examination of words identified in the keyword analysis, which were considered useful to uncovering the ways in which intertextual elements are reframed and warrants for expertise are expressed. Lexico-grammatical patterns related to expressions of factivity, to the framing of reported speech, and to the expression of identity were examined. Z-score was used as the measure of statistical significance with which to calculate the significant collocates of *that*, in order to identify the most statistically salient reporting verbs and status nouns.
- Where it was necessary to investigate the prosodies of certain words, their use in the *Bank of English* or the *ukWaC* corpus was examined. *UkWac* was accessed using the *Skylight* concordance at *Skylight*: http://www.skylight-to-english.co.uk/skylight. Where it was desirable to highlight register differences, the *Times, New Scientist* and *British Spoken* sub-corpora of the *Bank of English* corpus were consulted.
- In order to discover the sort of external references which were frequently drawn on by *JABS* writers, the number of references to specific publications and web addresses were recorded.

## 4.7 Conclusion

This chapter has discussed issues related to corpus compilation and analysis and has described the processes of compilation of the *JABS* and *NHSvax* corpora. The next three chapters present and discuss the results. Chapter Five examines the lexico-grammatical patterns most

frequently used to express notions of health, risk, immunity and the causal hypothesis in the *NHSvax* corpus. Chapter Six examines the expression of epistemological claims in the *JABS* corpus, considering the ways in which notions of immunity, risk and causation are expressed in the *JABS* corpus, and comparing and contrasting the findings with the patterns found in the *NHSvax* corpus. Chapter Seven examines the ways in which intertextual elements are reformulated and arguments framed in the JABS corpus.

# CHAPTER FIVE

# MEDICAL-SCIENTIFIC DISCOURSE: EVIDENCE FROM THE *NHSvax* CORPUS

## 5.1 Introduction

One of the objectives of this project is to discover the ways in which authors represented in the *JABS* corpus data express their beliefs about the supposed risks of the MMR vaccine and about health in general and the immune system. It has been suggested that parents often talk about risk, health and the immune system in distinctly different ways from the medical-scientific community. It has also been found that vaccine-critical groups often use medical-scientific discourses of health and risk and reframe them. In order to discover how risk, the MMR-autism causal hypothesis, and notions of health and immunity are commonly talked about in medical, scientific discourse, it is useful to identify the typical lexico-grammatical features which occur in the *NHSvax* corpus and are used to talk about these issues. The aims of this chapter are therefore to identify salient lexico-grammatical features in the *NHSvax* corpus which can be used as a point of comparison when analysing the contents of the *JABS* corpus in the following two chapters.

Section 5.2 examines the most frequent words and statistically significant keywords in the *NHSvax*. Keywords are identified for closer analysis. Section 5.3 presents an analysis of the ways in which notions related to the causal hypothesis are expressed in the *NHSvax* corpus. Section 5.4 examines expressions related to health, the immune system, and disease and

causation. Section 5.5 analyses the ways in which notions of risk are expressed. Section 5.6 summarizes the main conclusions drawn from the features highlighted in the analysis and points forward to the following chapter.

## 5.2 Frequency data and keywords in the *NHSvax* corpus

### 5.2.1 The 200 most frequent words in the *NHSvax* corpus

Appendix 1 shows the 200 most frequent words in the *NHSvax* corpus. As is typical of a small, specialized corpus, a number of lexical words appear unusually high up in the frequency list, compared with what one observes in a larger, general corpus. For the sake of clarity, the thirty most frequent words in the *BNC* and *NHSvax* corpus are shown in Table 5.1 below. While no lexical words appear among the top thirty words in the *BNC*, a number of words related to (childhood) vaccination and health appear among the top thirty in *NHSvax*, namely *vaccine, mmr, immunisation, vaccination, health, children*, and *vaccines*. The word *uk* ranks at 30, owing to the high frequency in the corpus of references to the United Kingdom and of its use in web addresses. However, owing to the lack of interactive data in the *NHSvax* corpus, personal pronouns are absent from among the thirty most frequent words. Many of the remaining lexical words among the top 200 (see Appendix 1) relate to the core business of the NHS websites, which is to supply information about vaccines and about immunization schedules. There are therefore terms such as *information* and *programme*. There are several words referring to diseases and vaccines, for example, *disease(s), measles, rubella, mumps, hib, hpv, flu, meningitis*. There are also words such as *dose* and *booster*, and a number of words relating to children's ages, such as *year(s), months, one, two,* and *age*.

| rank | ***BNC*** | | ***NHSvax*** | |
|---|---|---|---|---|
| | **word** | **frequency** | **word** | **frequency** |
| 1 | *the* | 6055105 | *the* | 31453 |
| 2 | *of* | 3049564 | *#* | 28633 |
| 3 | *and* | 2624341 | *of* | 19112 |
| 4 | *to* | 2599505 | *and* | 14542 |
| 5 | *a* | 2181592 | *to* | 14104 |
| 6 | *in* | 1946021 | *in* | 11744 |
| 7 | *#* | 1604421 | *a* | 9851 |
| 8 | *that* | 1052259 | *is* | 7355 |
| 9 | *is* | 974293 | *vaccine* | 7228 |
| 10 | *it* | 922687 | *for* | 5991 |
| 11 | *for* | 880848 | *that* | 5520 |
| 12 | *was* | 863917 | *be* | 5209 |
| 13 | *i* | 732523 | *are* | 4220 |
| 14 | *on* | 731319 | *or* | 4128 |
| 15 | *with* | 659997 | *this* | 3905 |
| 16 | *as* | 655259 | *mmr* | 3677 |
| 17 | *be* | 651535 | *with* | 3664 |
| 18 | *he* | 593609 | *immunisation* | 3447 |
| 19 | *you* | 588503 | *have* | 3337 |
| 20 | *at* | 524075 | *on* | 3229 |
| 21 | *by* | 513444 | *as* | 3133 |
| 22 | *are* | 458368 | *children* | 3031 |
| 23 | *this* | 454419 | *it* | 3010 |
| 24 | *have* | 448684 | *by* | 2958 |
| 25 | *but* | 446783 | *not* | 2867 |
| 26 | *not* | 431075 | *health* | 2837 |
| 27 | *from* | 425987 | *at* | 2763 |
| 28 | *had* | 413144 | *from* | 2625 |
| 29 | *his* | 410294 | *vaccines* | 2436 |
| 30 | *they* | 376289 | *uk* | 2365 |

**Table 5.1** The thirty most frequent words in the *BNC World* version and *NHSvax* corpus

While it may be useful to examine words which denote specific diseases or vaccines, if we are concerned with uncovering 'non-obvious meaning' (Partington, 2010: 88) and if we are interested in discovering the ways in which causation, risk and immune system function are talked about in the corpus data, other words might prove more productive. In order to bring to the fore words which might offer a more fruitful avenue of enquiry, we need to turn to the keywords.

### 5.2.2 Keywords in the *NHSvax* corpus

All the keywords in the *NHSvax* corpus which occur in at least 10% of corpus files are shown in Appendix 2. Keywords are ranked according to log-likelihood, and raw frequencies and log-likelihood statistics are given. The *BNC (World* version*)* was used as a reference corpus. Since log-likelihood was used as the measure of significance, many of the highly frequent lexical words described above (for example, *vaccine, mmr, immunisation, vaccines*, and so on) have risen towards the top of the list. Appendix 3 shows all of the keywords with 10% or more distribution across corpus files, arranged according to semantic and functional categories. Categories were identified according to the procedure described in Chapter Four. Each category set out in Appendix 3 is divided into sub-categories. Within each sub-category, words are ordered according to their relative statistical significance (log-likelihood). As explained in the previous chapter, a small set of candidate terms was selected from among the keywords for closer examination. The candidate keywords selected are those which are considered most likely to provide useful information regarding the ways in which notions of health risk and immunity are expressed or the ways in which warrants for expertise are expressed. Terms which are considered less central to expressions of these notions are rejected. Terms which are relevant but which have a strong collocational relationship with other terms which have been selected are rejected. The keyword categories are as follows:

1. vaccination and pharmaceutical products;
2. health, disease and medical conditions;
3. people, places and institutions;
4. science, medicine and healthcare;
5. sources
6. likelihood, causation and temporal relations;

7. expressions with modal meaning;
8. conditionality;
9. expressions of evaluation;
10. other nouns, verbs, and phrases referring to events, actions, and states of affairs;
11. other grammatical items;
12. miscellaneous items.

The contents of the first category are shown in Table 5.2.

| | | |
|---|---|---|
| **vaccination and pharmaceutical products** | vaccines | *vaccine*, *mmr*, *vaccines*, *ipv* (inactivated polio vaccine), *dtap* (diphtheria, tetanus and pertussis), *menc* (meningococcal conjugate C vaccine), *bcg* (anti-tuberculosis vaccine) |
| | vaccination | *immunisation, vaccination, immunisations, immunised, vaccinated, vaccinations* |
| | vaccination: pre-modifiers | *fully (immunised)* |
| | vaccination: post-modifiers | *component* |
| | vaccine types | *conjugate, inactivated (polio vaccine), live, oral (polio vaccine), single, combined, multiple, separate* |
| | mode of delivery | *separately* |
| | processes | *manufactured* |
| | other pharmaceutical products | *medicines, products, product, drugs* |
| | vaccine additives | *thiomersal* |

**Table 5.2** The *NHSvax* keywords in the 'vaccination and pharmaceutical products' category

From this category, the words *vaccine, mmr,* and *vaccines* are considered central to fulfilling the research aims of this thesis and so were selected as candidate keywords. Other terms denoting vaccines were not considered relevant to this project. Terms relating to vaccination in general, the pre- and post-modifiers listed above, and the terms *separately* and *manufactured* were not selected because where their use in the corpus is of relevance to this

study only in those instances where they collocate with other, more central, keywords. The sub-category which contains terms which refer to pharmaceutical products (*medicines, product(s)* and *drugs*) is potentially relevant, as is the word *thiomersal.*[53] However, concordance evidence shows that *medicines* and *product(s)* frequently occur in references to the safety and licencing of vaccines or, more usually, to institutional bodies who oversee checks on the safety of vaccines and license them. Examples are:

> [5.1] The vaccines are prepared in hens' eggs and should not be given to individuals with known anaphylactic hypersensitivity to egg **products**.
>
> [5.2] …the Committee on Safety of **Medicines**
>
> [5.3] … the **Medicines** and Healthcare **products** Regulatory Agency
>
> [5.4] … role is to assure the quality of biological **medicines** through a mixture of **product** testing

The word *drugs* is notable solely for the neutral prosody of the environment in which it is used. This is in contrast to its use in non-medical contexts. For example, according to the *Bank of English*, the most significant lexical collocates of *drugs* among the top twenty collocates are:

> *alcohol, use, drugs, anti, taking, illegal, sex, prescription, drink, used, using, take, crime, war, effects, drug.*[54]

With the exception of *prescription*, a clear reference to the use of drugs for medical purposes, *drugs* is associated with alcohol, sex, and illegality. The environment of *drugs* in the *NHSvax*

[53] *Thiomersal* is the British English variant of the more widely used term *thimerosal*. It refers to a mercury-based preservative which is added to some vaccines and which has been suggested as a cause of autism.
[54] The only measures of significance available with the Bank of English are t-score and MI. The measure of significance used here is t-score, which is similar to log-likelihood in that it suppresses grammatical items.

corpus, unsurprisingly, points to the use of the word in healthcare contexts. The lexical collocates among the top twenty collocates are:

> *drugs, antiviral, use, requiring, oral, insulin, diabetes, given, influenza, treatment, prevention, prevent, used, treat, development, immunosuppression, alcohol, organisms.*[55]

Although there is a reference to *alcohol* here, the collocational profile indicates that drugs are viewed as beneficial. A typical example of the use of *drugs* in *NHSvax* is:

> [5.5] Guidance on the use of **antiviral drugs** for the treatment of influenza

Thiomersal has been implicated as a causal factor in the development of autism, hence the 322 citations of *thiomersal* in the NHSvax corpus. As one might expect, instances of *thiomersal* occur in contexts in which the argument that the substance causes harm is rebutted, for example,

> [5.6] The Committee on Safety of Medicines (CSM) **has examined the link between thiomersal and neurotoxicity**. The CSM advised that there is **no evidence of harm caused by doses of thiomersal in vaccines**, except for hypersensitivity reactions (such as allergic skin reactions). There is **no evidence of a link between hypersensitivity reactions and the development of autism**.

While the use of *drugs* and *thiomersal* is relevant to the debate about the safety of MMR, they warrant no further examination here. The items from the first category of words which are selected as candidate terms for closer analysis are therefore *vaccine(s)* and *mmr*.

[55] In order to offer a valid comparison, t-score has been used as the measure of significance.

The items belonging to the 'health, disease and medical conditions' category are shown in Table 5.3.

| | | |
|---|---|---|
| **health, disease and medical conditions** | health and disease: general terms | *disease, health, infection, diseases, infections, illness, illnesses* |
| | health and disease: adjectives | *unwell, ill, healthy* |
| | diseases and infections | *measles*, *hib* (haemophilus influenzae type b), *rubella*, *mumps*, *hpv* (human papilloma virus), *pneumococcal* (*disease*/*infection*), *meningitis*, *flu*, *influenza*, *polio*, *virus*, *pertussis*, *tetanus*, *diphtheria*, *cervical (cancer)*, *viruses*, *whooping* (*cough*), *haemophilus (influenza(e) type b*), (*whooping*) *cough*, *influenzae*, *bacteria*, *cancer*, *pneumonia* |
| | diseases and infections: non-evaluative pre-modifiers | *infectious, congenital, communicable, chronic, bacterial, viral, seasonal, acute, long-term* |
| | diseases and infections: verbs | *infected, infect, catching* |
| | bacteria and virus types | *type, types, strains, b, c, strain* |
| | immune system and nervous system | *immune, immunity, antibodies, nervous (system), (immune/nervous) system* |
| | immune system processes | *response, respond* |
| | medical conditions: general terms | *disorders, syndrome, condition, conditions, disorder* |
| | specific conditions | *autism, autistic(spectrum disorders)* |
| | non-pathological conditions | *pregnant, pregnancy* |
| | parts of the body | *bowel, brain, glands, blood, skin, throat, body, chest, heart* |
| | symptoms and signs | *fever, symptoms, rash, redness, headache, swelling, signs, temperature, diarrhoea* |
| | symptoms and signs: pre-modifiers | *sore, swollen* |

**Table 5.3** The *NHSvax* keywords in the 'health, disease and medical conditions' category

The first sub-category in this category contains words which refer to disease and health in general, such as *disease (s), health, infection(s), illness(es),* while the second sub-category consists of the ajdectives *unwell, ill* and *healthy*. Since one of the objectives of this thesis is to examine the ways in which beliefs about health are expressed in the *JABS* and *NHSvax*

corpora, it might seem useful to examine this group of words in detail. However, in the vast majority of its 2,827 occurrences, the word *health* is used in compound nouns referring to institutions, professionals or fields of activity, for example, *department of health* <600>,[56] *health professionals* <390>, *health protection agency* <314 >, *public health* <214>, *health visitors* <173>. As for *infection(s)* and *illness(es),*examination of concordances reveals that use of these words is salient to the objectives of this thesis only in those instances in which they collocate with words which denote causation. They need not be examined separately. The adjectives *unwell, ill* and *healthy,* on the other hand, are more likely to yield interesting results. *Unwell* and *ill* are potentially interesting because some of their contexts of use relate to reactions to vaccines, while *healthy* is a significant keyword in the *JABS* corpus (as will be explained in Chapter Six) and examination of the use of this word in the *NHSvax* corpus is likely to provide a useful comparison.

The third sub-category is the largest in this category and relates to terms denoting diseases. As we see in Table 3, these terms refer to diseases for which a vaccine is available from the NHS. These terms include *measles, rubella, mumps, meningitis, flu, polio,* and so on. While these are relevant to the topic of this thesis, relevant aspects of their use are likely to be picked up by examining concordances of words related to causation and so these words are not included as candidate terms. The terms in the 'diseases and infections: non-evaluative pre-modifiers' and 'diseases and infections: verbs', and 'bacteria and virus types' sub-categories are also rejected. The words in the subsequent four sub-categories are potentially more useful for fulfilling the research objectives. The terms *immune* and *immunity* are of obvious relevance. *Antibodies* and *system* are keyword in the corpus because they a very strong collocational

[56] Throughout the rest of this thesis, figures given in angle-brackets refer to the raw frequencies of items.

bond with *immune* (and, in the case of *antibodies*, also *immunity*). It was therefore considered unnecessary to select these as candidate terms. Similarly, *respond* and *response*, although representing potentially fruitful avenues of enquiry in their own right, also have a very strong collocational bond with *immune*. From the remaining words in the 'health, disease and medical conditions' category, only the words *bowel* and *brain* were selected, since both occur in the context of references to the MMR-autism hypothesis. The words selected from the 'health, disease and medical conditions' category are therefore, *healthy, unwell, ill, immune, immunity, bowel* and *brain*.

The words belonging to the 'people, places and institutions' category are shown in Table 5.4.

| | | |
|---|---|---|
| **people, places and institutions** | nhs users/vaccination target groups | *children, child, babies, parents, girls, infants, baby, adults, (age/risk/patient) groups, individuals, (general/target) population, young, group, patients, patient, women, people, person* |
| | pre-modifiers referring to groups of people | *older, young, human, target (groups/population), vulnerable (patients)* |
| | healthcare practitioners and medical experts | *gp, professionals, gps, nurses, nurse, doctor, (health) visitor, experts, expert, (general) practitioners, doctors, colleague, (healthcare) workers, colleagues* |
| | named individuals | *(dr andrew) wakefield,(dr david) salisbury* |
| | job titles and honorifics | *cmo (chief medical officer), professor, dr, chief (medical officer), director* |
| | public institutions | *pcts, primary (car/nhse trusts), (primary care/nhs) trusts, nhs, hpa, agency, department, committee, organization, joint (committee), csm (committee on the safety of medicines), (primary care/meningitis/nhs) trust, centre* |
| | specific institutions | *(royal) free (hospital)* |
| | industries | *manufacturers* |
| | private companies | *(aventis/sanofi) pasteur* |
| | places | *laboratory, clinic, surgeries, surgery, hospital, school* |
| | geographical references | *uk, countries, england, london, wales, national, locally, worldwide* |

**Table 5.4** The *NHSvax* keywords in the 'people, places and institutions' category

As is evident, the largest sub-category here consists of words which refer to the groups of people with whom the healthcare practitioners and vaccination programme administrators are concerned. Thus, we find words such as *children*, *child*, *babies*, *parents*, *girls*, *infants*, *baby*, and so on. What is interesting about these terms is that most of them are plural forms, thus indicating the preoccupation of the NHS with groups of people en masse, rather than with individuals. More interesting, though, are the terms in this group which are even less specific than those above in who or what they refer to. These words are *group(s)* and *population*, and the technical term *target*. Concordance evidence, however, shows that these words have a strong collocational relationship with *risk* and so need not be examined separately. From the sub-category relating to healthcare practitioners and medical experts, the terms expert and experts were considered potentially interesting in light Taylor's (2010) observations of their use in news articles, as discussed in Chapter Two. The terms *wakefield* and *salisbury* were also considered potentially interesting. These terms refer, respectively, to Dr Andrew Wakefield and Dr Andrew Salisbury, the then Director of Immunisation. Although it might be interesting to see how Andrew Wakefield is represented in this corpus, given the centrality of his role in the MMR controversy. However, it was decided that it would be more useful to concentrate on the ways in which key concepts are represented. The terms *expert(s)*, *wakefield* and *salisbury* were therefore not selected. The remaining terms in this category were rejected on the grounds that most of them were insufficiently specific in what they refer to. The two exceptions are *free*, which occurs in references to the Royal Free Hospital, where Andrew Wakefield's research group was based, and *pasteur*, a reference to the pharmaceutical company known alternatively as Aventis Pasteur and Sanofi Pasteur. It was felt that neither of these terms was key to fulfilling the aims of the thesis. No terms were selected from this category.

The words belonging to the ‘science, medicine and healthcare’ category are shown in Table 5.5.

| | | |
|---|---|---|
| **science, medicine and healthcare** | general references to science, medicine and healthcare | *medical, healthcare, clinical, (general) practice, medicine, scientific, nursing, diagnosed, practices, treatment, treat* |
| | terms related to administration of vaccination programme | *programme, routine, update, ordering, campaign, supply, introduced, introduction (of), changes, schedule, delivery, funding, updates, orders, supplies, updated, supplied, public (health), order, programmes, resources, services, appointment, feedback, issued, routinely, catch(-up programme), ordered, responsible, stock, access, reference, (yellow) card, yellow (card scheme), measure* |
| | administration of vaccines | *dose, booster, doses, injection, jab, injections, (vaccination) site* |
| | references to age and time periods | *childhood, age, pre-school, months, aged, years, september, year, born, february, weeks, august, march, monthly, annual, month, october, date, april, november, january, birth, december, winter, monday, july, june, friday* |
| | references to epidemiological research and monitoring | *deaths, uptake, surveillance, coverage, outbreaks, epidemic, levels, monitoring, survey, spread, cases, incidence, rate, proportion* |
| | quantities and quantifiers | *number, numbers, amounts* |
| | reference to scientific research | *results, data, study, studies, findings, research, epidemiology, statistics* |
| | research processes | *conducted* |
| | nhs information and communication | *information, advice, details, support, questions, answers, guidance* |
| | scientific and medical procedures and checks | *confirmed, testing, tested, test, check* |

**Table 5.5** The *NHSvax* keywords in the ‘science, medicine and healthcare’ category

As Table 5.5 shows, many of the words in this category relate to the administration of the vaccination programme, the administration of vaccines, and references to the ages at which vaccines are administered to children or to dates and time periods. There are also words which relate to the gathering and dissemination of epidemiological data, to the communication of information to the public, and to scientific and medical procedures. None of these terms are

considered potentially useful for answering the research questions guiding this project. Some of the words in the first sub-category, though, are potentially more useful. These are the words *medical, clinical, medicine* and *scientific*. Concordance evidence reveals, however, that *medical* and *medicine* predominantly occur in the corpus as parts of longer expressions. *Medical* typically occurs in terms such as *chief medical officer, british medical journal,* and so on. These terms were rejected as candidate terms. *Clinical* tends to occur in expressions such as *clinical risk groups* and *clinical evidence*. Clinical was rejected, therefore, since evidence of its use comes to light when terms such as *risk* and *evidence* are examined. *Scientific*, on the other hand, is used in a range of contexts. Given Taylor's (2010) findings about the evaluative use of the word *science* in popular publications, it seems plausible that the adjective *scientific* might be used in a similar fashion in the *NHSvax* corpus. The sub-category of words which refer to scientific research are potentially more useful in that they can shed light on some of the ways in which references to scientific sources are used in the NHS texts. However, the same information can be gleaned by examining the use of relevant verbs, for example *SHOW* or *REPORT*, forms of which are keywords in this corpus. The word *scientific* is the only keyword from this category to be selected. No keywords were chosen from the category of 'sources' (see Appendix 3), which mainly comprises terms referring to documents, most of which are produced by the NHS, for example, *copies, document, factsheet, leaflet, leaflets, letter*, and so on.

The five categories which follow contain a greater number of keywords which merit closer examination. The first three of these categories are shown in Table 5.6 below. They are the closely related categories of 'likelihood, causation and temporal relations', 'expressions with modal meaning' and 'conditionality'. Since notions of risk, uncertainty and causation are

central to the MMR controversy, all the terms in these categories were selected, with the exception of *contact* and *discuss*, which occur in such contexts as 'Contact your doctor' or 'Discuss this with your GP' and merit no further examination.

| | | |
|---|---|---|
| **likelihood, causation, and temporal relations** | possibility | *risk, risks* |
| | cause and effect | *cause, link, caused , causes, associated, causing, due* (*to*), *impact, linked, association, associated , affects, affected, affecting, lead* (*to*), *affect, links, related* |
| | pre-modifiers | *causal* |
| | reason | *reasons* |
| | ergative verbs indicating change | *increase, develop, increased, developing, reduce, increasing, rise* |
| | temporal relations | *since, following, follow, followed, after, prior, before, previous, previously* |
| | danger | *complications, damage, catch, catching* |
| **expressions with modal meaning** | epistemic modality | *may, can, will, likely, possible, cannot* |
| | necessity/desirability | *recommended, should, recommend, advised, recommendations, need, advise, recommends, required, needed* |
| | directives | *contact, discuss* |
| **conditionality** | | *if* |

**Table 5.6** The *NHSvax* keywords in the 'likelihood, causation and temporal relations', 'expressions with modal meaning', and 'conditionality' categories

The next two categories are the 'expressions of evaluation' category and the category which consists of verbs, nouns and phrases which refer to actions, events and states of affairs. These are shown in Table 5.7 below. As we can see, there are numerous terms which occur in expressions used in contexts in which positive evaluation of vaccines is expressed, as well as terms used to refer to potential harm or sources of harm. There are a number of explicitly evaluative terms in this category, as well as the 'status' nouns *evidence, facts, issues, problems* and *concerns*. All terms in this category were selected as candidate terms.

| | | |
|---|---|---|
| **expressions of evaluation** | terms used to express positive evaluation of vaccines | *protection, protect, safety, protects, protected, effective, safest, efficacy, ensure, protecting, safe, prevent, prevention, preventing, benefits* |
| | potential ill-effects of vaccines | *adverse (events/effects/reactions), reactions, anaphylactic (shock/reaction), reaction, effects, side (effects), (adverse) events* |
| | sources of harm | *exposure (to mercury/thiomersal),exposed (to mercury)* |
| | diseases and infections: evaluative terms | *serious, common, mild, rare, severe, minor* |
| | diseases and infections: potential effects | *breathing (problems), painful (disease/swollen glands)* |
| | evaluation of information/research | *latest* |
| | other evaluative terms | *high, higher, important, key, low, properly* |
| | hedging and boosting | *usually, rarely, highly, frequently, extremely* |
| | status nouns: the 'sign' group | *evidence* |
| | status nouns: factivity | *facts, issues, problems* |
| | status nouns: affect | *concerns* |
| | comparison/contrast | *or, compared, other* |
| **other nouns, verbs, and phrases referring to events, actions, and states of affairs** | verbal process/ attribution | *reported, concluded, reporting, statement* |
| | showing | *shows, shown* |
| | discovering | *finding, found, identified, identify* |
| | cognition | *awareness, (BE) aware, recognised, estimated, considered* |
| | occurring | *occur, occurred, occurs* |
| | durative | *continues* |
| | giving and receiving | *given, offered, receive, received, provides, receiving, provided, giving, provide, providing* |
| | starting and stopping | *avoid, starts* |
| | material processes: miscellaneous | *use, used, visit, using* |
| | relational processes | *contains, contain, containing, contains, include, includes* |
| | expression with meta-discoursal function | *listed* |

**Table 5.7** The *NHSvax* keywords in the 'expressions of evaluation' category and the category of 'other nouns, verbs and phrases referring to events, actions, and states of affairs'.

From the category labelled 'other nouns, verbs and phrases referring to events, actions, and states of affairs', the terms denoting verbal processes were selected, as were the terms in the 'showing', 'discovering' and 'cognition' sub-categories. The rationale for selecting these is that they can be used in attribution. The forms of the verb *OCCUR*, listed in the 'occurring'

sub-category, were also chosen on account of Partington's (2004b) observation that the prosody of *OCCUR*, when it is used in academic texts, encodes uncertainty.

The two remaining categories listed in Appendix 3 are the set of those grammatical keywords which have not already been included in a semantic category and a small set of miscellaneous items. The miscellaneous items comprise words which were difficult to categorise in terms of their relevance to any of the categories already identified, a set of items related to electronic communication (these probably showed up as keywords owing to their absence in the *BNC*), and a small set of abbreviations. Words in this category are not candidates for further analysis. There are relatively few grammatical keywords listed in the grammatical items category. Only two are deemed relevant since they have a strong collocational relationship with *mmr, measles, mumps, rubella* and, in the case of one of them, *autism*. These words are *against* and *between*.

The keywords which have been selected for closer examination are shown in Table 5.8 below, organized according to their semantic and functional categories. As explained in the previous chapter, the keywords identified for closer inspection serve as a route to uncovering frequent recurrent patterns which are used to express particular notions and concepts. The discussion that follows does not describe the lexico-grammatical environment of each keyword separately. Instead it focuses on key themes which emerge from the analysis of the keywords and presents a discussion of the recurrent lexico-grammatical patterns which are used to realize particular propositions. These themes concern the ways in which propositions related to the causal hypothesis are expressed, focussing in particular on how notions of causation

and risk are expressed, the ways in which beliefs about health and the immune system are expressed, and the ways in which appeals to authority are made.

| **Semantic/functional category** | **Keywords** |
|---|---|
| vaccines | *vaccine, vaccines, mmr* |
| medical conditions | *autism, autistic, disorders, syndrome, condition, conditions, disorder* |
| parts of the body | *bowel, brain* |
| health and disease | *unwell, ill, healthy* |
| immune system | *immune, immunity* |
| science and medicine | *scientific* |
| risk | *risk, risks* |
| cause and effect | *cause, link, caused, causes, associated, causing, due (to), impact, linked, association, associated, affects, affected, affecting, lead (to), links, related* |
| items signalling temporal meaning | *since, following, follow, followed, after, prior, before, previous, previously* |
| cause and effect pre-modifier | *causal* |
| reason | *reasons* |
| dangers of diseases and potential harm | *complications, damage, catch, catching, damage* |
| epistemic modality | *can, may, will, cannot, likely, possible* |
| deontic modality | *recommended, should, recommend, advised, recommendations, need, advise, recommends, required, needed* |
| conditionality | *if* |
| positive evaluations of vaccination | *protection, protect, safety, protects, protected, effective, safest, efficacy, ensure, protecting, safe, prevent, prevention, preventing, benefits* |
| reactions to vaccines | *adverse, reactions, effects, reaction, events* |
| evaluation of research | *latest* |
| explicitly evaluative items (diseases) | *serious, common, mild, rare, severe, minor* |
| other evaluative terms | *high, higher, important, key, frequently, extremely* |
| hedging and boosting | *usually, rarely, highly, frequently, extremely* |
| status nouns: signs | *evidence* |
| status nouns: factivity | *facts, issues, problems* |
| status nouns: affect | *concerns* |
| comparison/contrast | *or, compared, other* |
| verbal processes | *reported, concluded, statement* |
| showing | *shows, shown* |
| discovering | *finding, found, identified, identify* |
| cognition | *awareness, aware, recognised, estimated, considered* |
| occurring | *occur, occurred, occurs* |
| other prepositions | *against, between* |

**Table 5.8** Keywords in the *NHSvax* corpus with at least 10% distribution across corpus texts, identified for closer investigation

## 5.3 Expressions related to the MMR-autism causal hypothesis in the *NHSvax* corpus

### 5.3.1 Clusters associated with *mmr* and *autism*

In order to discover how propositions relating to the causal hypothesis are expressed, it is useful to start by examining concordance data for the keywords *mmr* and *autism*, before going on to consider the keywords which express causation. Appendices 4 and 5 show, respectively, the one hundred most frequent clusters from the concordances for *mmr* and *autism*. Since *WordSmith* seeks clusters from within a window of five words either side of the node, not all the clusters contain the node word. Table 5.9 below shows the twenty most frequent meaningful clusters (with frequencies) which contain the nodes *mmr* or *autism* (as appropriate).[57] By 'meaningful' I am referring to what Sinclair (1991: 110) terms 'semi-pre-constructed phrases'. Some of these correspond to complete grammatical units, such as, for example, noun phrases or prepositional phrases. Others correspond to the core parts of recurrent expressions in which there are fixed elements plus one or more variable elements (we might term these semantic sequences). For example, the *NHSvax* corpus has 28 citations of *risk of catching*. In 22 instances, the string *risk of catching* is part of the longer expression *at risk of catching*. The string *(at) risk of catching* is invariably followed by a term which denotes a disease. The expressions which follow *(at) risk of catching* in the *NHSvax* corpus are *measles* <7>, *measles, mumps or rubella* <5>, *the diseases* <5>, *either measles or mumps* <4>, *rubella* <2>, and *them* <2>. The string *risk of catching* can therefore be considered a semi-pre-constructed unit, which, although it does not correspond to a grammatically complete unit, has meaning potential.

[57] *WordSmith Tools*' calculations of frequency are not 100% accurate, as can be seen in the discrepancy between the recorded frequencies of the string *mmr and autism* when calculated from the two concordances. However, where word counts vary, the difference in scores is usually only one, or, at the most, two.

| | Clusters with *mmr* as node | freq. | Clusters with *autism* as node | freq. |
|---|---|---|---|---|
| 1 | *the mmr vaccine* | 505 | *mmr and autism* | 161 |
| 2 | *mmr and autism* | 162 | *link between mmr and autism* | 90 |
| 3 | *link between mmr and autism* | 91 | *vaccine and autism* | 88 |
| 4 | *dose of mmr* | 85 | *mmr vaccine and autism* | 76 |
| 5 | *mmr vaccine and autism* | 77 | *children with autism* | 67 |
| 6 | *doses of mmr* | 68 | *prevalence of autism* | 41 |
| 7 | *two doses of mmr* | 52 | *incidence of autism* | 41 |
| 8 | *mmr vaccine was introduced* | 50 | *the incidence of autism* | 40 |
| 9 | *safety of mmr* | 47 | *increase in autism* | 38 |
| 10 | *mmr the facts* | 46 | *a link between mmr and autism* | 38 |
| 11 | *the mmr vaccination* | 39 | *risk of autism* | 36 |
| 12 | *the mmr vaccine was introduced* | 38 | *causes of autism* | 34 |
| 13 | *a link between mmr and autism* | 38 | *cases of autism* | 32 |
| 14 | *second dose of mmr* | 37 | *the mmr vaccine and autism* | 30 |
| 15 | *between mmr vaccine and autism* | 34 | *vaccination and autism* | 30 |
| 16 | *the mmr vaccine and autism* | 30 | *autism spectrum disorders* | 29 |
| 17 | *the safety of mmr* | 29 | *the prevalence of autism* | 28 |
| 18 | *immunised with mmr* | 28 | *diagnosis of autism* | 26 |
| 19 | *a second dose of mmr* | 28 | *between the mmr vaccine and autism* | 26 |
| 20 | *introduction of mmr* | 28 | *vaccines and autism* | 24 |

**Table 5.9** Top twenty meaningful three- to six-word clusters in *NHSvax* corpus with *mmr* and *autism* as node words

As is apparent from Table 5.9, the most frequent meaningful clusters with *mmr* and *autism* point to the high frequency of the string *link between mmr and autism*. The use of nominalization, typical of scientific literature, has carried over into this popularization genre. The connection between autism and inflammatory bowel disease is evident from the clusters, as is a preoccupation with rates of incidence of autism. The nominalized form *link* is evidently preferred over more explicit signals of causal relations, such as *CAUSE*. As mentioned in Chapter Two, there is a marked tendency in scientific writing to condense clausal meaning into nominalized forms (Halliday, 1998). This creates textual cohesion and drives the argument forwards. It is suggested that an effect of nominalization is to obscure agency (Veel,

1998). In the case of the MMR causal hypothesis, however, the use of nominalization serves not to obscure causal relations, instead, its use reflects the uncertainty of the science. We also see in Table 5.9 that *CAUSE* only appears in the most frequent clusters in the string *causes of autism*. However, it is worth examining concordance data, not just for *causes of autism*, but for *cause autism* and *causes autism* in order to gain a fuller picture of the ways in which the causal hypothesis is expressed. The string *risk of autism* also merits closer examination, in order to discover the extent to which, and how, the possibility of a cause-effect relation between vaccination and autism is expressed in terms of risk. These are discussed in the sections which follow.

### 5.3.2 Expressing the causal hypothesis using explicit markers of causation

There are 33 instances of *causes of autism*. These are shown in Concordance 5.1 below. As we can see, the expression *causes of autism* tends to occur in references to scientific research. There are, for example, four instances of *research into causes of autism*, four citations of *epidemiology and possible causes of autism*, two of *epidemiology and causes of autism* and two of *what scientific research has revealed about the occurrence and causes of autism spectrum disorders*. The environment in which *causes of autism* occurs also indicates a degree of uncertainty. For example, there are eight instances in which the expression *causes of autism* is pre-modified with *possible*, as Example 5.7 illustrates.

> [5.7] **The possible causes of autism spectrum disorders** are reviewed in a recent paper by Prof P. Szatmari.

The idea that the way in which parents view the onset of autism may not be fixed is expressed in the five repeated instances of the following example:

[5.8] The aim of this study was to estimate the prevalence of autism and to look at whether **parental perception regarding the onset or causes of autism** had changed over that time.

." Q7 If MMR does not cause autism, what does? A7 The MRC's report on the ?Epidemiology and Causes of Autism?, published in December last year, provides a clear picture of what scientific research
is to help them to take forward recommendations on research in their report on the ?Epidemiology and Causes of Autism?." Q7 If MMR does not cause autism, what does? A7 The MRC's report on the
last year, provides a clear picture of what scientific research has revealed about the occurrence and causes of autism spectrum disorders (ASDs). It concludes that the strongest evidence to date is for a
last year, provides a clear picture of what scientific research has revealed about the occurrence and causes of autism spectrum disorders (ASDs). It concludes that the strongest evidence to date is for a
of Autism-." Q7 If MMR does not cause autism, what does? A7 The MRC's report on the - and Causes of Autism-, published in December last year, provides a clear picture of what scientific
. This is to help them to take forward recommendations on research in their report on the - and Causes of Autism-." Q7 If MMR does not cause autism, what does? A7 The MRC's report on the -
rejection of a causal relationship between thiomersal-containing vaccines and autism". On research into causes of autism, the Committee said: "? more research must be conducted on autism. However, the
rejection of a causal relationship between thiomersal-containing vaccines and autism". On research into causes of autism, the Committee said: "& more research must be conducted on autism. However, the
rejection of a causal relationship between thiomersal-containing vaccines and autism". On research into causes of autism, the Committee said: "-- more research must be conducted on autism. However, the
rejection of a causal relationship between thiomersal-containing vaccines and autism". On research into causes of autism, the Committee said: "-- more research must be conducted on autism. However, the
based on observations of rats, this article began: "We have approached the possible neurochemical causes of autism by assuming that the fundamental problem of the autistic child is emotional. Some of
estimate the prevalence of autism and to look at whether parental perception regarding the onset or causes of autism had changed over that time. The results of this study showed a levelling in the
estimate the prevalence of autism and to look at whether parental perception regarding the onset or causes of autism had changed over that time. The results of this study showed a levelling in the
estimate the prevalence of autism and to look at whether parental perception regarding the onset or causes of autism had changed over that time. The results of this study showed a levelling in the
estimate the prevalence of autism and to look at whether parental perception regarding the onset or causes of autism had changed over that time. The results of this study showed a levelling in the
estimate the prevalence of autism and to look at whether parental perception regarding the onset or causes of autism had changed over that time. The results of this study showed a levelling in the
factors." Hershel Jick M.D. and James A. Kaye, M.D. D.P.H.(Dec 2003) Epidemiology and Possible Causes of Autism. Pharmacotherapy 2003;23(12):1524-1530 30
: Increase In Autism Most Likely Due To Changing Diagnostic Practices Epidemiology and Possible Causes of Autism: Hershel Jick, M.D. Pharmacotherapy Journal December 2003 Changes in diagnostics
factors." Hershel Jick M.D. and James A. Kaye, M.D. D.P.H.(Dec 2003) Epidemiology and Possible Causes of Autism. Pharmacotherapy 2003; 23 (12):1524-1530 No evidence for links between autism,
.htm Increase In Autism Most Likely Due To Changing Diagnostic Practices Epidemiology and Possible Causes of Autism: Hershel Jick, M.D. Pharmacotherapy Journal December 2003 Changes in diagnostics
autism. A study funded by The Medical Research Council, is currently under way investigating possible causes of autism. This study will examine whether autistic children have a history of other conditions or
al. ( 2003 ). The cause of autism spectrum disorders. British Medical Journal 326: 173-4. The possible causes of autism spectrum disorders are reviewed in a recent paper by Prof P. Szatmari. This review
. Recall bias, MMR and autism. Archives of Diseases in Children 87:493-4. Jan, 2003 The possible causes of autism spectrum disorders are reviewed in a recent paper by Prof P. Szatmari. This review
. Recall bias, MMR and autism. Archives of Diseases in Children 87:493-4. Jan, 2003 The possible causes of autism spectrum disorders are reviewed in a recent paper by Prof P. Szatmari. This review
any, from the thiomersal present in the vaccine. MMR not an environmental risk factor for autism The causes of autism spectrum disorders "In spite of recent publicity, there is good epidemiological evidence
syndrome cases. If MMR is not causing autism, what is? What studies are being done looking into the causes of autism The cause(s) of autism is not yet known and many factors are thought to play a role
in the Department of Clinical Medicine at the University of Bristol, is carrying out research into the causes of autism. She said: "These findings are in line with all of the other research that's been done.
research paid for by the LSC will be sent to the Medical Research Council who are investigating the causes of autism. ENDS Media information: Katherine Hartup Tel: 020 7759 0493 Richard Shand Tel:
research paid for by the LSC will be sent to the Medical Research Council who are investigating the causes of autism. ENDS Media information: Katherine Hartup Tel: 020 7759 0493 Richard Shand Tel:
has been raised, studies have shown those fears to be groundless. In its recent review of the causes of autism, the American Academy of Pediatrics considered around 1000 scientific papers, and
has been raised, studies have shown those fears to be groundless. In its recent review of the causes of autism, the American Academy of Pediatrics considered around 1000 scientific papers, and
disorders. The Committee believes the matter is likely to be clarified by a better understanding of the causes of autism. GACVS also concluded that there is no evidence to support the routine use of
disorders. The Committee believes the matter is likely to be clarified by a better understanding of the causes of autism. GACVS also concluded that there is no evidence to support the routine use of

**Concordance 5.1** *Causes of autism*

A search for *cause* and *causes* + *autism* revealed 38 instances of *cause autism* and 19 of *causes autism*. The contexts of use indicate that the proposition that MMR causes autism is typically negated, or it is expressed in the form of a question. Alternatively, it is attributed to a speaker or writer. There are 16 instances of *mmr/it does not cause autism (or inflammatory bowel disease).* However, these citations usually occur in longer strings which include the word *evidence*. Examples are:

[5.9] **The evidence** is that MMR vaccine does not cause autism or inflammatory bowel disease (IBD).

[5.10] There is **convincing evidence** that MMR does not cause autism or any particular subtypes of autistic spectrum disorder.

[5.11] The **weight of scientific evidence** strongly indicates that MMR does not cause autism.

As is evident from the examples above, the force of the proposition is often intensified. Evidence is sometimes pre-modified by terms such as *convincing* or the term *scientific* is used to lend authority to the assertion. In the example above, the expression *the weight of scientific evidence* lends extra emphasis. Elsewhere, *cause autism* occurs as part of a question: 'Does MMR cause autism?' or 'Doesn't MMR cause autism?' Sometimes, the proposition is attributed, and, in such instances, a low degree of commitment to the truth of the proposition is expressed. The following example is typical:

[5.12] … only a very small proportion of health professionals **thought that** MMR can cause autism or bowel disease (around 1% or 2%).

Appeals to scientific evidence are not uncommon, as in the following example, in which the status of the proposition is expressed as a hypothesis:

[5.13] **This study provides strong evidence against the hypothesis that** MMR vaccination causes autism.

### 5.3.3 Expressing the causal hypothesis using nominalization

As already noted, the most common way in which the causal hypothesis is referred to is through use of the expression *link between mmr and autism*. A concordance of *link between mmr and autism* returns 68 occurrences. The top twenty collocates of the string *link between*

*mmr and autism*, according to MI3 are shown in Table 5.10. MI3 values are shown in brackets.

| Rank | Collocate | MI3 value | Total no. of occurrences | Total to left of node | Total to right of node |
|---|---|---|---|---|---|
| 1 | *ever-increasing* | 19.84 | 4 | 0 | 4 |
| 2 | *finds* | 19.36 | 7 | 7 | 0 |
| 3 | *causal* | 19.14 | 11 | 11 | 0 |
| 4 | *credible* | 19.00 | 5 | 5 | 0 |
| 5 | *evidence* | 18.99 | 28 | 27 | 1 |
| 6 | *no* | 18.64 | 31 | 30 | 1 |
| 7 | *judgement* | 18.04 | 4 | 2 | 2 |
| 8 | *there* | 17.85 | 27 | 16 | 11 |
| 9 | *upholds* | 17.84 | 2 | 0 | 2 |
| 10 | *alleging* | 17.79 | 3 | 3 | 0 |
| 11 | *a* | 17.62 | 58 | 51 | 7 |
| 12 | *perception* | 17.01 | 3 | 0 | 3 |
| 13 | *mcgill* | 16.84 | 2 | 0 | 2 |
| 14 | *unequivocal* | 16.52 | 2 | 0 | 2 |
| 15 | *suggesting* | 16.45 | 4 | 4 | 0 |
| 16 | *separating* | 16.26 | 2 | 0 | 2 |
| 17 | *speculation* | 16.21 | 3 | 1 | 2 |
| 18 | *hypothesis* | 16.11 | 4 | 4 | 0 |
| 19 | *create* | 16.04 | 2 | 0 | 2 |
| 20 | *support* | 15.80 | 7 | 7 | 0 |

**Table 5.10** The 20 most significant collocates, according to MI3, of the string *link between mmr and autism*

The relatively high statistical significance of the left-hand collocates *causal, credible, evidence, no, there* and *a* results from the fact that *link between mmr and autism* frequently occurs in the longer string *there is no (credible) evidence of a (causal) link between mmr and autism*. This is the most frequent expression used to rebut the causal hypothesis. The left-hand collocate *finds* is relatively highly significant owing to its use in the expression *finds no causal link. Support* is statistically significant because the expression *the evidence did not support a link* also recurs. It is interesting to note the relatively high significance of the terms *alleging* and *suggesting* , which can be used to signal authorial distance, and the 'status'

nouns *perception*, *speculation* and *hypothesis*, which function as indicators that the proposition to which they refer is not 'directly aligned with the world' (Hunston, 2011: 28). Not only is the expression *link between mmr and autism* highly frequent in the *NHSvax* corpus, but concordance evidence suggests that the causal hypothesis is also sometimes expressed using the expression *association between mmr (vaccine/vaccination) and autism*. There are 48 occurrences of *association between mmr and x.* Typical examples of the ways in which the causal hypothesis is rebutted in the data are given below.

> [5.14] New study **finds no link between MMR vaccine and autism**.
>
> [5.15] The most comprehensive, independent research analysis yet undertaken has **found no link between the MMR vaccine and autism or inflammatory bowel disease.**
>
> [5.16] The Department of Health has welcomed this latest evidence which **finds no causal link between MMR and autism**.
>
> [5.17] New research involving half a million children **finds no evidence of MMR link to autism**
>
> [5.18] The inquiry, which was published in March 1998, came to the conclusion that there is **no evidence to indicate any link between MMR vaccination and bowel disease or autism.**
>
> [5.19] … the American Academy of Pediatrics considered around 1000 scientific papers, and concluded that **the evidence did not support a link between MMR and autism.**
>
> [5.20] 'I am certainly **not aware of any convincing evidence for the hypothesis of a link between MMR and autism**...' (Berelowitz, 2001).
>
> [5.21] MMR vaccination cause autism? There is **no credible evidence of a link between MMR and autism.**
>
> [5.22] The main objective of this study was to evaluate whether there was **an association between MMR vaccine and autism.**
>
> [5.23] **No credible scientific evidence** shows **an association between MMR and autism.**

As some of the examples above illustrate, the term *evidence* is often pre-modified and the effect of the pre-modifiers is to intensify the force of the proposition. The most common

strings are *no scientific evidence* <17>, *no credible scientific evidence* <7>, *no credible evidence* <6>, *no new scientific evidence* <4>, *no convincing evidence* <4>. The use of expressions such as *FIND no link between…* and *no (credible) evidence of a link/association between…* is typical of the ways in which scientists express claims. Although one can argue that *FIND no link between* expresses a strong degree of commitment to the truth of the proposition, it can be interpreted as a way of hedging the claim. The claim that a link was not found might be taken as implying that a link might exist and is yet to be uncovered. The expression *FIND no evidence of* can be interpreted in a similar way, whilst *FIND no credible evidence* implies that some evidence was found but was rejected by the scientist. It is easy to see how the meanings implied in expressions such as these might be exploited in vaccine-critical discourse.

### 5.3.4 Expressing the causal hypothesis in terms of risk

Concordance evidence shows that the 36 instances of *risk of autism* in the *NHSvax* corpus tend to occur in references to scientific papers. Typical examples are shown below.

> [5.24] **Paper examines whether, in the UK, there is an increased risk of autism** (AD) following exposures, in early life, to wild measles, live attenuated measles, alone or in combination as MMR, and the alteration of the mumps strain within MMR.
>
> [5.25] **The authors investigated whether MMR vaccination is associated with an increased risk of autism** or other pervasive developmental disorders.

Once again, the tendency is to rebut the causal hypothesis, often with reference to *evidence*, or *scientific evidence*:

> [5.26] Update of **scientific evidence in published studies have continued not to find an increased risk of autistic spectrum disorder** associated with MMR.

### 5.3.5 Summary of Section 5.3

Corpus evidence shows that the MMR-autism hypothesis is typically expressed in (popularized) medical-scientific discourse using the term *link between MMR and autism*. This reflects the preference in scientific writing for nominalization, but also reflects the uncertainty surrounding the MMR-autism causal connection. In the *NHSvax* corpus, unsurprisingly, the causal proposition is typically rebutted. Expressions which encode a degree of uncertainty, such as *(FIND) no (scientific/credible/convincing) evidence of a link between MMR and autism*, are most common.

## 5.4 Expressions related to health and the immune system in the *NHSvax* corpus

### 5.4.1 Disease and causation

The NHS websites from which the *NHSvax* data was gathered have a very narrow focus on promoting vaccination uptake. As a result of this, the data contains descriptions of the diseases, what causes them, and what effects the diseases can cause. As one would expect, the pathogenic theory, or germ theory, of disease (the idea that micro-organisms rather than genetics or lifestyle choices are the main factors causing disease) underpins the description. Concordance evidence shows numerous instances of the semantic sequence A BACTERIUM/VIRUS CAUSES A DISEASE, for example:

> [5.27] Rubella is a disease **caused by a virus**.
>
> [5.28] … **the bacteria that cause the disease** are still present in soil.
>
> [5.29] **Meningococcal infection can cause meningitis and septicaemia** (blood poisoning).
>
> [5.30] This vaccine protects against **the two virus types that cause over 70% of cervical cancer.**

[5.31] Pneumococcal infection **can lead to a range of diseases** including pneumonia, septicaemia and meningitis.

Related semantic sequences are A DISEASE/VIRUS CAUSES A PHYSICAL REACTION, A DISEASE CAUSES DAMAGE and A DISEASE CAUSES DEATH:

[5.32] Tetanus is a painful disease that **affects the muscles and can cause breathing problems.**

[5.33] encephalitis: Inflammation of the tissues of the brain **which can cause lasting brain damage.**

[5.34] Pneumococcal infection can lead to **a range of diseases including pneumonia, septicaemia and meningitis which may cause death**.

Thus we see that the risks of diseases are emphasized in the *NHSvax* data.

### 5.4.2 Health and vaccination

Vaccination is presented as the way to avoid the dangers posed by bacteria and viruses. There are 401 citations of *protection* (although results are skewed by just over 100 citations of *health protection agency* and *health protection informatics*) 308 citations of *protect*, and 97 of protects:

[5.35] The new vaccine **provides the same level of protection** as the old one.

[5.36] We want to provide children with **the safest and most effective protection from disease.**

[5.37] Research from around the world shows that immunisation is **the safest way to protect your child's health.**

[5.38] MMR is the combined vaccine against measles, mumps and rubella. It is **the safest way to protect your children against these diseases.**

[5.39] MMR vaccine **protects against measles, mumps and rubella**, and is routinely given to children at 13 months and again between 3 and 5 years of age.

Interestingly, a number of the contexts in which *protection* is used encode modal meaning. In Examples 5.40, 5.41 and 5.42, expressions of desirability and necessity, such as *would be a good thing, is recommended,* and *are required* are used to urge the reader to get their child vaccinated. To reinforce the message, the effects of vaccination are evaluated positively: *the benefits of boosting her protection; extra protection; good protection.* The implication is that, by accepting the vaccines for their child, parents are doing their children good.

> [5.40] Even if your daughter had received her first MMR, and a single rubella, **having another MMR would be a good thing** as it has the benefits of boosting her protection against measles and mumps as well.
>
> [5.41] **The vaccination is recommended** because it gives children extra protection against a serious disease.
>
> [5.42] **Two doses of MMR vaccine are required** to give good protection against measles and mumps and can be given 3 months apart.

Concordance evidence shows that *should* is not typically used to encourage parents to vaccinate. Concordance lines with *should* tend to occur in documents aimed at health professionals and refer to when particular vaccines should be administered or how they should be stored.

Since health advice on the NHS websites is aimed at encouraging people to accept vaccinations for their children or for themselves in order to avoid contracting diseases, there is little evidence in the corpus data as to what the NHS believes constitutes good health, other than an implicit understanding that it entails not suffering from disease. This is seen in concordance evidence for *unwell* and *ill*:

> [5.43] Occasionally, children may be **unwell and irritable and develop a temperature, headache, sickness and swollen glands**.

[5.44] Measles is caused by a very infectious virus […] Children often have to spend about five days in bed and may be off school for ten days. **Adults are likely to be ill for longer**.

[5.45] Most children who get Hib infections **become very ill and need hospital care**.

As discussed in Section 5.2.2, the word *health* was rejected as a keyword which warranted close analysis since it typically occurs in the *NHSvax* corpus as part of compound nouns and binomials denoting the titles of organizations and job titles (for example, *department of health, health protection agency, health visitors, world health organization*, and so on), or to refer to public health issues in general, for example, *occupational health, health and social care*, *public health problem*, and so on. However, very occasionally, *health* occurs in the environment of *PROTECT*, for example,

[5.46] Research from around the world shows that immunisation is **the safest way to protect your child's health.**

This reinforces the idea that good health equates to avoiding disease. Concordance evidence for *healthy* is a little more interesting. There are 48 citations of *healthy* and the most frequent right-hand lexical collocates are *children* <9>, *people* <7>, *adult* <3> and *adults* <3>. In the majority of occurrences, *healthy* equates to an absence of disease. For example:

[5.47] … even in countries such as the UK, **if healthy children catch measles**, especially if they are in their teens, they can still die.

[5.48] This means that the bacteria can be spread among **vaccinated and healthy people** without causing any illness.

[5.49] These initial trials are done on a small number of **healthy adult volunteers** (about 10-12 people).

[5.50] Clinical trials examined **healthy adults** who had received a full primary course of vaccine plus a booster dose of vaccine containing diphtheria and tetanus more than ten years ago.

However, there are a small number of examples in which *healthy* has somewhat different associations. The first example, shown below, comes from the FAQ section of the NHS immunisation website.

> [5.51] Surely if children are **healthy and well-nourished** they won't die of these diseases?

Here, the term *healthy* is associated with *well-nourished*, the implication being that good health is associated with a good diet. The proposition that healthy and well-nourished children will not die from certain diseases is attributed to a member of the public. As one might expect, the answer that follows this question rebuts the argument:

> [5.52] Unfortunately, even with modern medical care, we cannot prevent or treat the real risk of serious complications. For example, from 1970 to 1983, **more than half the deaths from measles in the UK occurred in previously healthy children**.

Finally, there are three instances of *healthy* which occur in contexts where the MMR-autism hypothesis is referred to. Two of these occurrences are a repeated line which comes from a report on a court hearing:

> [5.53] The court ruled that **prior to the MMR vaccination the girl diagnosed with early infantile autism was healthy and had normal development**.

The third occurs in the context of a rebuttal of one aspect of the hypothesis:

> [5.54] **The notion that a previously healthy child was "normal" one day and showed clear signs of autism the next day** is at odds with the clinical course of pervasive developmental disorder.

The notion of health which is implied in the *NHSvax* data is that good health consists of an absence of disease and that the best way to avoid contracting communicable diseases is through vaccination. A belief that people who are 'healthy' and are well-nourished do not need vaccinations is attributed to members of the public and is rebutted.

### 5.4.3 The immune system

#### 5.4.3.1 Meanings associated with the word *immune*

The top thirty collocates of *immune* in the *NHSvax* corpus, according to MI3, are shown in Table 5.11 below. In order to gain an idea of the normal collocational environment of *immune* in general English, the top thirty collocates of *immune* in the *UkWaC* corpus are shown.[58] As is evident, the collocational environment of *immune* in the *NHSvax* corpus is similar to that found when the word is used in general English. The top collocates in both (if we ignore *immune* in the *NHSvax* list) are *system* and *response*. We also find *systems* in both lists, as well as other forms of the lexeme *RESPOND*, and words which refer to disease, namely and *disease* and *infections* (*UkWaC*) and *bacterial* (*NHSvax*). Interestingly, in both corpora, terms which refer to problems have a strong collocational relationship with *immune* but there is a slightly different variety of words in each. In *UkWaC*, we find *Deficiency, weakened, deficiency, suppress* and *suppression*, whereas in *NHSvax*, we find *overload, weaken, weakened, weakening, overloads* and *suppressed.* This suggests that it is not unusual, when using the word *immune*, to refer to weakened immune systems or to some form of suppression of the immune system. However, there is a particular concern in the *NHSvax* data with the idea of 'overload'. Finally, there are certain words which have a strong collocational

[58] The *UkWaC* collocate list was compiled using Sketch Engine, http://www.sketchengine.co.uk, accessed 02/03/2013.

relationship with *immune* in one of the corpora but not the other. In *ukWaC*, we find the nouns *cells*, *function* [59]and *antigens* and the adjectives *humoral*, *innate*, *Acquired* and *cell-mediated*, while in *NHSvax*, we see words such as body and infant (both keywords in the corpus) and the verbs *produce*, *priming* and *meeting.*[60]

| | UkWaC | | NHSvax | |
|---|---|---|---|---|
| | **collocate** | **MI3** | **collocate** | **MI3** |
| *1* | *system* | 36.261 | *immune* | 27.96 |
| *2* | *response* | 32.121 | *system* | 25.08 |
| *3* | *cells* | 30.714 | *response* | 21.29 |
| *4* | *the* | 30.309 | *s* | 19.56 |
| *5* | *responses* | 30.132 | *the* | 19.25 |
| *6* | *To* | 28.820 | *overload* | 19.01 |
| *7* | *and* | 28.557 | *weaken* | 18.57 |
| *8* | *Of* | 28.163 | *to* | 18.47 |
| *9* | *Deficiency* | 27.883 | *body* | 18.32 |
| *10* | *systems* | 27.877 | *a* | 17.66 |
| *11* | *body* | 27.434 | *infant* | 17.66 |
| *12* | *weakened* | 27.345 | *systems* | 17.34 |
| *13* | *humoral* | 27.285 | *responds* | 17.31 |
| *14* | *deficiency* | 27.160 | *produce* | 17.05 |
| *15* | *function* | 26.878 | *someone* | 16.98 |
| *16* | *innate* | 26.832 | *meeting* | 16.39 |
| *17* | *Is* | 26.618 | *of* | 16.23 |
| *18* | *suppress* | 26.393 | *weakened* | 16.14 |
| *19* | *In* | 26.279 | *responses* | 16.13 |
| *20* | *Acquired* | 26.277 | *weakening* | 16.07 |
| *21* | *boost* | 26.110 | *overloads* | 16.01 |
| *22* | *antibodies* | 26.045 | *priming* | 15.81 |
| *23* | *immune* | 26.030 | *bacterial* | 15.73 |
| *24* | *are* | 25.808 | *an* | 15.73 |
| *25* | *infection* | 25.781 | *and* | 15.69 |
| *26* | *antigens* | 25.740 | *not* | 15.66 |
| *27* | *cell-mediated* | 25.616 | *with* | 15.66 |
| *28* | *suppression* | 25.579 | *is* | 15.48 |
| *29* | *that* | 25.428 | *are* | 15.37 |
| *30* | *disease* | 25.296 | *suppressed* | 15.34 |

**Table 5.11** The thirty most significant collocates of *immune* in *ukWaC* and *NHSvax*, according to MI3[61]

[59] Evidence from Sketch Engine shows that, when it collocates with *immune*, *function* typically occurs in the expression *immune system function*.
[60] Concordance evidence shows that *meeting* is used as a noun here.
[61] The collocation programme in Sketch Engine is case-sensitive, hence the occurrence of *Deficiency* and *deficiency* in the *UkWaC* collocate list.

The collocational profile revealed in Table 5.11 above indicates that there is concern in the *NHSvax* texts with talking about processes. It also indicates that there are aspects of knowledge about the immune system which are mentioned relatively rarely in the *NHSvax* texts or are not mentioned at all, hence the absence among the top collocates of *immune* in *NHSvax* of *humoral, innate, Acquired* and *cell-mediated*.

If we look closely at the contexts in which *immune* is most commonly used in the *NHSvax* corpus, we find, firstly, a number of descriptions of the way in which the immune system works and how vaccines interact with it. For example:

> [5.55] As soon as a child is born it comes into contact with thousands of bacteria and viruses. **A baby's immune system responds** to all these challenges and prevents them from causing harm.
>
> [5.56] By providing protection against a number of bacterial and viral pathogens, **vaccines prevent the 'weakening' of the immune system** and consequent secondary bacterial infections occasionally caused by natural infection.'
>
> [5.57] The vaccine is tested for safety and also to see that it does **produce the immune response needed to prevent the disease**.
>
> [5.58] Even if children have responded to the vaccine the first time, there will be no problem from being exposed to the viruses again. **It's like any one of us who is immune meeting someone with the disease** - the infection can't get established.
>
> [5.59] Antigen - A substance which, under appropriate conditions, **triggers an immune response.**

There are also some references to the fact that vaccination might be potentially harmful for certain groups of people. For example:

[5.60] **If your child's immune system is "suppressed" (because they are having treatment for a serious condition such as a transplant or cancer**), the doctor or practice nurse should get advice from a child health specialist.

The ways in which *OVERLOAD* is used when forms of the lexeme collocate with *immune* is interesting. The idea that combined vaccines, such as the MMR vaccine, might overload the immune system is a subject of scientific enquiry, as is evidence by the fact that immune overload appears in the titles of some journal articles, as in the following examples.

[5.61] E Miller, N Andrews, P Waight, B Taylor (2003). Bacterial infections, **immune overload, and MMR vaccine**

[5.62] An article in Pediatrics **examined whether vaccines overwhelm or weaken the infant's immune system.**

An assumption that the members of the public believe that multiple vaccines overload the immune system is implicit in some of the questions which occur in the FAQ sections of the website, such as:

[5.63] What about giving three live vaccines at once - **it overloads the immune system**?

The proposition that multiple vaccines overload the immune system is rebutted, although it is most frequently rebutted using the expression *(finds/is) no evidence*. It is interesting to note the differences in the degrees of commitment to the truth of the proposition expressed in Examples 5.64 and 5.65. The expression *finds no evidence*, which is used in Example 5.65, is more tentative than *there is no evidence*, in that it suggests that evidence exist but remain as yet undiscovered. Admittedly, the proposition expressed in Example 5.66 is attributed.

[5.64] New MMR research **finds no evidence of immune system overload**.

[5.65] **There is no evidence that the triple MMR (measles, mumps, and rubella) vaccine causes immune system overload** and make children more vulnerable to serious bacterial infection, finds a study in Archives of Disease in Childhood.

However, the proposition is also rebutted in ways in which a strong commitment to the truth of the proposition is expressed. We see this in the following example, which is taken from a concordance line which occurs four times.

> [5.66] In theory, a baby could respond effectively to around 10,000 vaccines at any one time. **The baby's immune system can and does easily cope with the MMR and pneumococcal vaccines at the same time.**

It is interesting to compare the phraseological patterns used in Examples 5.64, 6.65 and 5.66. The degree of commitment to the truth of the proposition is greater in each. In fact, in Example 5.66, the expression is intensified: we are not only told the baby's immune system *can* cope, we are told it *does easily cope*. This examination of *immune* shows us that the word is used in the *NHSvax* corpus very much in the way it is used in other contexts, however, the view of the immune system expressed here is more limited than what we find in a general corpus of English. There is also a particular concern in *NHSvax* with the concept of vaccine overload. Having examined the way the word *immune* in used in the *NHSvax* corpus, the discussion moves on to an analysis of *immunity*.

#### 5.4.3.2 Meanings associated with the word *immunity*

Table 5.12 below shows the thirty most significant collocates of *immunity* in the *NHSvax* and *UkWaC* corpora, according to MI3[62].

[62] The *UkWaC* collocate list was compiled using Sketch Engine, http://www.sketchengine.co.uk, accessed 02/03/2013.

| | UkWaC | | NHSvax | |
|---|---|---|---|---|
| | **collocate** | **MI3** | **collocate** | **MI3** |
| 1 | *Infection* | 29.842 | *immunity* | 27.26 |
| 2 | *cell-mediated* | 29.108 | *herd* | 22.03 |
| 3 | *humoral* | 27.045 | *passive* | 20.22 |
| 4 | *prosecution* | 26.814 | *generated* | 18.35 |
| 5 | *immunity* | 26.561 | *active* | 18.03 |
| 6 | *innate* | 26.259 | *to* | 18.01 |
| 7 | *and* | 26.162 | *produce* | 17.62 |
| 8 | *diplomatic* | 25.901 | *the* | 17.43 |
| 9 | *to* | 25.785 | *boost* | 16.44 |
| 10 | *of* | 25.767 | *impaired* | 16.18 |
| 11 | *Innate* | 25.498 | *person* | 15.99 |
| 12 | *privileges* | 25.409 | *lowered* | 15.93 |
| 13 | *the* | 25.302 | *receiving* | 15.88 |
| 14 | *from* | 25.037 | *conferred* | 15.79 |
| 15 | *infection* | 24.943 | *factors* | 15.63 |
| 16 | *Inflammation* | 24.912 | *cells* | 15.45 |
| 17 | *Crown* | 24.595 | *is* | 15.43 |
| 18 | *Humoral* | 24.566 | *animation* | 15.18 |
| 19 | *Immunity* | 24.408 | *check* | 14.99 |
| 20 | *vaccination* | 24.340 | *affecting* | 14.83 |
| 21 | *mucosal* | 24.276 | *from* | 14.39 |
| 22 | *protective* | 24.270 | *immunodeficient* | 14.35 |
| 23 | *herd* | 24.245 | *resist* | 14.35 |
| 24 | *Cell-mediated* | 23.987 | *halting* | 14.35 |
| 25 | *disease* | 23.500 | *longer-term* | 14.33 |
| 26 | *in* | 23.373 | *of* | 14.30 |
| 27 | *Interest* | 23.292 | *natural* | 13.92 |
| 28 | *PII* | 22.985 | *disease* | 13.75 |
| 29 | *is* | 22.759 | *phenomenon* | 13.61 |
| 30 | *natural* | 22.712 | *build* | 13.48 |

**Table 5.12** The thirty most significant collocates of *immunity* in *ukWaC* and *NHSvax*, according to MI3.

Since *UkWaC* is a corpus of general English, unlike *NHSvax*, we see evidence of uses of *immunity* which are not related to the medical sense of the word. Thus, there are words such as *prosecution, diplomatic, privileges, Crown, interest,* and *PII* (public interest immunity). If we focus on collocates related to the medical sense of *immunity*, it is interesting that we find more differences between each of the corpora than similarities. The two collocate lists share the word *immunity* as well as terms related to disease, such as *disease* (both corpora) and *infection(s)* (UkWaC). They also share the words *herd* and *natural*. However, the *NHSvax*

collocate list lacks the adjective *protective* and, more interestingly, given their occurrences as collocates of *immune* in *UkWaC*, *cell-mediated*, *humoral* and *innate*. The adjective *mucosal* is also present in the *UkWaC* list and absent from the *NHSvax* list. However, although *cell-mediated* does not appear as a highly significant collocate of *immunity* in the NHSvax corpus, the noun *cells* does. The relative lack of significance of the words *cell-mediated*, *humoral*, *innate* and *mucosal* as collocates of *immunity* in the *NHSvax* corpus suggests that certain aspects of the immune system are not of central concern in the NHS texts. The words which have a strong collocational bond with *immunity* in *NHSvax* but not *UkWaC* are as follows: the adjectives *passive*, *active*, *longer-term* and *immunodeficient*; the passive forms *generated*, *impaired*, *lowered*, *conferred*;[63] a set of verbs which refer to the function of the immune system or the function of vaccines on the immune system, namely *produce*, *boost*, *resist*, and *build*; the continuous forms, *receiving*, *affecting* and *halting*; the semi-technical term *factors*; and the words *person*, *animation* and *phenomenon*.

It is interesting to note the high significance of the word *herd* as a collocate of *immunity* in the *NHSvax* corpus. This reflects the central concern of public vaccination policy, which is to achieve what is known as 'herd immunity', that is, to ensure that a sufficient proportion of the population is vaccinated against a given disease so that even unvaccinated individuals are unlikely to catch it. There are 37 citations of *herd immunity*. It occurs most frequently in the context of definitions or explanations of the term, for example:

> [5.67] **Herd immunity - The protection conferred on individuals who have not been immunised** because sufficient numbers of the rest of the population have been immunised.

[63] Concordance evidence shows that these terms are used in passive constructions in the *NHSvax* corpus when they collocate with *immunity*.

The most frequent extended phraseological pattern, though, is *(factors) affecting herd immunity* <10>. *Herd immunity* rarely occurs in contexts where explicit evaluation is evident, although there are two instances of the string *it's important that herd immunity is maintained*. The adjectives *active, passive,* seen in Table 5.11 above, reflect the distinction made in the science between active immunity and passive immunity. The DoH evidently considers it important to explain the difference between different methods of vaccination and different types of vaccine. There are 13 instances each of *active immunity* and *passive immunity*. All occur as part of definitions of the terms:

> [5.69] **Active immunity is generated** by the body when the immune system is triggered to produce antibodies, either by the immunisation or the disease.

> [5.69] **Passive immunity is generated** when the body is given proteins that act as antibodies instead of making them itself.

Similarly, the six instances in which *cells* and *immunity* collocate, pertain to explanations of a type of vaccine:

> [5.70] Acellular vaccine Without whole cells. **An acellular vaccine contains only parts of cells which can produce immunity in the person receiving the vaccine** (see DTaP).

As mentioned above, the adjective *natural* occurs as a significant collocate of *immunity*. There are a mere five citations of the string *natural immunity*, but the concept would seem to be an important one in the science of vaccination. There are 109 citations of *natural* in the *NHSvax* corpus and it occurs most frequently in contexts in which an opposition is set up between immunity gained through catching a disease (natural immunity) and immunity gained through vaccination. *Natural* most frequently occurs in the following strings:

*natural infection* <23>; *natural measles* <22>; *natural history* (of the development of a disease) <10>; *natural disease* <8>; *natural immunity* <8>; *natural infections* <6>; *natural polio* <6>; *natural rubella* <6>.

Instances of *natural* + A DISEASE typically occur in the context of descriptions of the effects of diseases, for example:

> [5.71] Subacute sclerosing panencephalitis (SSPE) (A delayed complication of measles that causes brain damage and death.) is a rare condition that can develop some years after **natural measles infection** and causes brain damage and death.

However, the environments in which *natural immunity* occurs indicate that the DoH perceives a need to counter a widespread belief among the public that acquiring immunity to a disease through natural infection is better than gaining it through vaccination. The following question and answer sequences from an FAQ section of one of the websites illustrates this.

> [5.72] Is it true that **the immunity a person gets from natural infection** is likely to last for life? - **The problem is that to gain natural immunity** people have to catch the natural infection and the risk of complications is much greater than that of an adverse effect from a vaccine.
>
> [5.73] Wouldn't it be far better for children to **have the natural disease to give them longer-term immunity**? It is true that the immunity a person gets from natural infection is likely to last for life. **The problem is that to gain natural immunity** people have to catch the natural infection and the risk of complications is much greater than that of an adverse effect from a vaccine.

Parents are encouraged to accept vaccination, and to accept repeated doses of vaccines, even if the child already has natural immunity. One of the statistically most salient collocates of *immunity* is *boost*. *Boost* occurs in the environment of *immunity* in contexts in which the imperative to administer (or accept) repeated doses of a vaccine is

justified, as Example 5.74 shows. *Boost*, in turn, also frequently collocates with *antibodies*, as we see in Example 5.75.

> [5.74] Repeating a vaccination will **boost any existing immunity**.
>
> [5.75] Although your baby had measles at six weeks we would still advise that he or she should have MMR. This will help protect your child against mumps and rubella and will also **boost the antibodies your baby has already developed against measles.**

As was seen with the evidence for *immune* in the *NHSvax* corpus, there is acknowledgement that vaccines may be harmful to people with weakened immune systems:

> [5.76] In addition, influenza immunisation of staff may reduce the transmission of influenza to vulnerable patients, **some of whom may have impaired immunity** and thus reduced protection from any influenza vaccine they have received themselves.
>
> [5.77] The bacteria in the vaccine are alive but have been modified so that they do not cause disease (**except on occasion in people with lowered immunity**, for example due to HIV infection).

*Immunity* is therefore frequently used in the *NHSvax* corpus in contexts in which the processes of acquiring immunity, whether naturally or through vaccination, are explained. The idea that it is better to acquire immunity naturally rather than by means of vaccination is rebutted. *Immunity* is frequently used as part of the expression *herd immunity* and this expression is used mainly in order to define the term as part of a strategy to encourage vaccine uptake.

### 5.4.4 Summary of Section 5.4

The pathogenic, or germ, theory of disease is expressed through semantic sequences such as A BACTERIUM/VIRUS CAUSES A DISEASE, A DISEASE/VIRUS CAUSES A

PHYSICAL REACTION, A DISEASE CAUSES DAMAGE and A DISEASE CAUSES DEATH. Vaccination is presented as a means of avoiding potential harm. It is represented as enhancing, or 'boosting', the immune system. The terms *vaccination* and *immunisation* often occur in the environment of forms of the lexeme *PROTECT*. Propositions related to the safety of vaccines are often intensified through use of the superlative *safest*, in, for example, *vaccinations are the safest way*. The way in which the immune system is represented in the *NHSvax* corpus is similar in many respects to the way in which it is represented in *UkWaC*. The main differences are that certain collocates of *immune* and *immunity* which are highly statistically significant in *UkWaC*, such as *cell-mediated*, *humoral*, *innate* and *mucosal* are not significant collocates in *NHSvax*. There is concern in the *NHSvax* data, as one might expect, with assuring the public that acquitting immunity through vaccination is just as good as, if not better than, acquiring it naturally. Vaccines are said to 'boost' the immune system. It is acknowledged that the immune systems of some groups of people may be compromised, therefore there are several citations of *weakened immune system*. The producers of the texts in the *NHSvax* corpus hold that a widespread belief among the public is that multiple vaccines compromise the immune system. The expression *immune system*, therefore, collocates *with weaken* and *overload*. The concept of *natural immunity* is presented in binary opposition to immunity acquired through vaccination. The argument that acquiring immunity through catching a disease 'naturally' is better than vaccination is rebutted here.

## 5.5 Expressions of risk, safety, and danger in the *NHSvax* corpus

### 5.5.1 Vaccination as protection

As one might predict, vaccination is represented as something which is safe and which protects the individual. The verb *protect* frequently occurs in the same environment as the superlative *safest*. This is evident from the collocate lists in Table 5.13 below. The reason why *protect* and *safest* share so many collocates is that they occur again and again in the string *the safest way to protect*. The item *your* appears in the collocate lists because the writers of the NHS website material often appeal to the reader's own sense of self-interest or to their identity as a parent. This is shown in the following two examples, which occur 17 and 16 times, respectively, in the *NHSvax* corpus.

[5.78] … **the safest way to protect** your child for life.

[5.79] … **the safest way to protect** your health for life.

| | collocates of *protect* | MI3 | collocates of *safest* | MI3 |
|---|---|---|---|---|
| 1 | *protect* | 28.10 | *safest* | 25.75 |
| 2 | *against* | 23.36 | *way* | 23.41 |
| 3 | *way* | 21.91 | *protect* | 21.68 |
| 4 | *safest* | 21.88 | *your* | 19.32 |
| 5 | *to* | 20.94 | *effective* | 17.60 |
| 6 | *your* | 19.70 | *to* | 17.60 |
| 7 | *does* | 19.44 | *the* | 17.45 |
| 8 | *yourself* | 19.25 | *mmunisation* | 17.22 |
| 9 | *the* | 19.00 | *most* | 16.35 |
| 10 | *them* | 18.31 | *child* | 16.31 |
| 11 | *child* | 17.60 | *health* | 15.54 |
| 12 | *vaccine* | 17.21 | *immunisation* | 15.41 |
| 13 | *from* | 17.17 | *among* | 15.21 |
| 14 | *not* | 17.17 | *is* | 14.46 |
| 15 | *children* | 16.82 | *protecting* | 14.40 |

**Table 5.13** Top fifteen collocates of *protect* and *safest* in the *NHSvax* corpus, according to MI3

The most common phraseological pattern in which *protect* is used, however, is V *against* n. There are 144 instances of the string *protect against* in the *NHSvax* corpus. The noun group which follows *protect against* refers to a source of harm, specifically, a disease.

> [5.80] The overwhelming weight of evidence shows that **MMR is the safest way to protect against measles, mumps and rubella**, and the number of studies demonstrating this is growing.

Variations on this pattern are V n *against* n and V n *from* n. These are exemplified below.

> [5.81] MMR is **the safest way to protect** your child **against** measles, mumps and rubella.
>
> [5.82] Immunisation is **the safest way to protect** your baby **from** very serious diseases such as measles, meningitis, whooping cough and polio.

Of course, where the verb *protect* is concerned, the pattern V n *from* n differs from V n *against* n in that, in the case of the former, there is the potential for the noun group which forms part of the prepositional phrase to refer to an outcome rather than a source of harm. However, in the *NHSvax* data, whether *protect against* or *protect from* is used, the noun group in the prepositional phrase invariably refers to a disease.

### 5.5.2 Vaccination and risk

As was observed in Section 5.4, in the analysis of expressions of causation and disease, there is a lot of emphasis in the data on the risks posed by disease. The possible risks of vaccination are downplayed. The corpus evidence for *risk* and *risks* in the *NHSvax* corpus shows many of the patterns of use observed byHamilton, Adolphs and Nerlich (2007) in the health-related

pedagogical data of *CANCODE*. The top forty collocates of *risk* in *NHSvax,* according to MI3, are shown in Table 5.14. Reflecting the findings of Hamilton, Adolphs and Nerlich (2007), there are several items among the top forty collocates which indicate the assessment of degrees of risk, such as *increased, greater, higher, reduce, reduces, slight, increase,* and *high*. There are also a number of items denoting possible outcomes. These are words such as *developing, catching, disease, infection, paralytic* and *pneumococcal*. Of course, these words indicate effects, but do not hint at possible causes. Closer examination is necessary to uncover how the sources of risk are expressed, although we do see the word *associated,* which can be used to suggest a causal relationship. As we see in Table 5.14, a handful of collocates refers to groups of people. These are *groups* and *children*. Words such as these which denote specific groups of individuals reflect the main concern of epidemiology, which provides the methodological underpinnings of public health policy. Interestingly, the negative particles *not* and *no* appear in the collocate list.

| | **collocate** | **MI3** | | **collocate** | **MI3** |
|---|---|---|---|---|---|
| 1 | *risk* | 29.23 | 21 | *associated* | 17.65 |
| 2 | *at* | 22.87 | 22 | *not* | 17.44 |
| 3 | *of* | 22.76 | 23 | *children* | 17.42 |
| 4 | *the* | 21.81 | 24 | *and* | 17.37 |
| 5 | *increased* | 21.58 | 25 | *an* | 17.35 |
| 6 | *groups* | 21.12 | 26 | *most* | 16.86 |
| 7 | *developing* | 19.88 | 27 | *there* | 16.81 |
| 8 | *greater* | 19.05 | 28 | *be* | 16.62 |
| 9 | *is* | 18.99 | 29 | *reduces* | 16.30 |
| 10 | *a* | 18.96 | 30 | *slight* | 16.14 |
| 11 | *from* | 18.73 | 31 | *who* | 16.00 |
| 12 | *to* | 18.73 | 32 | *for* | 16.00 |
| 13 | *clinical* | 18.69 | 33 | *increase* | 15.98 |
| 14 | *higher* | 18.62 | 34 | *those* | 15.92 |
| 15 | *catching* | 18.61 | 35 | *paralytic* | 15.88 |
| 16 | *are* | 18.59 | 36 | *no* | 15.87 |
| 17 | *in* | 18.23 | 37 | *that* | 15.81 |
| 18 | *disease* | 17.83 | 38 | *greatest* | 15.63 |
| 19 | *reduce* | 17.70 | 39 | *high* | 15.57 |
| 20 | *infection* | 17.69 | 40 | *pneumococcal* | 15.56 |

**Table 5.14** Top forty collocates of *risk* in the *NHSvax* corpus, according to MI3

In order to gain a more detailed picture of the contexts in which *risk* and its collocates are used, it is useful to examine the clusters in which *risk* occurs. Table 5.15 below shows all of the meaningful two- to six-word clusters containing *risk* in the *NHSvax* corpus. The first thing which is apparent from the data shown in Table 5.15 is that *at risk* <298> is the most frequent phraseological pattern. As Hamilton, Adolphs and Nerlich (2007) observe, *at risk* is used to refer to groups of people who are considered especially vulnerable to developing a disease or condition. We see this in particular in the profile of *risk* in the *NHSvax* corpus. The expression *at risk* frequently occurs in longer strings such as *at risk groups* <48>, *at risk group* <10>, and *at risk children* <27>.Other clusters referring to groups of people are *risk groups* <151>, *clinical risk group* <15> and *high risk groups* <9>. The high frequency of expressions referring to groups reinforces the impression that a major concern in the *NHSvax* data is identifying target groups of people. Concordance evidence shows that the expressions *(at) risk groups/children* tend to be used in policy documents. Typical examples of *(at) risk groups/children* are:

> [5.83] the JCVI recommended that **the following risk groups be offered BCG vaccination**
>
> [5.84] It is important to make progress towards the WHO recommended level of 75% uptake for those aged 65 years and over **and improve uptake rates for the under-65 clinical risk groups**, and healthcare workers.
>
> [5.85] If the programme is to be truly effective then we need to increase this figure and also **improve markedly upon the uptake rates seen in the targeted at risk groups** and in health professionals.
>
> [5.86] Also, unprotected children can pass on these diseases to other **at risk children** and pregnant women.

| | Cluster and raw frequency | | | Cluster and raw frequency | |
|---|---|---|---|---|---|
| 1 | *at risk* | 298 | 51 | *at higher risk* | 11 |
| 2 | *the risk* | 261 | 52 | *risk factors* | 11 |
| 3 | *risk groups* | 151 | 53 | *do not increase the risk* | 10 |
| 4 | *increased risk* | 90 | 54 | *in a clinical risk group* | 10 |
| 5 | *high risk* | 68 | 55 | *the risk of bleeding* | 10 |
| 6 | *clinical risk* | 57 | 56 | *given the risk* | 10 |
| 7 | *risk of developing* | 52 | 57 | *increased the risk* | 10 |
| 8 | *an increased risk* | 51 | 58 | *risk of getting* | 10 |
| 9 | *at risk groups* | 48 | 59 | *risk of causing* | 10 |
| 10 | *a risk* | 47 | 60 | *risk from flu* | 10 |
| 11 | *greater risk* | 42 | 61 | *at most risk* | 10 |
| 12 | *most at risk* | 41 | 62 | *at risk group* | 10 |
| 13 | *risk group* | 40 | 63 | *risk of bleeding* | 10 |
| 14 | *risk of autism* | 36 | 64 | *their risk* | 10 |
| 15 | *clinical risk groups* | 35 | 65 | *most risk* | 10 |
| 16 | *higher risk* | 35 | 66 | *risk types* | 10 |
| 17 | *reduce the risk* | 31 | 67 | *may be at risk* | 9 |
| 18 | *risk of catching* | 28 | 68 | *presents a greater risk* | 9 |
| 19 | *at risk children* | 27 | 69 | *the risk of polio* | 9 |
| 20 | *be at risk* | 26 | 70 | *at an increased risk* | 9 |
| 21 | *increase the risk* | 25 | 71 | *the risk of autism* | 9 |
| 22 | *to reduce the risk* | 24 | 72 | *the risk groups* | 9 |
| 23 | *children at risk* | 24 | 73 | *high risk groups* | 9 |
| 24 | *a greater risk* | 23 | 74 | *more at risk* | 9 |
| 25 | *at risk of catching* | 22 | 75 | *risk of polio* | 9 |
| 26 | *high risk hpv* | 20 | 76 | *risk of exposure* | 9 |
| 27 | *risk of infection* | 18 | 77 | *risk of complications* | 9 |
| 28 | *slight risk* | 18 | 78 | *a risk factor* | 9 |
| 29 | *increased risk of autism* | 17 | 79 | *low risk* | 9 |
| 30 | *at greater risk* | 17 | 80 | *greatest risk* | 9 |
| 31 | *associated with an increased risk* | 16 | 81 | *immunisation presents a greater risk* | 8 |
| 32 | *are at risk* | 16 | 82 | *in clinical risk groups* | 8 |
| 33 | *the risk of developing* | 15 | 83 | *the following risk groups* | 8 |
| 34 | *is at risk* | 15 | 84 | *the at risk groups* | 8 |
| 35 | *reduces the risk* | 15 | 85 | *the clinical risk groups* | 8 |
| 36 | *clinical risk group* | 15 | 86 | *be at greater risk* | 8 |
| 37 | *there is a risk* | 14 | 87 | *at greatest risk* | 8 |
| 38 | *a high risk* | 14 | 88 | *at risk patients* | 8 |
| 39 | *is a risk* | 14 | 89 | *risk from hib* | 8 |
| 40 | *at increased risk* | 14 | 90 | *unnecessary risk* | 8 |
| 41 | *increased risk of developing* | 13 | 91 | *subcutaneous injection to reduce the risk* | 7 |
| 42 | *no increased risk* | 13 | 92 | *vaccines do not increase the risk* | 7 |
| 43 | *there is an increased risk* | 12 | 93 | *paper concludes no increased risk* | 7 |
| 44 | *a clinical risk group* | 12 | 94 | *who is most at risk* | 7 |
| 45 | *a slight risk* | 12 | 95 | *will be at risk* | 7 |
| 46 | *risk factor* | 12 | 96 | *leave children at risk* | 7 |
| 47 | *particular risk* | 12 | 97 | *reducing the risk* | 7 |
| 48 | *any risk* | 12 | 98 | *at risk from flu* | 7 |
| 49 | *not associated with an increased risk* | 11 | 99 | *is most at risk* | 7 |
| 50 | *at risk of developing* | 11 | 100 | *the overall risk* | 7 |

**Table 5.15** Meaningful two- to six-word clusters containing *risk* in *NHSvax*

As one would expect, the potential harm associated with risk is the development of a specific disease. The data shows that one of the main messages the NHS aims to convey is that it is not only vulnerable groups who are at risk, but people who have not been vaccinated. The following examples illustrate this.

> [5.87] **… anyone who is not fully protected against tetanus is at risk from the disease.** The only way to protect yourself from tetanus is by immunisation.
>
> [5.88] Although all three diseases are uncommon in the UK, **children who are not immunised are still at risk of catching them.**

Where the MMR vaccine is concerned, the NHS tries to counter arguments that the vaccine poses risks. We see expressions such as *no increased risk of* and *no evidence/proof that the (ADJECTIVE) risk is (of x) is higher)*:

> [5.89] The paper concludes **no increased risk of autism (AD) following exposures to wild measles and vaccinations with monovalent measles**, and Urabe or Jeryl-Lynn variants of MMR
>
> [5.90] There is **no proof that the overall risk of autism is higher in children who were vaccinated with MMR** or of an increase in autism prevalence associated with the use of the MMR vaccine.

The argument that single vaccines are a safer alternative to MMR is often countered by expressing the proposition that single vaccines pose a different sort of risk. The potential outcome is of harm occurring as a result of the risks posed by disease:

> [5.91] Giving the vaccines separately **leaves children at risk of catching measles, mumps or rubella** between the three injections.
>
> [5.92] The recommendations of intervals between vaccines by those doctors offering single doses are arbitrary and **leave children at risk of catching either measles or mumps**.

There is evidence in the data of tension between the public good and the concerns of the individual. The main concern of public health policy regarding vaccination is to reduce cases of diseases in the population as a whole. This is illustrated in the following example.

> [5.93] **We need to see** continued increases in uptake to reduce the risk of further measles cases.

What is interesting in the above example, is the ambiguity implicit in the use of the exophoric reference *we*. It is unclear if inclusive or exclusive *we* is intended. *We* might refer to the NHS alone or it might refer to the country in general, thus including the reader. This example might therefore be interpreted as constituting an appeal to the reader's sense of public duty. The following example is interesting in terms of the way it puts forward an argument for promoting herd immunity by implicitly appealing to the reader's sense of self-interest. Again, there is ambiguity in the use of exophoric reference.

> [5.94] **If a child is not immunised, they** will be at risk from catching the disease and **will rely on other people immunising their children** to avoid becoming infected.

In Example 5.94, the reader is not addressed with the pronoun *you* and references to *your child* and *he* or *she* are avoided. Instead, reference is made to a hypothetical child and the generic pronoun *they* is used. It is not made clear whether the reader is expected to interpret *a child* and *they* as referring to their own child or someone else's. Nor is it made clear whether the reader is positioned as belonging to the group of parents who immunise their child and thus help protect other children or to the group of parents who leave their child unimmunised and thus leave their child relying on other parents. This ambiguity makes it possible for the writers to simultaneously appeal to parents' sense of civic duty and to their feelings of responsibility to their own child.

### 5.5.3 The dangers of vaccination

Finally, it is acknowledged that vaccination is not entirely risk-free. It is acknowledged that some children may come to harm as a result of vaccination. The term *damage* is not normally used to refer to undesirable effects, though. There are 173 citations of *damage* and the fact that the government acknowledges that damage from vaccines occurs is reflected in the 13 occurrences of the string *vaccine damage payment scheme*. However, the remaining citations of *damage* occur in the context of warnings about the risks of particular diseases. A typical example is:

> [5.95] **The complications include chest infections, fits, encephalitis (swelling of the brain), and brain damage**. In very serious cases, measles kills.

Instead of using damage to refer to undesirable effects, the prosodically more neutral terms *reaction(s)* or *(side) effect(s)* are used, often together with words which signal temporal relations, such as *following* or *after*. The words *reaction(s)* and *side-effects* often collocate with forms of the lexeme *OCCUR*, which is also a prosodically neutral term. There are 427 citations of *effects*, of which 331 citations are for *side-effects* and 27 for *adverse effects*, and 397 of *OCCUR*. Examples of *side-effects*, *reactions* and *OCCUR* are:

> [5.96] **More serious side effects are rare** but include: fever headaches dizziness nausea and vomiting swollen glands.
>
> [5.97] **Local reactions**, such as mild soreness, redness and hardening where the injection is given **may occur**, lasting no longer than 1-3 days. Occasionally **a mild fever or muscle pain may occur**.

Interestingly, the term *adverse effects* occurs mainly in the context of a rebuttal, as the example below shows:

> [5.98] In those studies where MMR has been compared with the component vaccines **there is no suggestion that MMR causes more adverse effects than measles vaccines alone.**

Where it is acknowledged that vaccines can pose risks, the risks of the vaccine are usually presented as standing in a binary opposition to the risks posed by the disease. The risks of the disease are unfavourably compared with the risks of the vaccine, as corpus evidence for *risks* shows:

> [5.99] Comprehensive analysis of the reported adverse reactions established that serious events causally related to MMR vaccine **are rare and greatly outweighed by the risks of natural MMR diseases.**
>
> [5.100] **The benefits of vaccination far outweigh the risks**, if any, of exposure to thiomersal-containing vaccines.
>
> [5.101] **The benefits of DTwP outweigh any risks**, if any, from the thiomersal present in the vaccine.

Use of the expression *the risks* presupposes that there are risks. However, as we see from the three examples above, the proposition that there are risks is often hedged with the phrase *if any*, while the proposition that the benefits outweigh the risks is often intensified with terms such as *greatly* or *far*.

### 5.5.4 Summary of Section 5.5

The mainstream medical-scientific discourse relating to risk, as evidenced in the *NHSvax* corpus, displays the sorts of patterns one would expect. Assessments of degrees of risk are expressed using forms of the lemmas *INCREASE* and *HIGH*, and so on. The risks posed by

diseases are represented as greater than those posed by vaccines. One of the typical phraseological patterns used to express this proposition *is the benefits (…) outweigh the/any risks* or *the risks are (…) outweighed by the benefits*. Vaccination is represented as safe and forms of the lexeme *PROTECT* collocate with *RISK*. Arguments that vaccines are risky are rebutted with expressions such as *(there is) no evidence of increased risk*. Where it is acknowledged that there is an element of risk posed by vaccines, prosodically neutral terms such as *reactions*, *side effects* and *OCCUR* are used to refer to the outcomes. Some of the data shows evidence of tension between concern for the public good. There is a high frequency of expressions such as *(at) risk groups*, and so on, owing to the fact that the aim of public health policy regarding vaccination is to ensure herd immunity. As we have seen, there is some evidence in the *NHSvax* texts of tension between concerns for the public good and perceptions of parents' concerns for their own children. This is witnessed in particular in the case of Example 5.94, discussed in Section 5.5.2 above, in which an assertion was expressed in such a way as to appeal simultaneously to a parent's sense of responsibility for their child and to their sense of civic responsibility.

## 5.6 Conclusion

The analysis here has shown that mainstream medical-scientific discourse is characterised by patterns of language use which reflect the pathogenic model of disease. Thus, we see repeated instances of the semantic sequence A BACTERIUM/VIRUS CAUSES A DISEASE, A DISEASE/VIRUS CAUSES A PHYSICAL REACTION, A DISEASE CAUSES DAMAGE and A DISEASE CAUSES DEATH. Furthermore, explanation of the way in which the immune system functions is largely limited to descriptions of antibody response. Other aspects of knowledge

about the immune system are overlooked. Vaccination is evaluated positively and it represented as enhancing the immune system. We see this in expressions such as *the benefits of boosting her protection; extra protection;* and *good protection.* Statistically salient collocates are therefore *trigger, response* and *respond*. Natural immunity is presented in opposition to vaccine-acquired immunity. The argument that it is better to acquire immunity naturally than to have it conferred via vaccination is rebutted. As for expressions of risk, these conform to the patterns of language which are typical of risk discourse in the field of healthcare. We therefore see assessments of degrees of risk expressed using forms of the lemmas *INCREASE* and *HIGH*. We also find expressions such as *at risk (groups/children),* which are used to identify vulnerable groups. The causal hypothesis is typically expressed using the nominalized form *link* as part of the expression *link between mmr and autism*. The proposition is usually rebutted using the expression *there is no evidence of a link.* Use of these expressions emphasizes the uncertainty of the science. Potential harm from vaccines is expressed using the terms *side effects* and *adverse reactions* and the verb *OCCUR* is often used in such contexts. This chapter has highlighted phraseological patterns which are typical of medical-scientific discourse about vaccination, risk, health, immunity, and the MMR-autism causal hypothesis. Conclusions drawn in this chapter on these topics are drawn on to inform the analysis in the next chapter, in which the typical discoursal patterns evident in the *JABS* corpus are analysed. The ways in which people express warrants for expertise and the ways in which attribution is expressed are discussed in the subsequent chapter.

# CHAPTER SIX

# THE EXPRESSION OF EPISTEMOLOGICAL CLAIMS IN THE *JABS* CORPUS

## 6.1 Introduction

One of the objectives of this study is to discover the extent to which, and how, writers represented in the *JABS* corpus draw on different discourses in pursuit of their interactional goals. The previous chapter examined the ways in which notions of health, immunity, and risk, in relation to vaccination and the MMR vaccine, are expressed in the *NHSvax* corpus. The objective was to identify the sort of lexico-grammatical patterns which typify mainstream medical-scientific discourse about health and immunity, and about the safety and risks of MMR in particular, and vaccination in general. The aim of the current chapter is to describe the typical lexico-grammatical patterns evident in the *JABS* corpus, and to identify the similarities and differences between salient patterns uncovered here and those which have been identified in the *NHSvax* corpus. By highlighting the similarities and differences between the two corpora, we will be able to distinguish the extent to which, and how, mainstream scientific discourse is drawn on in the *JABS* corpus, and the extent to which other discourses are apparent. The process of identifying the similarities and differences between the two corpora starts with a comparison of word frequencies and keywords. The aim of the keyword analysis is to highlight items which warrant closer examination. Keywords relating to the realization of propositions concerning the causal hypothesis, health, immunity, and risk

are examined in this chapter. Keywords relating to the expression of warrants for expertise and the framing of intertextual elements of text are examined in the following chapter.

This chapter is structured as follows. Section 6.2 examines the most frequently occurring words in the corpus, before moving on to an examination of the socially salient keywords. Section 6.3 examines the ways in which propositions relating to the causal hypothesis are expressed in the *JABS* corpus. Section 6.4 examines the ways in which beliefs about health and the immune system are expressed. Section 6.5 considers the ways in which notions of risk are expressed. Section 6.6 concludes the chapter. With each topic which is examined, comparisons are drawn with the findings from the analysis of the *NHSvax* corpus.

## 6.2 Frequency data and keywords in the *JABS* corpus

### 6.2.1 The 200 most frequent words in the *JABS* corpus

On examining the 200 most frequent words in the *JABS* corpus (see Appendix 6), we find several similarities between these and the most frequent 200 words in the *NHSvax* corpus. Once again, as is typical of a small, specialized corpus, certain lexical words come relatively high up the frequency list, when compared with what one might expect to see in a larger, general corpus. To illustrate this point more clearly, the top fifty words in the *BNC World*, and in the *JABS* and *NHSvax* corpora are shown, with their frequencies, in Table 6.1 below.

| rank | **BNC** | | **JABS** | | **NHSvax** | |
|---|---|---|---|---|---|---|
| | **word** | **frequency** | **word** | **frequency** | **word** | **frequency** |
| 1 | *the* | 6055105 | *the* | 202217 | *the* | 31453 |
| 2 | *of* | 3049564 | *#* | 131420 | *#* | 28633 |
| 3 | *and* | 2624341 | *of* | 108591 | *of* | 19112 |
| 4 | *to* | 2599505 | *to* | 102319 | *and* | 14542 |
| 5 | *a* | 2181592 | *and* | 90767 | *to* | 14104 |
| 6 | *in* | 1946021 | *a* | 73498 | *in* | 11744 |
| 7 | *#* | 1604421 | *in* | 71240 | *a* | 9851 |
| 8 | *that* | 1052259 | *that* | 51729 | *is* | 7355 |
| 9 | *is* | 974293 | *is* | 49065 | *vaccine* | 7228 |
| 10 | *it* | 922687 | *i* | 35571 | *for* | 5991 |
| 11 | *for* | 880848 | *for* | 34366 | *that* | 5520 |
| 12 | *was* | 863917 | *it* | 30464 | *be* | 5209 |
| 13 | *i* | 732523 | *on* | 28530 | *are* | 4220 |
| 14 | *on* | 731319 | *have* | 28367 | *or* | 4128 |
| 15 | *with* | 659997 | *with* | 27195 | *this* | 3905 |
| 16 | *as* | 655259 | *are* | 25251 | *mmr* | 3677 |
| 17 | *be* | 651535 | *was* | 24899 | *with* | 3664 |
| 18 | *he* | 593609 | *not* | 24706 | *immunisation* | 3447 |
| 19 | *you* | 588503 | *by* | 24637 | *have* | 3337 |
| 20 | *at* | 524075 | *this* | 24427 | *on* | 3229 |
| 21 | *by* | 513444 | *as* | 23856 | *as* | 3133 |
| 22 | *are* | 458368 | *be* | 23494 | *children* | 3031 |
| 23 | *this* | 454419 | *vaccine* | 23182 | *it* | 3010 |
| 24 | *have* | 448684 | *you* | 22045 | *by* | 2958 |
| 25 | *but* | 446783 | *from* | 20117 | *not* | 2867 |
| 26 | *not* | 431075 | *children* | 16430 | *health* | 2837 |
| 27 | *from* | 425987 | *at* | 16056 | *at* | 2763 |
| 28 | *had* | 413144 | *or* | 16033 | *from* | 2625 |
| 29 | *his* | 410294 | *they* | 15886 | *vaccines* | 2436 |
| 30 | *they* | 376289 | *has* | 15190 | *uk* | 2365 |
| 31 | *or* | 370166 | *but* | 14892 | *disease* | 2324 |
| 32 | *which* | 366196 | *had* | 13024 | *will* | 2323 |
| 33 | *an* | 338743 | *an* | 12913 | *has* | 2246 |
| 34 | *she* | 325351 | *vaccines* | 12758 | *can* | 2213 |
| 35 | *were* | 308363 | *been* | 12438 | *you* | 2115 |
| 36 | *her* | 304311 | *we* | 12260 | *was* | 2097 |
| 37 | *we* | 300833 | *who* | 12024 | *been* | 1904 |
| 38 | *one* | 290466 | *if* | 11846 | *an* | 1839 |
| 39 | *there* | 285870 | *mmr* | 11742 | *if* | 1826 |
| 40 | *all* | 277566 | *autism* | 11680 | *there* | 1813 |
| 41 | *been* | 260360 | *â* | 11632 | *measles* | 1784 |
| 42 | *their* | 254603 | *all* | 11489 | *should* | 1770 |
| 43 | *if* | 253804 | *he* | 11380 | *your* | 1730 |
| 44 | *has* | 252703 | *my* | 11271 | *vaccination* | 1686 |
| 45 | *will* | 251179 | *there* | 11129 | *who* | 1654 |
| 46 | *so* | 239549 | *were* | 10924 | *s* | 1640 |
| 47 | *would* | 229699 | *about* | 10813 | *they* | 1617 |
| 48 | *no* | 229618 | *their* | 10524 | *information* | 1608 |
| 49 | *what* | 225524 | *which* | 10516 | *their* | 1553 |
| 50 | *can* | 211093 | *one* | 10253 | *may* | 1530 |

**Table 6.1** The fifty most frequent words, with frequencies, in the *BNC*, *JABS* and *NHSvax* corpora

As was observed with the *NHSvax* corpus, a small number of lexical items relating to the topic of childhood vaccination are highly frequent. The fact that there are fewer lexical words among the fifty most frequent words in *JABS* than in *NHSvax* is due, at least in part, to the relative differences in size between the two corpora. Alternatively, it might also be due to differences in the type/token ratio (TTR). The standardized TTR of the *JABS* corpus is 42.17, while the standardized TTR of the *NHSvax* corpus is 36.37.[64] The lower TTR of the *NHSvax* corpus may result from the greater lexical density of the *NHSvax* corpus, compared with *JABS*, possibly resulting from the lack of interactive data in the former. When the *JABS* wordlist is compared with the *BNC*, noticeable features are the relatively high frequencies of *I* and *my*, suggesting that the corpus data contains a higher than average occurrence of accounts of personal experience, and the relatively high frequencies of *have*, *are*, *has*, and *been*, and corresponding low frequencies (relatively) of *was*, *had*, and *were*, which suggest a concern with present time events. The negative particle *not* is also relatively frequent in *JABS*.

The lexical words which rank among the 200 most frequent words in the *JABS* corpus belong to many of the same semantic categories identified in the *NHSvax* corpus. There are therefore several words relating to childhood vaccination, the MMR-autism causal hypothesis, and health, for example *vaccine, children, vaccines, mmr, autism, health, vaccination, measles, child, disease, virus, vaccinated, vaccinations, autistic*, and *diseases*. While there is much concern with vaccination and vaccines in general, *mmr* is the only term among the top 200 words which refers to a specific vaccine. Similarly, *measles* is the only communicable disease referred to in the wordlist. This reflects the fact that, although the JABS group is concerned

[64] Because type/token ratios vary according to the length of text, *WordSmith Tools* calculates a standardized type/token ratio, based on the average TTR per 1,000 words of text.

with vaccine damage in general, the MMR-autism controversy was highly topical at the time the data was gathered. As was seen when the *NHSvax* corpus frequency list was examined, the words *risk* and *safety* occur among the top 200 words, as do a number of words related to causation, such as *cause, effects, link,* and *damage*.

### 6.2.2 Keywords in the *JABS* corpus

To gain a clearer picture of the words which are statistically significant, we need to turn to the keywords. Appendix 7 shows the keywords in the *JABS* corpus (according to log-likelihood, and using the *BNC World* as reference corpus) with a distribution across at least 10% of corpus files. Appendix 8 shows the same keywords grouped according to semantic and functional categories. Many of the categories are identical to those used to describe the contents of the *NHSvax* corpus, although one category (the 'concerns and business of JABS' category) was added to reflect a key difference in the *JABS* corpus. The criteria for selecting *JABS* corpus keywords for closer examination are the same as those used when selecting candidate keywords from the *NHSvax* keyword list: the terms selected are those which are considered most likely to provide useful information regarding the ways in which notions of health risk and immunity are expressed or the ways in which warrants for expertise are expressed; terms which are considered less central to expressions of these notions are rejected; terms which are relevant but which have a strong collocational relationship with other terms which have been selected are rejected. The keyword categories are as follows:

1. vaccination and pharmaceutical products;
2. health, disease and medical conditions;
3. people, places and institutions;

4. science, medicine and healthcare;
5. sources;
6. the concerns and business of JABS;
7. likelihood, causation and temporal relations;
8. expressions with modal meaning;
9. conditionality;
10. expressions of evaluation;
11. other nouns, verbs, and phrases referring to events, actions, and states of affairs;
12. other grammatical items;
13. miscellaneous items.

The words in the first category of keywords listed in Appendix 8 are shown in Table 6.2 below. These words are similar to those in the same category in the *NHSvax* corpus, although there are fewer terms which refer to vaccines and there is less concern with types of vaccines, other than a preoccupation with single, combined and multiple vaccines. As was the case when examining the *NHSvax* corpus, it was decided to include the words *vaccine, vaccines* and *mmr* in the candidate list. Reflecting the fact that the *JABS* data represents the voices of a wider range of people than are represented in the *NHSvax* data, there are more words here which belong in the sub-category 'administering vaccines', since the informal terms *jab, shots* and *shot* appear. The term *jabs* is also highly statistically significant, partly because it also refers to the name of the JABS group. This term is the only keyword to be listed in two categories. This is because 50% of the occurrences of *jabs* in the *JABS* corpus refer to injections and 50% to the *JABS* group and website. The words in this sub-cteogry have been

rejected from the candidate list on the same grounds for which similar terms were excluded from the *NHSvax* collocate list.

| | | |
|---|---|---|
| **vaccination and pharmaceutical products** | vaccines | *vaccine, vaccines, mmr* |
| | vaccination | *vaccination, vaccinations, vaccinated, immunisation, vaccinate, immunization, vaccinating* |
| | vaccination: pre-modifiers | *fully (vaccinated)* |
| | vaccine types | *single, combined, multiple* |
| | administering vaccines | *jabs, jab, dose, doses, injection, shots, shot* |
| | other pharmaceutical products | *drugs, medicines, products, product* |
| | vaccine additives | *mercury, thimerosal* |
| | toxicity | *toxic* |

**Table 6.2** The *JABS* keywords in the 'vaccination and pharmaceutical products' category

The words *drugs, medicines, products* and *product* were selected, however. When the collocational environments of the words *drugs, medicines, products* and *product* in the *NHSvax* corpus were examined, it was found that the use of these words was unremarkable. However, a cursory glance at the collocational environments of these terms in the *JABS* corpus indicates that they are frequently used in contexts in which a lack of trust in authority is suggested. These terms merit further investigation. Also potentially interesting for the same reasons are the words *mercury, thimerosal*[65] and *toxic*. The keywords selected from this category are:

> *vaccine, vaccines, mmr, drugs, medicines, products, product, thimerosal, mercury, toxic*.

The words included in the 'health, disease and medical conditions' category are shown in Table 6.3.

[65] The American name *thimerosal* is evidently preferred here over the British name *thiomersal*, which is a keyword in the *NHSvax* corpus.

| | | |
|---|---|---|
| **health, disease and medical conditions** | health and disease: general terms | *health, disease, diseases, infection, infections, illness, illnesses* |
| | health and disease: adjectives | *healthy, ill, sick* |
| | diseases and infections | *measles, virus, mumps, flu, cancer, polio, rubella, viruses, tetanus, meningitis* |
| | diseases and infections: non-evaluative pre-modifiers | *chronic, viral, infectious, genetic* |
| | disease types | *type* |
| | immune system and nervous system | *immune, immunity, (immune/nervous) system* |
| | immune system processes | *response* |
| | medical conditions: general terms | *disorders, syndrome, disorder, condition, conditions* |
| | specific conditions | *autism, autistic (spectrum), spectrum* |
| | conditions: misc. pre-modifiers | *neurological* |
| | diseases and conditions: verbs | *suffering (from), suffer* |
| | parts of the body | *brain, bowel, blood, cells, cell, body, heart* |
| | symptoms and signs | *symptoms, fever, pain* |
| | diet | *food* |

**Table 6.3** The *JABS* keywords in the 'health, disease and medical conditions' category

Once again, although most of the words listed here are identical to the keywords from the *NHSvax* corpus listed in the same category, there are fewer of them. In particular, there are fewer keywords which refer to specific diseases or which are used to pre-modify such terms. A similar set of *JABS* keywords were selected from this category as were selected from the *NHSvax* keywords belonging to this category. Thus, the words *immune* and *immunity* were selected, along with *disorders, syndrome, disorder, condition, conditions, autism, autistic, brain* and *bowel*. The adjectives *healthy, ill* and *sick* were also selected. The word *health* was rejected from the candidate list on same grounds for which it was rejected from the *NHSvax* candidate list: it mainly occurred in strings such as *department of health, public health,* and so on. One final word which has been grouped in this category and is potentially of interest is

*food*. This is worth examining because its use might shed light on notions of health. The words in this category which have been selected for closer examination, therefore, are:

> *immune, immunity, healthy, ill, sick, disorders, syndrome, disorder, condition, conditions, autism, autistic, brain, bowel, food.*

The keywords in the 'people, places and institutions' category are shown in Table 6.4.

| | | |
|---|---|---|
| **people, places and institutions** | nhs users/laypeople | *children, parents, child, son, infants, babies, daughter, kids, baby, parent, girls, patients, adults, families, patient, (the) public, population, people, individuals, mother, family, group, boy, women* |
| | pre-modifiers referring to people | *infant, human, older* |
| | healthcare practitioners and scientific and medical experts | *doctors, gp, doctor, researchers, scientists, experts, expert, officials* |
| | miscellaneous | *author* |
| | named publicly known individuals | *(dr andrew) wakefield, andrew (wakefield), david (kirby/ southall/ salisbury,/ thrower/ elliman)* |
| | JABS forum posters | *aasa,(john) stone, john (stone)* |
| | job titles and honorifics | *dr, professor, director* |
| | public institutions: general | *center, centers, department, institute, agency, government, authorities* |
| | public institutions: specific | *nhs* |
| | industries | *drug (companies), pharma, pharmaceutical (companies) manufacturers, companies* |
| | private companies | *merck* |
| | places | *hospital, university, school* |
| | geographical references | *usa, countries* |
| | nationalities | *american* |

**Table 6.4** The *JABS* keywords in the 'people, places and institutions' category

As one might expect, there are fewer words here than in the equivalent category in the *NHSvax* data which refer to healthcare professionals or public institutions. However, the terms *researchers, scientists, expert* and *experts* are interesting, given Taylor's (2010) findings. These terms were selected from this sub-category. Concordance evidence for some of the words in the sub-categories which refer to public and private institutions indicated that

they often occur in contexts where evaluation is implied and so might provide indication of the ways in which issues of trust are expressed. For example, the term *pharma*, a colloquial term used to refer to the pharmaceutical industry, occurs most frequently in the string *big pharma* and is used to express a negative evaluation. For example,

[6.1] … bad corporate science from **big pharma**.
[6.2] … corrupt officials who take bribes from **big pharma**.

The terms chosen are *government, nhs, pharma* and *pharmaceutical*. The terms selected from the category of 'people, places and institutions' are therefore as follows:

*scientists, expert, experts, nhs, government, pharma, pharmaceutical*.

The keywords listed in the 'science, medicine and healthcare' category are shown in Table 6.5.

| | | |
|---|---|---|
| **science, medicine and healthcare** | general medical, healthcare and scientific terms | *medical, medicine, diagnosed, clinical, scientific, science, diagnosis, treatment, alternative, care, treat, treated* |
| | terms related to administration of vaccination programme | *program, schedule, introduced, programme* |
| | references to age and time periods | *childhood, months, aged, age, years, birth, weeks, year, february, born, month, daily, days, september, november, today, times* |
| | references to epidemiological research and monitoring | *cases, deaths, epidemic, incidence, case, rates, levels, rate, spread* |
| | other technical terms | *factors* |
| | measurement | *percent* |
| | research processes | *conducted* |
| | miscellaneous | *et, al* |
| | tests and investigations | *trials, tested, testing, tests, trial, investigation confirmed, test, check* |

**Table 6.5** The *JABS* keywords in the 'science, medicine and healthcare' category

In the ‘science, medicine and healthcare’ category, there is, as one might expect, less evidence among the keywords of a preoccupation with administration of vaccine programmes, although the words *program, schedule, introduced* and *programme* occur. There is also a set of words which relate to epidemiological research and monitoring, for example, *cases, deaths, epidemic*, and so on, and there is a set of words related to tests and investigations, such as *trials, tested, tests*, and so on. None of the words in these sub-categories is likely to prove particularly fruitful for fulfilling the research objectives. It is interesting to note the substantial sub-category of words which indicate references to scientific research, such as *study, research, studies*, and so on, but, as was argued in the previous chapter, analysis of the verbs with which they collocate is likely to prove more productive. Some of the words which have been placed in the ‘general medical, healthcare and scientific terms’ sub-category are potentially interesting, though. Terms such as *diagnosed, care, treat* and *treated* are unlikely to prove fruitful and *medical, medicine, scientific* and *clinical* are rejected on the grounds that they collocate with other terms of interest. *Science,* on the other hand, is worth closer investigation, in view of Taylor’s (2010) findings. The word *alternative*, which is a keyword in *JABS* but not in *NHSvax*, is also potentially interesting. Many of the 1,042 citations of *alternative* refer to alternative medicine or alternative healthcare. However, the significance of this term in the data is skewed by 418 occurrences of the string *aged survivor of many years of alternative health care...and one vaccine, administered by a doctor without the consent of my parents, 50 years ago*, which occurs at the end of all the posts posted by one particular forum participant. Only the word *science* was therefore selected from this category.

The categories labelled ‘sources’ and ‘the concerns and business of JABS’ are very small categories. No words were selected from these categories. The first category comprises words

such as *news, media, press, source, website* and *site*. It also contains a small set of words which refer to specific sources of information. These are *jabs*, *cdc* (Centers for Disease Control and Prevention), *fda* (Food and Drug Administration), *whale* (a medicine-critical and vaccine-critical website) and *bbc.* It is interesting to note that references to these sources are frequent enough for them to be keywords in the corpus. Leaving aside *jabs*, the frequencies of the terms are as follows: *cda* <1,662> ; *fda* <1,127>; *whale* <893>; *bbc* <666>. They also represent different types of sources. The CDC and FDA are both American public institutions. The BBC is a mass media source. Whale, on the other hand, is a vaccine-critical internet source. Although these terms are potentially interesting from the point of view of examining attribution, they were not selected, on the grounds that examining reporting verbs was likely to prove more fruitful. The 'concerns and business of JABS' category comprises words related to requesting and offering information and advice, for example *info, information, advice, help, question, answer* and *support,* and to political action and legal issues, namely *campaign, hearing* and *legal*. These terms were not considered potentially useful.

The categories labelled 'likelihood, causation and temporal relations', 'expressions with modal meaning' and 'conditionality', on the other hand, yielded several terms. These are shown in Table 6.6 below. All items in this category were selected, with the exception of *current* and *currently*, which are less likely than the other terms in the 'temporal' sub-category to indicate causal relations, and *exposure* and *exposed*, which collocate with *thimerosal* and *mercury*.

| | | |
|---|---|---|
| **likelihood, causation, and temporal relations** | possibility | *risk, risks, possibility* |
| | cause and effect | *cause, link, caused , causes, linked, causing, due (to), links, affected, association, connection, because, lead (to), result (in)* |
| | adverbs and prepositions: temporal | *following, after, prior, recently, recent, currently, previously, since, current* |
| | reason | *reason* |
| | ergative verbs indicating change | *develop, developed, developing, , increased, increase, develop* |
| | effects, danger and harm | *adverse (events/effects/reactions), reactions, damage, effects, (vaccine) damaged, death, complications, harm, dangerous, injury, (adverse) events, side (effects), effect* |
| | sources of harm | *exposure (to mercury/thimerosal), exposed (to)* |
| **expressions with modal meaning** | epistemic modality | *'ll, potentially, can, potential, likely, possible, cannot, possibly, maybe, likely* |
| | necessity/desirability | *recommended, should* |
| **conditionality** | | *wouldn', if, whether, would* |

**Table 6.6** The *JABS* keywords in the categories 'likelihood, causation and temporal relations', 'expressions with modal meaning' and 'conditionality'

There were found to be many more keywords in *JABS* than in *NHSvax* which could be categorised in the 'expressions of evaluation' category (see Table 6.7). It was decided not to include all these terms as candidate keywords. Terms belonging to the sub-category 'terms used to express evaluation of vaccines' were selected, since they proved of central importance in the discussion of causation presented in Chapter Five. The terms which can be used as 'status nouns'[66] were also selected on account of the role they play, when they are followed by a projecting *that*-clause, in creating cohesion at the same time as expressing an evaluation of the proposition expressed in the *that*-clause. Other expressions of evaluation were not selected on the grounds that the use of such terms comes to light when examining many other terms.

[66] Admittedly, some in this group can also be used as verbs.

| | | |
|---|---|---|
| **expressions of evaluation** | terms used to express evaluation of vaccines | *safety, safe, protect, protection, prevention, prevent, effective, benefits* |
| | terms used to express evaluation of science/authority | *lack (of evidence/ information/ understanding/ knowledge), failed* |
| | evaluative terms: disease and reactions | *serious, severe, rare* |
| | evaluation of quality of information | *accurate, latest* |
| | evaluation: clarity | *apparently, clearly* |
| | other evaluative terms | *wrong, interesting, bad, unfortunately, sure, interested, normal, properly, seriously, significant, worse, correct , concerned, proper, higher, low, high, real, positive, responsible, benefit* |
| | hedging and boosting | *actually, extremely, completely, just, huge, highly, very, even, ever* |
| | status nouns: the 'sign' group | *evidence, proof* |
| | status nouns: factivity | *truth, problems, facts, issue, opinion, fact, problem, issues* |
| | status nouns: affect | *concerns, fear, concern, hope* |
| | comparison/contrast | *compared, similar, other, or, same* |
| | neutrally connotative terms used evaluatively in some contexts | *natural* |

**Table 6.7** The *JABS* keywords in the 'expressions of evaluation' category

The terms selected from this category are therefore:

> *safety, safe, protect, protection, prevention, prevent, effective, benefits, evidence, proof, truth, problems, facts, issue, opinion, fact, problem, issues concerns, fear, concern, hope.*

The category 'other nouns, verbs and phrases referring to events, actions and states of affairs' also yielded a number of useful terms. This category is shown in Table 6.8 below.

| | | |
|---|---|---|
| **other nouns, verbs, and phrases referring to events, actions, and states of affairs** | verbal processes | *reported, comments, reporting, quote, comment, claims, stated, according (to), states, mention, debate, suggest, statement, claim, concluded, agree, explain, saying, mentioned, say, told, asking, suggests, ask* |
| | recounting | *story, stories* |
| | written communication | *published, read, wrote, reading* |
| | making decisions | *informed (consent/choice/decision),(informed) consent, choice, decision* |
| | perception and cognition | *believe, wonder, known, note, trust, understand, know, noted, knows, (BE) aware, considered, hear, consider, doubt, watch, believed* |
| | seeming | *seems, appears, seem* |
| | showing | *prove, showed, revealed, showing, shows* |
| | discovering | *found, find, finding, identified, discovered* |
| | occurring | *occur, occurred, happened, happen* |
| | death | *died, kills, die* |
| | durative | *continue* |
| | adding | *add, added* |
| | other material processes | *try, removed, release, using* |
| | existential/relational processes or states | *live, lives, contain, contains* |

**Table 6.8** The *JABS* keywords in the 'other nouns, verbs, and phrases referring to events, actions, and states of affairs' category

The largest sub-category in this category comprises words which refer to verbal processes and which are potentially productive terms because they can be used in attribution. The words in this category were selected. The next sub-category comprises the words *story* and *stories*. It was decided to select these words since they are often used to introduce accounts of personal experience or of other people's experiences. The words in the sub-categories of 'perception and cognition', 'seeming', 'showing' and 'discovering' were selected on the grounds that words of this kind can be used as reporting verbs and can indicate speaker attitude. The words in the 'occurring' sub-category were selected on the grounds that these too can encode evaluation. The final pair of words from this category to be selected are *die* and *died* from the 'death' set. The verb *kills* <543> was rejected on the grounds that its frequency in the corpus is skewed by the presence of 502 occurrences of the string *vaccinations cause chronic*

*diseases is as fully an accurate things to say as "smoking kills" warnings on cigarette packs*, which occurs at the end of every post posted by one of the forum participants. The other items were rejected. We now turn to the 'other grammatical items' category. This is shown in Table 6.9 below.

| | | |
|---|---|---|
| **other grammatical items** | negative polarity | *'t, don', doesn', didn', isn', don't, not, wasn', haven', none, never, no* |
| | *that* | *that* |
| | abbreviations | *re* |
| | determiners and pronouns | *my, anyone, i, this, your, any, many, our, these, us, everyone, they, someone, we, several, anything, several, every* |
| | spatial deixis | *here* |
| | prepositions and adverbs | *regarding, about, against, also, by, ago, including, from, despite, without, among, yet* |
| | WH words | *who, why, how,* |
| | delexicalised verbs/ auxiliaries / cupola | *'ve, have, am, is, are, has, does, being, having, been, do, done* |
| | numerals | *million* |
| | quantities and quantifiers | *thousands, numbers, millions, hundreds, number, majority* |

**Table 6.9** The *JABS* keywords in the 'other grammatical items' category

The first thing of note here is the relatively high number of pronouns and determiners and of expressions indicating negative polarity. While it is interesting to note the statistical significance of terms which indicate negative polarity, one can only make useful observations about the expressions of negative polarity if one also considers expressions of positive polarity too. The terms indicating negative polarity were not included in the list of candidate terms. The first person pronouns *I* and *we* and the determiners *my* and *our* are potentially useful ways of uncovering how people talk about their own experience. These were included in the list of candidate terms. *That* was also included among the candidate terms, owing to its usefulness in bringing to light terms used to express evaluations of status or to frame reported

speech which frequently occur in the vicinity of *that*, even though they may not be keywords themselves. It was also decided to include *it* and *what*, in order to bring to light examples of introductory-*it* constructions and cleft and pseudo-cleft constructions. These, like constructions involving the N-*that* pattern, contribute to the cohesion of an argument and act as presupposition triggers.

The complete list of *JABS* corpus candidate keywords is shown in Table 6.10 below. Since close examination of keywords serves to bring to light central themes which are of concern in the data, the discussion which follows does not describe the lexico-grammatical environment of each keyword separately but focuses on the key themes which emerge from the analysis of the keywords and presents a discussion of the recurrent lexico-grammatical patterns which are used to realize particular propositions. These themes are causation, risk, health, the immune system, ways of expressing one's warrant for expertise, and the evaluation of discursive objects and sources of authority. The current chapter continues with an examination of the ways in which notions of health, immunity and risk are expressed. Chapter Seven examines the ways in which people express their warrant for expertise and the ways in which they express their evaluation of discursive objects and sources of information.

| Semantic/functional category | Keywords |
|---|---|
| **vaccines and pharmaceutical products** | *vaccine, vaccines, mmr, drugs, medicines, products, product, thimerosal, mercury, toxic* |
| **health, disease and medical conditions** | *immune, immunity, healthy, ill, sick, disorders, syndrome, disorder, condition, conditions, autism, autistic, brain, bowel, food* |
| **people, places and institutions** | *scientists, expert, experts, nhs, government, pharma, pharmaceutical* |
| **science, medicine and healthcare** | *science* |
| **likelihood, causation and temporal relations** | *risk, risks, possibility, cause, link, caused , causes, linked, causing, due (to), links, affected, association, connection, because, lead (to), result (in), following, after, prior, recently, recent, previously, since, reason, develop, developed, developing, , increased, increase, develop, adverse (events/effects/reactions), reactions, damage, effects, (vaccine) damaged, death, complications, harm, dangerous, injury, (adverse) events, side (effects), effect* |
| **epistemic modality** | *'ll, potentially, can, potential, likely, possible, cannot, possibly, maybe, likely* |
| **deontic modality** | *recommended, should* |
| **conditionality** | *wouldn', if, whether, would* |
| **expressions of evaluation** | *safety, safe, protect, protection, prevention, prevent, effective, benefits, evidence, proof, truth, problems, facts, issue, opinion, fact, problem, issues, concerns, fear, concern, hope* |
| **other nouns, verbs and phrases referring to events, actions and states of affairs** | *reported, comments, reporting, quote, comment, claims, stated, according (to), states, mention, debate, suggest, statement, conclusion, claim, concluded, agree, explain, saying, mentioned, say, told, asking, suggests, ask, story, stories, believe, wonder, known, note, trust, understand, know, noted, knows, (BE) aware, considered, hear, consider, doubt, watch, believed, seems, appears, seem, prove, showed, revealed, showing, shows, found, find, finding, identified, discovered, occur, occurred, happened, happen, died, die* |
| **grammatical items** | *that, it, what* |

**Table 6.10** Keywords in the *JABS* corpus with at least 10% distribution across corpus texts, identified for closer investigation

## 6.3 Expressions related to the MMR-autism causal hypothesis in the *JABS* corpus

### 6.3.1 Summary of conclusions drawn in Chapter Five regarding expressions relating to the causal hypothesis in the *NHSvax* corpus

When expressions relating to the causal hypothesis in the *NHSvax* corpus were discussed in Chapter Five, it was found that, when the causal hypothesis is referred to, nominalizations are preferred over explicit markers of causation. Thus we found expressions such as *link between mmr and autism* and *association between mmr and autism* were highly frequent. The proposition is typically rebutted using expressions containing the word *evidence*. Thus, we saw expressions such as *there is no evidence of a link* and *no evidence to suggest a link*, and so on. Where the causal hypothesis is expressed using an explicit marker of causation, such as *CAUSE*, the proposition is typically attributed to another speaker or writer and often it is hedged. It was suggested in Chapter Five, that, in the case of discourse about the causal hypothesis in *NHSvax*, the tendency, which is typical of scientific discourse, to condense clausal meaning into nominalized forms, and which we see in expressions such as *no evidence of a (causal) link between mmr and autism*, reflects the uncertainty of the science relating to the causal hypothesis. It is reasonable to hypothesize that propositions relating to the MMR-autism hypothesis in the vaccine-critical *JABS* corpus data might be expressed with a greater degree of certainty about the truth of the causal relation than was seen in the *NHSvax* corpus.

### 6.3.2 Phraseological patterns in the *JABS* corpus related to the causal hypothesis

Appendices 9 and 10 show, respectively, the fifty most frequent three- to six-word clusters from concordances of *mmr* and *autism* from the *JABS* corpus, alongside the fifty most

frequent three- to six-word clusters from concordances of *mmr* and *autism* from the *NHSvax* corpus. As might be expected, in both corpora, the most frequent cluster related to *mmr* and *autism*, respectively, are *the mmr vaccine* and *mmr and autism*. As with the *NHSvax* corpus, there is a high frequency of clusters from the concordances for *mmr* and *autism*, from the *JABS* corpus, which contain the words *link* and/or *between*. The lexeme *CAUSE* occurs in relation to the concordance *autism* as part of the cluster *cause of autism* <126> in *JABS* and *causes of autism* <34> in *NHSvax*. This might suggest that there is greater concern in the *JABS* corpus data with identifying a single cause of autism.

To get a clearer picture of the most frequent phraseological patterns with *mmr* and *autism* in *JABS*, it is useful to examine closely the twenty most frequent meaningful three- to six-word clusters. These are shown in Table 6.11 below. Of course, not all the highly frequent clusters listed in Table 6.11 are relevant to the present discussion. Clusters such as *(mmr rip) edited by gus* and *page 1 of 1* occur merely because most of the data comes from an online discussion forum. There are also 128 occurrences of the string *age of autism* and 91 occurrences of *the autism epidemic*. These relate to the online publication 'Age of Autism: Daily Web Newspaper of the Autism Epidemic'.[67] The remaining clusters shown in Table 6.10 merit consideration, though. Understandably, many of the clusters which are related to *mmr* reflect a preoccupation with the vaccine itself and with its administration. Hence, we see *the mmr vaccine, the mmr jab, the mmr vaccination, dose of mmr*, and *the mmr booster*. The cluster *the mmr controversy* is doubtless frequent owing to the high proportion of news text in the corpus. Interestingly, clusters which relate to expressions of causation in Table 6.10, do not contain the word *CAUSE*, instead we find *link between mmr and autism*, and *mmr autism link*.

[67] http://www.ageofautism.com

The issue of safety is obviously of great concern in the corpus data, thus we find *safety of mmr* and *the safety of mmr*. It is surprising, though, to find 73 occurrences of *mmr is safe*. Only closer observation will reveal how instances of *mmr is safe* are framed.

| | **Clusters with *mmr* as node** | **freq.** | **Clusters with *autism* as node** | **freq.** |
|---|---|---|---|---|
| 1 | *the mmr vaccine* | 1005 | *children with autism* | 556 |
| 2 | *the mmr jab* | 404 | *mmr and autism* | 202 |
| 3 | *mmr and autism* | 200 | *vaccines and autism* | 183 |
| 4 | *the mmr vaccination* | 151 | *autism spectrum disorders* | 157 |
| 5 | *edited by gus* | 121 | *of children with autism* | 137 |
| 6 | *mmr rip edited by gus* | 120 | *diagnosed with autism* | 135 |
| 7 | *mumps and rubella* | 109 | *age of autism* | 128 |
| 8 | *measles mumps and rubella* | 102 | *cause of autism* | 126 |
| 9 | *link between mmr and autism* | 100 | *between mmr and autism* | 126 |
| 10 | *mmr is safe* | 73 | *autism spectrum disorder* | 116 |
| 11 | *vaccine and autism* | 68 | *vaccine and autism* | 116 |
| 12 | *safety of mmr* | 64 | *page 1 of 1* | 111 |
| 13 | *mmr vaccine and autism* | 64 | *cases of autism* | 105 |
| 14 | *the safety of mmr* | 60 | *link between mmr and autism* | 101 |
| 15 | *dose of mmr* | 58 | *in children with autism* | 101 |
| 16 | *the mmr booster* | 57 | *between vaccines and autism* | 95 |
| 17 | *page 1 of 1* | 56 | *with regressive autism* | 92 |
| 18 | *the mmr controversy* | 54 | *the mmr vaccine* | 92 |
| 19 | *mmr was introduced* | 54 | *the autism epidemic* | 91 |
| 20 | *mmr autism link* | 51 | *incidence of autism* | 84 |

**Table 6.11** Top twenty meaningful three- to six-word clusters in *JABS* corpus with *mmr* and *autism* as node words

Turning now to the clusters related to *autism*, the preoccupation of the JABS group with discussing the problems of children with autism and with autism in general is reflected in the high occurrence of the clusters *(of/in) children with autism, autism spectrum disorder(s), diagnosed with autism, cases of autism, with regressive autism,* and *incidence of autism*. Once again, we find references to the causal hypothesis, and indications are that it is expressed in much the same way as was seen in the *NHSvax* corpus. Therefore, we find *mmr and autism,*

*vaccines and autism, between mmr and autism, link between mmr and autism,* and *between vaccines and autism,* as well as a reference to *the mmr vaccine*. However, closer examination of concordance evidence reveals greater variety in the expressions used to realize the causal proposition in this corpus than in *NHSvax*.

Table 6.12 below shows the results of a manual search through concordances of *LINK, CAUSE, ASSOCIATE, RELATE,* and *CONNECT*, for references to the causal proposition. It was found that a total of 51.6% of references to the causal proposition involve the use of nominalization, realized through the expressions *link between, association between, relationship between, connection with,* and so on. A further 13.1% involve verb uses of *LINK, ASSOCIATE,* and so on. Examination of the contexts of use of these expressions shown in Table 6.12 reveals that they tend to occur overwhelmingly in articles, uploaded onto the site or pasted into posts. This suggests that the tendency towards nominalization, typical of scientific discourse and of the popularized scientific discourse observed in the *NHSvax* corpus, is a frequent feature of media and other popularized accounts online, when the MMR-autism causal proposition is discussed. These phraseological patterns are thus transferred from the scientific domain into popularizations, and from there they are imported wholesale into discussion forum posts.

| Expression | Freq. |
|---|---|
| ***link between (the) (...) MMR / vaccine(s)/ vaccination/ (triple) jab and*** *(...) autism (spectrum disorders)/autistic disorders/regression* | 386 |
| ***(the) MMR (vaccine) /vaccines/vaccination (...) (does)/ (does not) cause/ causes/ causing*** *(bowel disease and/ all/ your child's) autism/ (regressive) autistic spectrum disorder/ asd/ problems* | 240 |
| ***caused by*** *vaccines/ (the) MMR* | 113 |
| ***MMR-autism link/ vaccine-autism link*** | 101 |
| ***linked to*** *(an increased risk of/the rise of/late onset) autism* | 60 |
| ***link(s) to*** *(bowel disorder and) autism* | 58 |
| ***autism after*** *(...) MMR/jab/vaccination* | 54 |
| ***association between*** *MMR (vaccine/vaccination)/measles (virus) vaccine/measles, mumps, rubella vaccine (and) (regressive) autism* | 53 |
| ***linking*** *(the) MMR (vaccine/jab)/vaccinations/vaccines/immunizations/the measles (mumps and rubella) vaccine/a common children's vaccine/it* ***with/to/and*** *(bowel disease and) autism/the disorder* | 52 |
| ***MMR (vaccine)/ vaccines can/could/may/might cause autism*** | 50 |
| ***link between autism and (...)*** *MMR/ measles, mumps and rubella vaccine/ vaccines/ vaccination(s)* | 48 |
| ***link(s) with*** *(regressive/rising) autism* | 46 |
| ***link(s) with*** *(the) (controversial) MMR (vaccine)/vaccine/triple jab* | 37 |
| ***linked to*** *(the)(...) MMR (vaccine/jab)/vaccines/vaccination* | 35 |
| ***autism following*** *(...) MMR/ vaccination/immunisation/jabs* | 34 |
| ***MMR (...) the/a (...) cause of*** *(...) autism/ bowel damage/problems* | 34 |
| ***connection between*** *(the) MMR (vaccine/jab)/ vaccine/vaccination/vaccines and (bowel disease and/later degeneration into) autism* | 32 |
| ***relationship between (the) MMR (...)/ vaccines/ vaccination and*** *autism* | 28 |
| ***associated with*** *(...) autism* | 25 |
| ***MMR (jab)/vaccines (has/had) caused*** *autism* | 25 |
| ***MMR (...)/ vaccines/ vaccination (...) lead(s)/leading to*** *autism* | 22 |
| ***associated with*** *(the) MMR* | 16 |
| ***connection between*** *(bowel disease and)/(their child's) autism and MMR/(vaccine strain) measles virus/ vaccines/vaccinations* | 15 |
| ***linkage/link to*** *MMR/the jab/vaccination/vaccines* | 15 |
| ***linking*** *(bowel disease or) autism/their regression/it with/to/and (the/their) MMR (vaccine)/ vaccine* | 11 |
| ***related to (...)*** *MMR/ vaccine(s)* | 11 |
| ***association between*** *autism/ASD/their child's illness/pervasive developmental disorder/inflammatory bowel disease and (the) MMR (vaccine)/vaccines* | 10 |
| ***connection with*** *the MMR vaccine/the jab* | 8 |
| ***MMR (...)/ vaccination(s)/ vaccines as (a/the) (...) cause of*** *autism* | 7 |
| ***MMR-autism connection*** | 6 |
| ***MMR (...) / vaccines/ vaccine related to*** *(...) autism* | 6 |
| ***association with*** *MMR* | 5 |

**Table 6.12** Realizations of the causal proposition in the *JABS* corpus, with a frequency of five or more occurrences

### 6.3.3 Recontextualization and modification of expressions of causation with nominalization

It would be inaccurate to give the impression that propositions expressed using phrases such as *link between, association between*, and so on, undergo no process of accommodation as they are recontextualized in popularizations and media texts. On the contrary, if we compare the use of such terms in the *NHSvax* and *JABS* corpora, as well as finding a number of similarities, we can see some significant differences. Of course, references in the *NHSvax* corpus to the causal hypothesis occurred in contexts of rebuttals of the proposition. We saw that the phraseological pattern most frequently used to rebut the hypothesis was *(FIND) no (evidence) of a link/association between mmr and autism*. We find a lot of evidence of these patterns in the *JABS* corpus, for example:

> [6.3] This time, researchers **found no link between autism and the measles-mumps-rubella vaccine.**
>
> [6.4] ... a series of epidemiological studies **has failed to find any evidence of** a **link between MMR and autism**.
>
> [6.5] A study by the respected Cochrane Library (October 2005) has said, on the basis of 31 pieces of research into the possible side effects of MMR, that it **found no association between MMR, autism, Crohn's disease and long-term disability**.

However, examination of extended context reveals that assertions such as these tend to come from news articles, or texts from similar online sources. Assertions that there is no evidence of a link are usually attributed. Examination of extended context, however, often reveals interesting patterns of framing. Examples 6.3 and 6.4, for example, both come from articles included in posts by vaccine-critical posters. The title of the article from which Example 6.5 is taken is ‘Expert sees no link between vaccines and autism’. The article is about a recently published book which counters vaccine-critical arguments. The article is unbalanced,

devoting most of its space to the book's author and little space to vaccine-critical voices. It also represents vaccine-critical groups in a less than flattering light: vaccine-critical individuals are reported to have insulted the author. This is illustrated in the opening paragraph of the article, which is shown in Example 6.6.

> [6.6] **They liken him to a prostitute. Someone with blood on his hands, who doesn't care about the health of children.**
> **Those are among the insults that Paul Offit gets by e-mail each week at Children's Hospital of Philadelphia.**
> He should probably expect to start getting a lot more.
> Offit, 57, has been defending the safety of vaccines for years, in response to beliefs that they are tied to autism-related disorders. He continues in the same vein with his new book - Autism's False Prophets: Bad Science, Risky Medicine, and the Search for a Cure - which is already generating heat.

Example 6.6 comes from an article about Andrew Wakefield's upcoming hearing with the General Medical Council. This article shows bias towards pro-vaccination arguments, as the wider context, shown in Example 6.7, demonstrates. The fact that several authoritative studies have failed to find evidence to support the hypothesis, has, we are told, 'failed to satisfy parents'. The implication is that the parents are unreasonable. This impression is reinforced through use of the verb *refused*. Wakefield is described as 'unrepentant about his research'. The fact that many parents mistrust the authorities is given a negative evaluation by being afforded the status of a 'puzzle'. Andrew Wakefield is also evaluated in negative terms, since he is reported as being 'unrepentant about his research'. Finally, the reporting verb *claim*, which signals authorial distance, is used to frame the assertion, attributed to his supporters, that many studies confirm his findings. The extended context is shown in Example 6.7:

> [6.7] Since The Lancet printed his paper in February 1998, a series of epidemiological studies has failed to find any evidence of a link between MMR and autism. **This has failed to satisfy parents, predominantly middle-class and well-educated, who have refused to allow their children to be immunised or demanded single vaccines, as Dr Wakefield recommended.**
> **One of the greatest puzzles of the saga** is what has sustained this level of mistrust in the

medical authorities. Unlike most scientific controversies which flare up and die away, this one has simmered for a decade and is now to be fired up again by the GMC case. **Dr Wakefield is said to be unrepentant about his research** and remains convinced that some children are vulnerable to damage by the MMR vaccine. **His supporters claim** that independent studies from different countries have confirmed his findings but clinical investigation of individual cases which could provide the proof has been blocked in the UK.

Articles such as those in which Examples 6.3 (and 6.6) and 6.4 (and 6.7) occur are selected by forum participants not because they, the posters, (necessarily) endorse the message encoded in the text, but in order to draw other forum users' attention to an article they consider of interest. Examination of the extended context of Example 6.5 revealed that the article in which this example occurs adopts a vaccine-critical stance, though. Example 6.8 comes from a text authored by the JABS administrators which criticizes the government's view that scientific evidence offers decisive proof that there is no link between MMR and autism, bowel disease, and long-term disability.

[6.8] A study by the respected Cochrane Library (October 2005) has said, on the basis of 31 pieces of research into the possible side effects of MMR, that it found no association between MMR, autism, Crohn's disease and long-term disability. **The Department of Health is hailing it as another 'final nail' in the MMR controversy but there is another side to this that they have missed.**

What we see in the first sentence in Example 6.8 is a summary of the findings of the Cochrane Review. By framing the reported section with the verb *has said*, the author neither endorses nor casts doubt on the veracity of the proposition that the Cochrane Review found no evidence of a connection. In fact, the author expresses a positive evaluation of the Cochrane Review through use of the pre-modifier *respected*. However, the response of the Department of Health to these findings is evaluated negatively. The juxtaposition of the quasi-scientific register used in the first sentence and the literary/journalistic register suggested by the

expression *hailing it as another final nail* flouts the maxim of manner and thus suggests irony. Similarly, the quotation marks around the expression *final nail* signal authorial distance from the claim, on the part of the DoH, that the findings of the study vindicate the view that there is no causal link between MMR, autism, and neurological disorders. The author also counters the argument put forward by the DoH but stating that they have overlooked a crucial aspect to the issue.

We find similar expressions of authorial distance in uses of *link between* in news texts, when reference to the causal hypothesis is made and it is not rebutted. Again, the proposition is typically attributed, and expressions of commitment to the truth of the proposition are often hedged. *Link* is frequently pre-modified with *possible*. There are 125 occurrences of *possible link*. The proposition is also often modified with forms of the lexeme *SUGGEST*. In the examples, below, attributed sources are highlighted with underlining.

> [6.9] Some of the 12 children whose medical history featured in the controversial 1998 Lancet paper, drawn up by Dr Wakefield and his colleagues and which **suggested a possible link between the jab and bowel disease and regressive autism**, had received the Urabe-strain vaccine
>
> [6.10] Dr Wakefield said: "The study has identified **a possible link between gut disorders in children and autism.**
>
> [6.11] Following the publication of an earlier 1995 paper on **a possible link between a measles vaccine and inflammatory bowel disease**, Dr Wakefield said he received several calls from parents.
>
> [6. 12] The purpose of the Lancet study was to **show a link between MMR and autism**.
>
> [6.13] Dr Andrew Wakefield, the doctor who first **suggested a link between the MMR vaccine and autism** ...
>
> [6.14] Confidence in the mumps, measles and rubella vaccine fell after researchers published a 1998 paper in The Lancet medical journal **suggesting a link between MMR and autism.**

The use of attribution evident in the news reports exemplified in this section is typical of media reporting. Media reports, especially hard news reports, rely heavily on sources. The reliance on scientific and DoH sources might account for the similarities observed here between the *JABS* and *NHSvax* corpora. Of course, as we have seen, claims relating to the causal hypothesis expressed using a nominalized form are more likely to be hedged in the *JABS* corpus than in the *NHSvax* corpus. Since they tend to occur in news texts, they are typically framed in ways that suggest authorial distance.

### 6.3.4 Expressions of causal relations using explicit markers of causation

As was in the discussion in Section 6.3.1 above, propositions relating to the causal hypothesis are sometimes made using explicit markers of causation. In fact, in 28.4% of instances relating to the causal proposition, forms of the lemma *CAUSE* are used. In 3% of these cases, *CAUSE* is used as a verb and is hedged, typically by *can*. Instances such as these are more frequent in news texts than elsewhere. Examples are:

> [6.15] The group's draft report recommends no change in the immunisation policy despite growing fears that **the MMR triple vaccine can cause autism** in some children.

> [6.16] Theory one - that **MMR vaccines and thimerosal containing vaccines can combine to cause autism -** Theory two - that **thimerosal containing vaccines alone can cause autism -** Theory three - **MMR vaccines alone can cause autism**

In Examples 6.15 and 6.16, authorial distance is also signalled through attribution. However, in both examples, the source of the attribution is generalized, not specific. In Example 6.15, it is attributed to the fears of unspecified people. In Example 6.16, it is attributed to an unidentified source by labelling it a theory.

In texts which take a clear vaccine-critical stance, claims that MMR does not cause autism are, obviously, always attributed. The use of reported speech allows the writer to reframe the proposition and, if they are so minded, to alter the illocutionary force of the original proposition. Indirect speech is often used and, in such instances, writers tend to express the causal hypothesis using unhedged forms of the lexeme *CAUSE*. We see this in Examples 6.17 and 6.18 below. It is possible that the sources expressed their claims in this way. Alternatively, they may have hedged their claims or used the expression *no evidence of a link*, or something similar.

> [6.17] Julie Gerberding, MD, the CDC director, also supported her staff and the 2004 IOM Committee report: the MMR vaccine **did not cause autism ever**. Period.

> [6.18] Accordingly, in every case under the Vaccine Act, DVIC has maintained the position that **vaccines do not cause autism**, and has never concluded in any case that autism was caused by vaccination.

Individual posters on the forum or message board make stronger or weaker claims depending on the identity they wish to project. The following post comes from a message board post by an infrequent poster, who declares some knowledge of the subject: as well as the prior knowledge she presumably has, she has watched a *Panorama* report. However, she concedes that her expertise is limited since she describes herself as 'confused'. She asks if the single vaccine *may* cause autism:

> [6.19] Im getting really confused. is it believed that the risk of autism is only from the MMR or if as Dr Wakefield suggested in Panorama that measles virus has been found in the bowel does this mean that **the single vaccine may also cause autism**?

It is the frequent posters, those who present themselves as having sound knowledge and expertise, who are more likely to make stronger claims. Take the following example:

> [6.20] It isn't just mercury, **MMR kills and causes autism**, yet doesn't have any mercury in it.

However, it is predominantly the JABS forum participants who are not vaccine-critical but whose stance is pro-science[68] who make strong claims in this way:

> [6.21] **MMR does not cause autism**. The claim that it does has been studied in great detail and it is simply a hypothesis that has been proven to be false.
>
> [6.22] ... put simply: **mmr does not cause autism**

Vaccine-critical posters are more likely to express themselves in ways which indicate less certainty. However, they sometimes use the uncertainty of the science strategically. We see this in the three examples below, which exploit the fact that scientific studies rebut the hypothesis by stating that there is no evidence of a link. The poster in Example 6.23 states that the studies have failed to draw a definitive conclusion, while the posters in Examples 6.24 and 6.25 claim that there is a lack of evidence to show that there is no causal link.

> [6.23] The studies that the pharmaceuticals companies and various bodies on the MMR **have never come to a proper conclusion regarding whether the MMR causes Autism or not.**
>
> [6.24] There **may not be evidence to show that Mmr definetly causes autism**, but there is also no research to show that it definetly does not.
>
> [6.25] Firstly **there is no scientific proof that the MMR causes Autism** as there is also no scientific proof it doesn't.

Having examined expressions of causation, we now turn to expressions related to safety.

[68] As explained in Chapter Four, a small number of people who espoused pro-science views joined JABS and posted on the forum in order to counter the vaccine-critical views expressed on the JABS site.

### 6.3.5 MMR and safety

There are 71 occurrences of the string *mmr is safe*. As might be predicted, in the majority of these occurrences the proposition that MMR is safe is framed in such a way as to express a negative evaluation. The assertion is usually attributed to another speaker. In Example 6.26, it is attributed to Tony Blair. We see authorial distance and a negative evaluation expressed through the use of the reporting verb *boasted*.

> [6.26] Mr Blair declared the matter settled back in October 2002, when **his official spokesman boasted** at a Downing Street media briefing: **'We are winning the argument that MMR is safe.'**

In Example 6.27, which comes from a message board post, the author distances him- or herself from the claim that MMR is safe, at the same time as expressing mistrust of the medical profession and the government, which are represented as conspiring together to mislead the public.

> [6.27] I do not trust this government and their spin and it is obvious that **the medical profession is just spinning the government line that MMR is safe** because when you try and ask your Doctor questions about possible side effects of MMR they dont have any answers.

Example 6.28, on the other hand, is one of the few examples of use of *mmr is safe* by a pro-science poster. The assertion that MMR is safe is intensified through the expression *in the face of overwhelming evidence*.

> [6.28] **In the face of overwhelming evidence that MMR is safe**. That Wakefield was wrong. Is it not arrogant to continue to assert the obvious? Is it not arrogant to state, as one regular did in a recent post, 'At last the tide may be turning' and thereby assume intellectual superiority over the medical profession and the great British Public.

### 6.3.6 Summary of Section 6.3

It was hypothesized that the *JABS* data would show a far greater tendency among writers to express claims related to the causal proposition in ways which expressed a greater degree of certainty than was seen in the *NHSvax* corpus. This was found to be true only to a certain extent. It was found that the most frequent discourse patterns which occur in the *NHSvax* data are those which occur most frequently in the *JABS* data. As a result, there are numerous instances in *JABS* of *link between mmr and autism* and *no evidence of a link between mmr and autism*, and so on. However, occurrences such as these typically occur in the news articles and official documents which are represented in the *JABS* corpus data. The fact that forms of expression which are typical of mainstream medical-scientific discourse recur with high frequency in an online discussion forum, even if they are reproduced in the *JABS* data largely in the form of verbatim reproductions of media texts, undoubtedly results from two social factors working in concert: the fact that medical-scientific discourse is a dominant discourse (cf. page 68) and the fact that the mass media plays a central role in mediating knowledge in the public sphere. Media discourse about the MMR hypothesis was found to display other patterns of usage, though. Expressions of causation involving forms of *CAUSE* occur relatively more frequently in the *JABS* data than in *NHSvax*. These expressions also typically occur in news texts. Assertions of causation expressed using *CAUSE* are typically hedged with the modal *can*. However, when the voices of JABS forum and message board posters are examined, we find patterns which are in marked contrast to those found in the *NHSvax* data. Individual posters show a preference for expressing claims using *CAUSE* rather than using nominalization. Furthermore, low-frequency posters tend to hedge their claims while high-frequency posters are more likely to make claims which express a higher degree of certainty. This tendency is particularly marked among pro-science posters.

## 6.4 Expressions related to health and the immune system in the *JABS* corpus

### 6.4.1 Summary of conclusions drawn in Chapter Five regarding expressions related to health and the immune system in the *NHSvax* corpus

In Chapter Five, it was concluded that, in the *NHSvax* corpus, diseases are represented as potentially harmful to health and well-being, while vaccination is represented as offering protection. The frequent semantic sequences used to refer to the causes of disease and the effects of disease were found to be A BACTERIUM/VIRUS CAUSES A DISEASE, A DISEASE/VIRUS CAUSES A PHYSICAL REACTION, A DISEASE CAUSES DAMAGE and A DISEASE CAUSES DEATH. The terms *PROTECT* and *safest* were found to occur frequently in the environment of *immunisation* or *vaccine(s).* A strong collocate of the word *immunity* in the *NHSvax* corpus is *boost*, and expressions containing *boost* and *immunity* were found to be used in contexts which express deontic meaning. It was also concluded in Chapter Five, that discourse about health and immunity in the *NHSvax* corpus is characterized by a notion of the immune system as an animate object which responds to particular stimuli. This is reflected in the collocates *system, response, body, systems, produce, disease, responds* and *bacterial.* The words *STIMULATE* and *TRIGGER* often occur in the environment of *immune response*. There is evidence, though, that the medical-scientific community acknowledges that immune system responses are not always effective. In particular, there were found to be a number of references to groups of people with compromised immune systems. The expression *weakened immune system* is frequently used to refer to groups of people for whom vaccination is risky. There were also found to be a number of citations of the expressions *overload the immune system* and *weaken the immune system*. These are used to refer to the belief, presumed to be

widespread among the public, that multiple vaccines compromise the individual's immune system. Another belief which is held to be widespread among the public, and which is rebutted in some of the *NHSvax* texts, is that *natural immunity*, that is, immunity acquired through contracting a disease, is better than vaccine-acquired active immunity or vaccine-acquired passive immunity. Finally, there is a concern evident in the *NHSvax* data with the notion of herd immunity. It was found that the term *herd immunity* rarely occurs in an environment where an evaluative stance is expressed, but, where this does happen, a positive evaluation is expressed.

### 6.4.2 Vaccination, disease, and causation

When the keywords in the *JABS* corpus were examined, it was noticed that there are far fewer keywords here than among the *NHSvax* corpus keywords which refer to diseases. Obviously, *measles, mumps*, and *rubella* are statistically highly significant (ranking at positions 11, 31, and 39, respectively, in the JABS keyword list), and *virus* (16), *disease* (24), *diseases* (38) and *viruses* (41) are also significant (see Appendix 8). One particular keyword which one might not expect to find ranking so high up the keyword list is *cancer* (35). Since JABS group members are preoccupied mainly with the harm vaccines can potentially cause, it is not surprising that, when the *JABS* corpus is compared with *NHSvax*, there is a relatively high occurrence of the semantic sequence A VACCINE CAUSES DAMAGE, and relatively few occurrences, of the semantic sequence A BACTERIUM OR VIRUS CAUSES A DISEASE. The most frequent realization of the semantic sequence A VACCINE CAUSES DAMAGE is (a vaccine) *(can/could) CAUSE (permanent) brain damage*.[69] Concordance 6.1 below shows

[69] Instances of CAUSE (permanent) brain damage are hedged with *can/could* in approximately 33% of occurrences. For this reason it was decided to label the semantic sequence A VACCINE CAUSES DAMAGE rather than A VACCINE CAN/COULD CAUSE DAMAGE.

lines for *CAUSE + brain damage*. As is evident from the concordance lines, the proposition that vaccines cause brain damage is modified in about half the instances.

| N | Concordance |
|---|---|
| 1 | as the normal vaccinations. Be wary of the DTP though, it caused fits and brain damage to one of my children, leaving him with severe learning disabilitie |
| 2 | the live measles virus in the MMR triple vaccine may be causÂ-ing gut and brain damage. In an article published in the Lancet medical journal in 1998, Dr |
| 3 | the live measles virus in the MMR triple vaccine may be causÂ-ing gut and brain damage. In an article published in the Lancet medical journal in 1998, Dr |
| 4 | st disagreeing. So are these so called sceptics denying vaccines can cause brain damage and other health problems? </thread> □□ □ |
| 5 | US Centers for Disease Control accept inprinciple that vaccines can cause brain damage? (Q2) Do these bodies accept that parents< reports of children< |
| 6 | to get her son vaccinated and was concerned as she had read it could cause brain damage and convulsions and asked him was that possible. "He said the |
| 7 | following concerns that the whooping cough (pertussis) vaccine could cause brain damage (the vaccine has since been changed). However, compensation |
| 8 | that MMR or any other vaccine causes autism. But vaccines do cause brain damage so perhaps our children have been wrongly labled 'autistic' |
| 9 | sent in most childhood vaccines. He says: "It's highly toxic, is known to cause brain damage and has been implicated in behavioural problems in children." On |
| 10 | showing the court how vaccines are routinely used in the laboratory to cause brain damage in rats (I am appending a couple of examples). I may also refer |
| 11 | showing the court how vaccines are routinely used in the laboratory to cause brain damage in rats (I am appending a couple of examples). I may also refer |
| 12 | vaccination that most people don't know but should know: 1. Vaccines cause brain damage - and can kill. Since the 1970s the British Government has very |
| 13 | brain damage. And remember no 1 on my list of facts: vaccines cause brain damage. Now, here's the killer fact: when a research project was set up |
| 14 | is to stop the spread of disease. Whooping cough vaccine will cause brain damage in 1/600000 children. ie one child every year in the UK. Choose |
| 15 | they said vaccines were too dangerous: the whooping cough vaccine caused brain damage, the chickenpox vaccine caused paralysis, and the Hib vaccine |
| 16 | I know , because I have personal experience . Bottom line :- Mercury causes brain damage , mental illness and of course Autism. Make your own minds up |
| 17 | strong believers in homoeopathy), for an action claiming that MMR causes brain damage, epilepsy, arthritis, and autoimmune disease." |
| 18 | measles component and/or the combined dosage of mercury which causes brain damage so there is a risk with either choice. It's easy for parents to ge |
| 19 | from scientists worldwide that the measles, mumps and rubella jab is causing brain damage in certain children. But he added: "There are very powerful people |
| 20 | umulation of evidence' from scientists worldwide that the triple jab is causing brain damage in certain children. Questioning the government's stance has |
| 21 | from scientists worldwide that the measles, mumps and rubella jab is causing brain damage in certain children. "But he added: "There are very powerful peopl |
| 22 | from scientists worldwide that the measles, mumps and rubella jab is causing brain damage in certain children. He called the sudden termination of legal aid |
| 23 | children and pregnant women. Thimerosal in vaccines is suspected of causing brain damage and weakening the immune system, making some children |
| 24 | scientists around the world was showing that the MMR vaccine was causing brain damage in certain children. He furthermore made it clear why so many |
| 25 | alled it Black Death. The other vaccines cause their own reactions ? paralysis, brain damage, lockjaw, etc. When doctors had tried to suppress the symptoms |
| 26 | rubella) vaccine can cause brain inflammation and permanent brain damage. There have been more than 46,000 reports of health problems |
| 27 | if they have to go back a century. "Measles vaccine can cause permanent brain damage and SSPE (rare but fatal brain disease), though almost certainly |
| 28 | vaccinated, writes Dr Halvorsen. "Measles 4vaccine can cause permanent brain damage and SSPE (rare but fatal brain disease), though almost certainly |
| 29 | claiming that the whooping cough (pertussis) vaccine caused permanent brain damage. Wilson reported the stories of 22 children who suffered from epil |
| 30 | the jab's pertussis component has been implicated as a cause of permanent brain damage and death. Details of the risk had been outlined in an |
| 31 | by vaccines? The fact that vaccines can cause temporary or permanent brain damage is an established fact. Even the manufacturers admit it. For |
| 32 | by vaccines? The fact that vaccines can cause temporary or permanent brain damage is an established fact. Even the manufacturers admit it. For |
| 33 | that measles vaccination can cause serious disorders leading to permanent brain damage. Dr Halvorsen also thinks vaccinating against mumps has done |
| 34 | that measles vaccination can cause serious disorders leading to permanent brain damage. Dr Halvorsen also thinks vaccinating against mumps has done |

**Concordance 6.1** Lines for *CAUSE (permanent) brain damage* in the *JABS* corpus

In twelve of the thirty-four examples above, an epistemic modal is used to hedge the claim. There are two instances of *may be causing gut and brain damage* (Lines 2 and 3), two of *could cause brain damage* (Lines 6 and 7), six of *can cause (permanent/ temporary or permanent) brain damage* (Lines4, 5, 27, 28, 31, 32), and two of *can cause serious disorders leading to brain permanent damage* (Lines 33 and 34). In one instance, commitment to the truth is attenuated with the phrase *suspected of*, as in *thimerosal in vaccines is suspected of causing brain damage* (Line 23) and in another case (Line 30), the term *implicated* is used:

*the jab's pertussis component has been implicated as a cause of permanent brain damage*. Elsewhere, strong commitment to truth is expressed, for example, *(the) jab is causing brain damage* (Lines 19-22), *vaccines cause brain damage* (Lines 12 and 13), *vaccines do cause brain damage* ( Line 8), and *whooping cough vaccine will cause brain damage* (Line 14), and so on. This is in stark contrast to the findings in the *NHSvax* corpus, where terms such as *adverse events* were preferred. While *CAUSE* constitutes an 'explicit causative' (Xuelan and Kennedy, 1992), *adverse events* corresponds to Xuelan and Kennedy's (ibid.) category of 'ambiguous causatives'. The terms *adverse* and *events*, considered separately, do not belong to the set of explicit causatives, but the expression *adverse events* has a semantic prosody of causation. The most frequent collocates which occur 1 word to the right of *adverse events* in the *NHSvax* corpus are *following, causal, after, see* and *associated*. Four of these terms are used to express causation either explicitly or implicitly. Use of the term *adverse events* might therefore be considered euphemistic. Use of *CAUSE* is clearly not.

Turning now to A BACTERIUM OR VIRUS CAUSES A DISEASE , the vast majority of occurrences of this semantic sequence concern the relationship between the human papilloma virus (HPV) and cervical cancer. There are several references in the JABS corpus to the causal relationship between the virus and cervical cancer. As we see in Table 6.13, *cancer* occurs among the top twenty collocates of *cause, causes*, and *causing*, and *cervical* is one of the top collocates of *causes*. The reason for the relatively high number of references to cervical cancer is that, in 2008, the UK government decided to extend its immunization programme to include vaccination for teenage girls against HPV. A lot of discussion ensued on the JABS discussion forum about the safety of the HPV vaccines. In most instances in

which *CAUSE* collocates with *cervical cancer*, it is presupposed that HPV causes cervical cancer.

| | ***cause*** | **MI3** | ***causes*** | **MI3** | ***caused*** | **MI3** | ***causing*** | **MI3** |
|---|---|---|---|---|---|---|---|---|
| 1 | *cause* | 33.91 | *causes* | 32.28 | *caused* | 32.98 | *causing* | 31.64 |
| 2 | *chronic* | 26.70 | *autism* | 22.24 | *by* | 26.87 | *the* | 20.34 |
| 3 | *fully* | 26.05 | *the* | 21.92 | *the* | 23.71 | *damage* | 19.17 |
| 4 | *diseases* | 25.97 | *that* | 21.46 | *that* | 22.31 | *of* | 18.95 |
| 5 | *vaccinations* | 25.52 | *of* | 21.38 | *damage* | 21.62 | *in* | 18.89 |
| 6 | *the* | 24.74 | *and* | 20.31 | *autism* | 21.38 | *and* | 18.60 |
| 7 | *of* | 24.45 | *cancer* | 20.13 | *vaccine* | 20.80 | *to* | 18.45 |
| 8 | *can* | 23.91 | *virus* | 19.34 | *vaccines* | 20.79 | *is* | 18.45 |
| 9 | *as* | 23.80 | *to* | 18.77 | *of* | 20.44 | *disease* | 17.88 |
| 10 | *autism* | 23.67 | *cervical* | 18.65 | *believed* | 20.18 | *autism* | 17.86 |
| 11 | *that* | 23.33 | *a* | 18.39 | *and* | 20.03 | *brain* | 17.62 |
| 12 | *is* | 23.33 | *in* | 18.31 | *to* | 19.90 | *cancer* | 17.38 |
| 13 | *death* | 23.27 | *which* | 18.20 | *a* | 19.87 | *be* | 17.16 |
| 14 | *to* | 23.00 | *natural* | 18.09 | *severe* | 19.73 | *problems* | 17.16 |
| 15 | *and* | 22.14 | *disease* | 18.06 | *be* | 19.43 | *a* | 17.10 |
| 16 | *vaccines* | 22.10 | *death* | 17.49 | *was* | 19.19 | *are* | 17.01 |
| 17 | *a* | 21.84 | *syphilis* | 17.34 | *mmr* | 19.11 | *capable* | 16.92 |
| 18 | *cancer* | 21.60 | *it* | 17.31 | *in* | 19.11 | *vaccines* | 16.78 |
| 19 | *in* | 20.90 | *what* | 17.26 | *illness* | 19.07 | *harm* | 16.70 |
| 20 | *not* | 20.87 | *mmr* | 17.13 | *have* | 18.93 | *viruses* | 16.39 |

**Table 6.13** The twenty most statistically significant collocates of *cause*, *causes*, *caused* and *causing*, according to MI3

The presupposition that HPV causes cancer is particularly evident in the use of the expression *cancer-causing HPV*. This is illustrated in Concordance 6.2.

| N | Concordance |
|---|---|
| 3 | y 31 July 2008 Gardasil, the vaccine designed to protect against the cervical cancer-causing HPV virus, is responsible for the deaths of 10 adolescent girls a |
| 4 | hat the benefits of the Government's national vaccination programme against the cancer-causing HPV virus far outweigh the risks. But equally, it would be wrong |
| 5 | and fatigue since being administered three Cervarix jabs, against the cervical cancer-causing HPV virus, between September 2008 and March this year. She |
| 6 | 2, 2008 Deborah Kotz About a month after being vaccinated against the cervical cancer-causing HPV virus, 13-year-old Jenny Tetlock missed the lowest hurdle in |
| 7 | to 45 years old. In addition, the study will compare immune responses to other cancer-causing HPV types. "Vaccination to prevent cervical cancer may be the |
| 8 | has also shown â¬Sefficacyâ¬ cancer-causing HPV types. </thread> □□ □□</thread> |
| 9 | rdasil The vaccines are very similar, Both are designed to protect against the cancer causing HPV types 16 and 18, which cause about 70% of the cervical |
| 10 | . The vaccine â¬ called Cervarix â¬ offers protection against two strains of cancer-causing HPV. The drive to vaccinate girls will cost the Government Â£100 |
| 11 | s vaccine Cervarix provides significant protection against the four most common cancer-causing HPV strains for 6.4 years, according to data from a follow-up stu |
| 12 | shows the vaccine protects women against one of the most common types of cancer-causing HPV â¬ types 16 and 18 â¬ that are responsible for more than 70 |

**Concordance 6.2** Concordance lines for *cancer-causing HPV* in the *JABS* corpus

However, there are a number of instances in which the proposition is negotiated, as Concordance 6.3 shows. In line 262, for example, the proposition that HPV causes cancer is hedged through use of the expression *it is unclear if*. In lines 268 and 269, the poster says, *I find this debunking of the HPV causes cancer theory quite compelling.*

| N | Concordance |
|---|---|
| 261 | would be if used in males, said Markowitz, a medical epidemiologist. HPV causes at least 20,000 cases of cancer in the U.S. each year. Cervical |
| 262 | these women. A "quick fix" technology won't work. It's also unclear if HPV "causes" the cancer or, rather, is "associated" with it (e.g. already existing c |
| 263 | receiving the jab in England alone. The sexually-transmitted infection, HPV, causes most cases of cervical cancer - the second most common cancer in |
| 264 | Other ongoing trials for both drugs are looking at vaccinating men. HPV causes cancers of the head, neck, mouth and anus in men, but overall |
| 265 | Prize in Physiology or Medicine in 2008, discovered that oncogenic HPV causes cervical cancer.1-4 His discovery led to characterization of the |
| 266 | repeat the meme on a regular basis. And what is that meme? That HPV causes cervical cancer, and, therefore, HPV vaccinations could halt cervical |
| 267 | last year â¬ the highest figure since records began in 1995. The illness causes fever which can lead to serious complications. Cases of mumps have |
| 268 | 06/09/2008 12:32:45 I find this debunking of the HPV causes cancer theory quite compelling. The author is equally unimpressed |
| 269 | Originally posted by I find this debunking of the HPV causes cancer theory quite compelling. The author is equally unimpressed |

**Concordance 6.3** Concordance lines for *HPV causes (cervical) cancer* in the *JABS* corpus

Typical examples of *CAUSE + cervical cancer* are:

> [6.29] Gardasil protects women from **strains of the HPV (human papillomavirus) that cause 70 percent of all cervical cancer**. But numerous cases of young women being stricken with various potentially deadly complications have arisen all over the world.
>
> [6.30] Asked about Gardasil on Wednesday, health officials said the vaccine was safe and should continue to be administered to women ages 9 to 26. The vaccine is the first known preventive against **the human papillomavirus, which can cause cervical cancer**.
>
> [6.31] Cheryl Cave says allegations that she is abusing her 13-year-old daughter Ashleigh are an attempt to “shut her up" as she is demanding an investigation into the safety of a drug used to protect girls against **the HPV virus that causes cervical cancer**. Ashleigh collapsed shortly after being given the jab at school and has spent nearly a year in hospital.

References to the causal relationship between HPV and cervical cancer represent the majority of realizations of the semantic sequence A BACTERIUM OR VIRUS CAUSES A DISEASE. However, there are other, potentially more interesting realizations of this semantic sequence.

For example, there are a number of instances in which doubts are raised about the causal relationship between HIV and AIDS, as Concordance 6.4 shows.

| N | Concordance |
|---|---|
| 245 | one day prior to his submitting his most important paper asserting HIV causes AIDS. Exhibit Two: |
| 246 | 26 years after its discovery, no scientist has been able to explain how HIV causes AIDS. The latest theory put forward, the so-called "Viral Load" |
| 247 | and doctors with good university positions who are questioning if HIV causes AIDS (many coming under quite nasty attacks for doing so.) This |
| 248 | the key experiments on HIV - including the one cited today as proving HIV causes AIDS. I show that it proved nothing of the sort - the scientist who did |
| 249 | US National Cancer Institute (NCI) laboratory chief famed for proving HIV causes AIDS, for which he has been awarded 27 honorary degrees and the |
| 250 | advice to another parent. That remark stood. John Stone denied that HIV causes AIDS. That remark stood. Less hypocrisy please, more facts. jdc |
| 251 | advice to another parent. That remark stood. John Stone denied that HIV causes AIDS. That remark stood. Less hypocrisy please, more facts. jdc |
| 252 | go to http://www.virusmyth.net/aids/award.htm "If there is evidence that HIV causes AIDS, there should be scientific documents which either singly or |
| 253 | evaluated. The authors concluded that: it is abundantly evident that HIV causes disease and death in hemophiliacs, a group generally lacking |
| 254 | using molecular cloning. Therefore it is highly presumptuous that HIV causes AIDS. There is no mention of this in your website link Occam48 |
| 255 | es, have you Gus..? Kind regards, Becky Is a picture of HIV proof that HIV causes AIDs? Please provide the photographic evidence. No, but it's pretty |
| 256 | es, have you Gus..? Kind regards, Becky Is a picture of HIV proof that HIV causes AIDs? Please provide the photographic evidence. |
| 257 | rocked to discover that the key HIV research, that said to prove that HIV causes AIDS, was investigated for scientific fraud by very powerful US |
| 258 | rocked to discover that the key HIV research, that said to prove that HIV causes AIDS, was investigated for scientific fraud by very powerful US |
| 259 | viewpoint to the good Dr CT over his "agnosticism" about whether HIV causes AIDS? I'll throw your first comment back at you there seems to be a |

**Concordance 6.4** Concordance lines for *HIV causes AIDS* in the *JABS* corpus

Here there is a lot of evidence of negotiation of meaning. This is expressed through use of expressions such as *his most important paper asserting that HIV causes AIDS, doctors … are questioning if HIV causes AIDS, John Stone denied that HIV causes AIDS, if there is evidence that HIV causes AIDS, it is abundantly evident that HIV causes diseases, the key HIV research that said to prove that HIV causes AIDS,* and *his agnosticism about whether HIV causes AIDS*. Assertions that HIV does not cause AIDS, or that the scientific community has no evidence that it does, as in Examples 6.32 and 6.33 below, are made in order to add weight to the evidence that the medical-scientific community cannot be trusted and thus justify adopting a vaccine-critical stance.

> [6.32] **If there is evidence that HIV causes AIDS**, there should be scientific documents which either singly or collectively demonstrate that fact, at least with a high probability. There is no such document.
>
> [6.33] Important articles show **HIV was never proven, isolated, or duplicated in lab setting to be a cause of AIDS**

Denials of the causal connection between HIV and AIDS tend to provoke responses from the pro-science posters on the forum and lead to protracted debate. Pro-science posters typically use references to denials of the HIV/AIDS connection to delegitimize vaccine-critical arguments, as we see in the following example, in which the poster accuses vaccine-critical JABS members of offering tacit support to members who express wrong-headed views:

> [6.34] Gus offered dangerous medical advice to another parent. That remark stood. **John Stone denied that HIV causes AIDS. That remark stood**. Less hypocrisy please, more facts.

Other frequent expressions of the semantic sequence A BACTERIUM OR VIRUS CAUSES A DISEASE relate to a supposed connection between the simian 40 virus (SV40) and cancer. This is illustrated in Examples 6.35 and 6.36 below. SV40 in turn has a causal connection with the polio vaccine, owing to the incident in the 1960s when batches of IPV got contaminated with SV40. We therefore find references to SV40 being used to realize the semantic sequence A VACCINE CAUSES DAMAGE, as Example 6.36 demonstrates.

> [6.35] Despite mounting evidence, **the US government denies that SV-40 causes cancer**
>
> [6.36] As I said before where are the figures to show people **how many children have contracted SV40 [cancer] from the polio vaccines**? **How many cancers are actually recorded as being caused by SV40?**

The semantic sequence A VACCINE CAUSES DAMAGE also overlaps with the sequence A VACCINE CAUSES DISEASE. We see this in the following two examples.

> [6.37] If live viruses used as a vaccine can cause polio today when hygiene is generally high, it may well be that the **polio epidemics 40 to 50 years ago were also caused by immunization against polio** while hygiene, sanitation, housing, and nutritional standards were still very low.

[6.38] my son has **vaccine induced polio & as a result has SV40 & other monkey virus contamination**, & he was vaccinated in the late 80s

The examples above in which reference is made to HIV/AIDS, SV40 and cancer, and to the polio vaccine provide evidence for the view that, in discussion of health or scientific issues in the public sphere, writers draw on an interdiscursive memory bank in order to further their communicative goals (Beacco *et al.*, 2002; Moirand, 2003). Furthermore, the use of realizations of the semantic sequences A BACTERIUM OR VIRUS CAUSES A DISEASE A VACCINE CAUSES DAMAGE and A VACCINE CAUSES DISEASE to challenge medical-science or to express mistrust in it reflects Hobson-West's (2005) finding that vaccine-critical groups often link science with issues of trust and use the connection strategically.

### 6.4.3 The immune system

6.4.3.1 Meanings associated with the word *immune*

The top thirty collocates of *immune* in the *JABS* corpus (according to MI3) are shown in Table 6.14 below. The top thirty collocates of *immune* in *UkWaC* are shown alongside in order to illustrate how immune is generally used in English. The top thirty collocates of *immune* in the *NHSvax* corpus are shown in order to compare and contrast uses in *JABS* and *NHSvax*. As one might expect, the words *system(s)* and *response(s)* are the top collocates of *immune* in *JABS*, exactly as they are in *UkWaC* and *NHSvax*. In common with the collocational profile of *immune* in *UkWaC*, the words *function*, *cells*, *boost* and *suppression* occur as significant collocates of *immune* in *JABS*. These words are not among the most significant collocates of *immune* in the *NHSvax* corpus, although the notion of suppression is implicit in the presence of the collocate *suppressed*. This suggests that there are some aspects

of knowledge about the immune system which are of concern in the *JABS* corpus data but are of marginal, or no, concern in the *NHSvax* corpus data.

| | UkWaC | | JABS | | NHSvax | |
|---|---|---|---|---|---|---|
| | **collocate** | **MI3** | **collocate** | **MI3** | **collocate** | **MI3** |
| *1* | *system* | 36.261 | *immune* | 33.52 | *immune* | 27.96 |
| *2* | *response* | 32.121 | *system* | 30.30 | *system* | 25.08 |
| *3* | *cells* | 30.714 | *response* | 25.16 | *response* | 21.29 |
| *4* | *the* | 30.309 | *the* | 24.99 | *s* | 19.56 |
| *5* | *responses* | 30.132 | *systems* | 24.98 | *the* | 19.25 |
| *6* | *To* | 28.820 | *to* | 22.95 | *overload* | 19.01 |
| *7* | *and* | 28.557 | *and* | 22.82 | *weaken* | 18.57 |
| *8* | *Of* | 28.163 | *of* | 21.82 | *to* | 18.47 |
| *9* | *Deficiency* | 27.883 | *a* | 21.81 | *body* | 18.32 |
| *10* | *systems* | 27.877 | *responses* | 21.36 | *a* | 17.66 |
| *11* | *body* | 27.434 | *compromised* | 21.07 | *infant* | 17.66 |
| *12* | *weakened* | 27.345 | *that* | 21.00 | *systems* | 17.34 |
| *13* | *humoral* | 27.285 | *function* | 20.93 | *responds* | 17.31 |
| *14* | *deficiency* | 27.160 | *is* | 20.53 | *produce* | 17.05 |
| *15* | *function* | 26.878 | *an* | 20.23 | *someone* | 16.98 |
| *16* | *innate* | 26.832 | *in* | 20.13 | *meeting* | 16.39 |
| *17* | *Is* | 26.618 | *stimulation* | 20.10 | *of* | 16.23 |
| *18* | *suppress* | 26.393 | *cells* | 20.01 | *weakened* | 16.14 |
| *19* | *In* | 26.279 | *with* | 19.98 | *responses* | 16.13 |
| *20* | *Acquired* | 26.277 | *suppression* | 19.92 | *weakening* | 16.07 |
| *21* | *boost* | 26.110 | *by* | 19.73 | *overloads* | 16.01 |
| *22* | *antibodies* | 26.045 | *s* | 19.61 | *priming* | 15.81 |
| *23* | *immune* | 26.030 | *activation* | 19.29 | *bacterial* | 15.73 |
| *24* | *are* | 25.808 | *are* | 19.16 | *an* | 15.73 |
| *25* | *infection* | 25.781 | *auto* | 18.88 | *and* | 15.69 |
| *26* | *antigens* | 25.740 | *which* | 18.78 | *not* | 15.66 |
| *27* | *cell-mediated* | 25.616 | *stimulating* | 18.56 | *with* | 15.66 |
| *28* | *suppression* | 25.579 | *stronger* | 18.48 | *is* | 15.48 |
| *29* | *that* | 25.428 | *innate* | 18.41 | *are* | 15.37 |
| *30* | *disease* | 25.296 | *boost* | 18.40 | *suppressed* | 15.34 |

**Table 6.14** The thirty most significant collocates of *immune* in *ukWaC* and *NHSvax*, according to MI3[70]

The main differences between the collocational profiles of *immune* in the *JABS* and *NHSvax* data, though, lie in the absence from the *JABS* profile of the words *overload*, *weaken* and

[70] The collocation programme in Sketch Engine is case-sensitive, hence the occurrence of *Deficiency* and *deficiency* in the *UkWaC* collocate list.

*weakened*, and the presence of the words *compromised, function, stimulation, cells, activation, auto, stimulating, stronger, innate*, and *boost*. Some of these differences are relatively insignificant. For example, concordance evidence shows that where the *NHSvax* texts refer to a *weakened immune system*, in *JABS* the term *compromised immune system* is preferred. Others are not, in themselves significant, but, when concordance evidence is examined, differences can be used in the wider contexts of use. An example is *immune suppression*. Some occurrences in which the term is used indicate that the term is used in much the same way as forms of *SUPPRESS* are used in *NHSvax*. For example, it often occurs in medical-scientific contexts to refer to an immune system which is weakened owing to the presence of disease:

> [6.39] Paper by P. G. Auwaerter and Diane Griffin, (source: Clinical Immunolgy and Immunopathology, 79(2): 163-70, May 1996):
> This found that:
> * **measles produces immune suppression** which contributes to an increased susceptibility to other infections

Often, however, it is used in contexts in which it is asserted that vaccination suppresses the immune system, for example:

> [6.40] **Vaccines and Immune Suppression**
> Vaccines, all vaccines, are immune suppressing; that is they depress our immune functions.

It was noted above that *overload*, which is among the top collocates of *immune* in *NHSvax*, is not among the top collocates of *immune* in *JABS*. Concordance evidence reveals that it is not the case that the writers represented in the *JABS* texts are not concerned with the idea of vaccine overload, it is simply that a wider range of terms is used with which to refer to the issue. *Immune stimulation* and *immune activation* are frequently used in place of *immune*

*system overload*. *Immune stimulation* is interesting in that the contexts of use of this expression is most revealing of the ideological stance of many of the JABS online community. *Immune stimulation* is often pre-modified, especially with the words *excess* or *excessive*. The causes of immune stimulation are sometimes attributed to diseases, but more frequently, it is vaccines which are represented as being responsible.

> [6.41] … essential to this process is a combination of pre-existing or vaccine-induced immune dysfunction and **excess immune stimulation by a crowded vaccine schedule**.
>
> [6.42] There is evidence that with repeated and **excessive vaccine-triggered immune stimulation**, the microglia do not shut down

What is clear from looking at the examples above is that they come from medical texts. Example 6.42, for instance, comes from an article, which has been copied into a post, entitled, *The Danger of Excessive Vaccination During Brain Development: The Case for a Link to Autism Spectrum Disorders*, and authored by Russell L. Blaylock, M.D., a vaccine-critical doctor. It is apparent that JABS forum members often draw on the sort of mainstream medical-scientific discourse which is not reproduced in scientific popularization texts such as those which comprise the *NHSvax* corpus. However, they do so by copying sections of texts or entire texts into their posts, rather than incorporating this highly technical terminology into their own speech.

Although *overload* is not a highly statistically significant collocate of *immune* in the *JABS* corpus, it does occur relatively frequently in the vicinity of *immune*. There are 56 citations of *overload* in the *JABS* corpus, but it attracts a broader range of collocates in this corpus than it does in the *NHSvax* corpus. There are eight citations in *JABS* of *overload the immune system*, six of *vaccine overload*, two of *overload infants' immune systems*, one of *overload babies' immune systems*, and one citation of *immune system overload*. *Overload* also occurs in the

expressions *toxic (mercury) overload* <4>, *chemical overload* <3>, *overload of toxic substances* <2>, and *aluminium overload* <2>. Furthermore, when *overload* is used in the *JABS* corpus, it often collocates with the keywords identified in the 'vaccine additives' and 'toxicity' sub-categories of the 'vaccination and pharmaceutical products' category (see Appendix 8).

In a number of the contexts in which *immune* is used, we see attempts at accounting for the processes which lead to neurological damage. Example 6.43 of the use of *immune activation* is one such occurrence. This example comes from the 'articles' sub-corpus and the article in question is written by a vaccine-critical author.

> [6.43] We know that **immune activation of the brain**, **especially when intense and prolonged**, can precipitate the release of excitotoxins from astrocytes and microglia. Excitotoxicity is now known to be a major mechanism of neural destruction in cases of viral infections of the brain

The following example of the use of *immune cells* is similar.

> [6.44] Multiple sclerosis is also associated with the expression of human endogenous retroviruses. Since the research shows that **ERVs contribute to the development of human degenerative and inflammatory nervous system diseases, including also having a direct effect on nervous system cells and immune cells**.
> THEREFORE.....We need some URGENT research to determine the role that ERV's play in the development of Autism.

There is also concern with auto-immune conditions, as Examples 6.45 and 6.46 illustrate. Example 6.46 also illustrates the expression of a lack of trust in the authorities:

[6.45] Glad to see that this has reached the forum, **I served in the 1990 Gulf conflict and had all the injections (Some legal some decidedly suspect) and now suffer a number of auto immune problems.**

[6.46] **I dont see many cures for the plagues of auto immune diseases and other chronic health problems** but i do see their symtoms being treated but the patient is never cured of the underlying cause

A similar lack of trust in the medical-scientific community is expressed in Example 6.47, in which alternative medicine, rather than vaccination, is suggested as a way of protecting the body from viruses.

[6.47] The only way to rid the body of viruses is to **improve immune system function- increase immune cells & immune chemicals through the use of herbal medicine & homeopathy.**

The concordance evidence for *immune* in the *JABS* corpus thus shows that the belief is often expressed that multiple vaccines may prove an excessive load for the immune system. Where the beliefs expressed here differ from those expressed in *NHSvax* is that there is greater concern with identifying possible causes of neurological damage (and the processes through which they occur) and there is concern with auto-immune conditions.

### 6.4.3.2 Meanings associated with the word *immunity*

The top thirty collocates of *immunity* in the *JABS* corpus, according to MI3, are shown in Table 6.15 below. The top thirty collocates of *immunity* in *UkWaC* and the *NHSvax* corpus are given alongside, for ease of comparison. As was noted in Chapter Five, there are a number of words with a strong collocational association with *immunity* in the *UkWaC* corpus which relate to meanings of *immunity* other than the medical sense. These do not concern us here.

| | UkWaC | | JABS | | NHSvax | |
|---|---|---|---|---|---|---|
| | **collocate** | **MI3** | **collocate** | **MI3** | **collocate** | **MI3** |
| 1 | *Infection* | 29.842 | *immunity* | 32.45 | *immunity* | 27.26 |
| 2 | *cell-mediated* | 29.108 | *herd* | 27.69 | *herd* | 22.03 |
| 3 | *humoral* | 27.045 | *to* | 21.63 | *passive* | 20.22 |
| 4 | *prosecution* | 26.814 | *lifelong* | 21.45 | *generated* | 18.35 |
| 5 | *immunity* | 26.561 | *the* | 21.31 | *active* | 18.03 |
| 6 | *innate* | 26.259 | *natural* | 21.30 | *to* | 18.01 |
| 7 | *and* | 26.162 | *and* | 19.97 | *produce* | 17.62 |
| 8 | *diplomatic* | 25.901 | *of* | 19.67 | *the* | 17.43 |
| 9 | *to* | 25.785 | *cellular* | 19.34 | *boost* | 16.44 |
| 10 | *of* | 25.767 | *is* | 18.71 | *impaired* | 16.18 |
| 11 | *Innate* | 25.498 | *confer* | 18.34 | *person* | 15.99 |
| 12 | *privileges* | 25.409 | *that* | 18.12 | *lowered* | 15.93 |
| 13 | *the* | 25.302 | *test* | 17.97 | *receiving* | 15.88 |
| 14 | *from* | 25.037 | *build* | 17.85 | *conferred* | 15.79 |
| 15 | *infection* | 24.943 | *lasts* | 17.76 | *factors* | 15.63 |
| 16 | *Inflammation* | 24.912 | *not* | 17.71 | *cells* | 15.45 |
| 17 | *Crown* | 24.595 | *for* | 17.65 | *is* | 15.43 |
| 18 | *Humoral* | 24.566 | *a* | 17.60 | *animation* | 15.18 |
| 19 | *Immunity* | 24.408 | *life-long* | 17.52 | *check* | 14.99 |
| 20 | *vaccination* | 24.340 | *confers* | 17.42 | *affecting* | 14.83 |
| 21 | *mucosal* | 24.276 | *check* | 17.25 | *from* | 14.39 |
| 22 | *protective* | 24.270 | *vaccine* | 17.19 | *immunodeficient* | 14.35 |
| 23 | *herd* | 24.245 | *in* | 17.07 | *resist* | 14.35 |
| 24 | *Cell-mediated* | 23.987 | *humoral* | 17.05 | *halting* | 14.35 |
| 25 | *disease* | 23.500 | *disease* | 17.01 | *longer-term* | 14.33 |
| 26 | *in* | 23.373 | *measles* | 16.93 | *of* | 14.30 |
| 27 | *Interest* | 23.292 | *long* | 16.91 | *natural* | 13.92 |
| 28 | *PII* | 22.985 | *acquired* | 16.88 | *disease* | 13.75 |
| 29 | *is* | 22.759 | *against* | 16.79 | *phenomenon* | 13.61 |
| 30 | *natural* | 22.712 | *induced* | 16.75 | *build* | 13.48 |

**Table 6.15** The thirty most significant collocates of *immunity* in *ukWaC, JABS* and *NHSvax*, according to MI3.

If we compare the collocational profiles of *immunity* in *UkWaC*, *JABS* and *NHSvax*, we notice a number of interesting similarities and differences. The concept of herd immunity is particularly significant in the *JABS* corpus, as it is in *NHSvax*. *Herd* is the most frequent collocate (after *immunity*) in both, while it appears lower down in the *UkWaC* collocate list. Also occurring in both the *JABS* and *NHSvax* collocate lists, although not in the *UkWAC* list, are the words *provide* and *against*. Interestingly, the *JABS* and *UkWaC* collocate lists share the words *humoral* and *cellular*. These are absent from the *NHSvax* list, suggesting that these

refer to an area (or areas) of knowledge about the immune system which is not of central concern in the *NHSvax* data. Focusing now on a comparison between *JABS* and *NHSvax* alone, we see that *natural* is a collocate of *immunity* in *JABS*, as it is in *NHSvax*. However, the words *active* and *passive* do not appear in the *JABS* collocate list, implying that this distinction is not considered important. Another interesting difference concerns the presence of *lifelong* among the collocates of *immunity* in *JABS*. If we start by examining the uses of the expression *herd immunity* in the *JABS* corpus, we find that the concept is often contested. Where it is not contested is when it occurs in media reports, as can be seen in example 6.48.

> [6.48] But immunisation rates are still **well below the 95% needed for herd immunity** and are particularly low in London.

In most instances, though, either a negative evaluation of *herd immunity* is signalled, or, at the very least, authorial distance is signalled. In 17% of the 295 occurrences of *herd immunity*, for example, inverted commas are used:

> [6.49] The purpose was to achieve an **"herd immunity"** that would be as high as possible.

In the following examples, negative evaluation is expressed in more than one way. The status of *herd immunity* is evaluated as a *theory* (Example 6.50) or a *myth* (Example 6.51) or evaluated as *so-called herd immunity* (Example 6.52). Other signals of negative evaluation are explicitly marked, for example, in the expression *complete and utter bunkum* or the assertion that *no mother worth her salt would put her own child at risk.*

> [6.50] I have proved to you that the **theory of herd immunity** via vaccination is complete and utter bunkum
>
> [6.51] I dont believe in **the myth of vaccine** induced herd immunity

[6.52] No mother worth her salt would put her own child at risk (and I believe it is a very real one) **for the sake of so-called herd immunity**.

Even where a writer endorses a belief in herd immunity, the contexts of use show explicit signs of dialogicality and negotiation of the status or desirability of herd immunity. In Example 6.53, the writer engages with the (imagined) readers. The writer starts with the words *and yes*, as though simulating agreement with an interlocutor. The writer addresses the audience with the words *you guys*. He or she also attributes opinions to the readers, implying they are unlikely to endorse the concept of herd immunity. In Example 6.54, the writer addresses a specific individual whose son has a compromised immune system.

[6.53] And yes there was a subsequently a vaccine induced polio outbreak when immunization was reintroduced, the uptake was so low, c 35% (musn't say that 'cause it somehow supports **the concept of herd immunity** which is something else some of you guys dont seem to believe in).

[6.54] Also, your son's immune-compromised state is one that **would benefit from herd immunity**.

It is in discussions about herd immunity that we see evidence of the tension between the public good and the concerns of the individual. In contrast to the idea of herd immunity, which puts the needs of society in general above the needs, or desires, of the individual, the idea of natural immunity privileges the individual. If the individual can acquire immunity naturally, the rest of society need not concern him/her. The contexts of use of *natural immunity* show that the term has a largely positive semantic prosody. The concept of natural immunity is often represented in binary opposition to that of vaccine-induced immunity. Unsurprisingly, natural immunity is evaluated more positively than vaccine-induced immunity. In Examples 6.55 and 6.56, natural immunity is evaluated as more effective than vaccination.

[6.55] **If vaccination was as effective as natural immunity** (as provaxer would have you believe) then a vaccination rate of 68% would be enough to stop outbreaks.

[6.56] i dont want to detract from why IMO **an immune system bought on breast milk and natural immunity cannot be beaten by any scientist or product** in building up a strong immune system generally.

In many contexts, the proposition is expressed that natural immunity confers lifelong immunity:

[6.57] **If a child has the actual illness they will have lifelong immunity**. The vaccine doesn't provide this.

Meanwhile, the claims made by the medical-scientific community that vaccination confers lifelong immunity are challenged:

[6.58] **They start with a claim of one shot gives lifelong immunity** and then...everyone is being reminded to get shots, forever and ever and ever.

One of the most interesting differences between the two corpora regarding the ways in which notions of immunity are represented is signalled by the presence among the significant collocates of *immunity* in the *JABS* corpus of *humoral* and *cellular*, and by their corresponding absence in the *NHSvax* profile. There are 31 occurrences of *humoral* in the *JABS* corpus and 241 of *cellular*. Although all occurrences of *humoral* refer to the immune system, not all occurrences of *cellular* do. However, there are 28 occurrences of *cellular immunity* and 26 occurrences of *cellular immune (system/ defense/ responses/ recognition/ deficiencies/ function)*. One cannot claim that these expressions occur with high frequency in the data, however, *humoral* and *cellular* occur frequently enough in the vicinity of *immunity* to be statistically significant. Where *humoral immunity* and *cellular immunity* are used in the *JABS* data, the terms typically occur in uploaded texts, as the following example illustrates:

> [6.59] **Cellular (or cell-mediated) immunity** refers to the part of the immune system that is engaged during natural exposure to any given pathogen or virus. **Humoral immunity** refers to the part of the immune system that is present in internal fluids, and rather simply put, it is responsible for developing antigen specific antibodies.

In most occurrences of *humoral immunity* and *cellular immunity*, the terms occur in contexts in which an argument is put forward that vaccinations harm the immune system:

> [6.60] As many Doctors and Scientists have shown for example **vaccines can diminish cellular immunity** in favour of an unnaturally high antibody responce leaving the person far more prone to chronic diseases and auto immune diseases in the future.

> [6.61] I'm sure your and your faiths response is, well vaccinate em younger and more often. That seems quite unwise, but hell let's risk it, vaccines are entirely safe right?): **..."Humoral immunity was deficient in 6-month-old infants given measles vaccine**, even in the absence of detectable passively acquired neutralizing antibodies.

The fact that vaccine-critical forum posters argue that vaccines damage the immune system is not remarkable in itself. What is interesting is the way in which discursive resources from the domain of medical-science are drawn on in order to advance an argument. It is also interesting that, in so doing, these individuals represent a more complex view of the immune system than that which is represented in the *NHSvax* data.

### 6.4.4 Health and food

The words which were identified among the *JABS* keywords as relating to notions of health are *healthy, ill, sick*, and *food*. The context in which these words are used shed little light on notions of health, but rather on notions of illness. *Healthy* and *ill* typically occur in contexts in which the safety of vaccines is challenged:

> [6.62] The MMR court cases were and still are vital not only to the families involved in the pursuit of justice for their children, but for **all parents who are concerned about whether the vaccines they are giving their healthy children are safe.**

> [6.63] However when **I inquired of the Department of Health about how many children were still seriously ill from the after effects of the November immunisation campaign,** their spokesman insisted that no children at all are suffering today from any after effects

*Sick*, on the other hand, is more usually used in order to criticise the behaviour of certain people, for example:

> [6.64] Steve I followed the link to your blog. **You get amusement from insulting and mocking the parents of sick children right?**

*Food* occurs frequently in references to the Food and Drug Administration. Of the 966 occurrences of *food*, 173 occur in references to the Food and Drug Administration. Where it is not used in this context, it is used to talk about problems with food. For example, the collocates which occur most frequently 1 space to the right of *food* are *allergies* <70>, *allergy* <23>, *intolerances* <18> and *intolerance* <16>. This results from the fact that many of the parents of autistic or otherwise neurologically damaged children report that their children suffer from bowel problems, something which inspired Andrew Wakefield's research.

### 6.4.5 Summary of Section 6.4

The analysis in this section has shown that there are many similarities between the ways in which concepts of health and immunity are expressed in the *JABS* and *NHSvax* corpora. A lot of the similarities are due to the fact that the only resources we have with which to express beliefs about the immune system are those which emerged from the medical-scientific domain. One interesting feature is that there is a concern evident in both corpora with the idea that multiple vaccines can adversely affect immune system function, but, where this is realized in the *NHSvax* data through expressions containing forms of the word *OVERLOAD*, in *JABS*, the expressions *immune activation* and *immune stimulation* are preferred. The main

differences between the discoursal patterns found in each of the two corpora, though, are due to differences in the ideological stance taken in each and to differences in the communicative aims of each. The main concern of the vaccine-critical voices evident in the *JABS* data is to resist the imperative to vaccinate. The idea that achieving herd immunity is a desirable goal is resisted. Sometimes, this is done by expressing a negative evaluation of the concept; sometimes this is done by contesting the notion that herd immunity is an achievable goal. It is also sometimes evaluated as a *myth*, a *theory* or a *concept*. More interestingly, the data reveals that the writers represented in the *JABS* corpus sometimes draw on knowledge about the immune system which originates in medical-science but which is very rarely referred to in the *NHSvax* data, if at all. We therefore see references to different types of immunity, namely *humoral immunity* and *cellular immunity*. In this way, vaccine-critical writers exploit medical-scientific discourse in order to challenge the science which is used to justify the government's vaccination policy. In some instances, the trustworthiness of the medical-scientific establishment is questioned. We see this in the way that the notion that HIV causes AIDS is sometimes challenged. We also see, in the mix of mainstream medical-scientific discourse, evaluative expressions, and references to diet and lifestyle which is apparent in the data, evidence of the patchwork of discourse types which is said to characterize contemporary debate about health and science issues (Beacco *et al.*, 2002; Moirand, 2003). Finally, the references to vaccine scares of the past or to theories about disease held by only a minority of people, point at the exploitation of an interdiscursive memory bank (Moirand, 2003; Beacco *et al.*, 2002).

## 6.5 Expressions of risk, safety, and danger in the *JABS* corpus

### 6.5.1 Summary of conclusions drawn in Chapter Five regarding expressions related to health and the immune system in the *NHSvax* corpus

In Chapter Five, it was concluded that the lexico-grammatical patterns identified in the *NHSvax* corpus to express notions of risk are typical of discourse from the field of healthcare. Assessments of degrees of risk are expressed using terms such as *HIGH* and *INCREASE*. The risks posed by diseases are represented as greater than those posed by vaccination. This idea is frequently expressed using the expression *the benefits … outweigh the risks* or *the risks … are outweighed by the benefits*. In order to counter the belief that vaccines are potentially harmful, the expression *there is no evidence of (increased) risk* is often used. The idea that vaccines are safe is sometimes expressed by referring to their role in protecting the individual from diseases. In fact, there is a strong collocational relationship between *RISK* and *PROTECT* in the *NHSvax* corpus. Undesirable health outcomes caused by vaccines are usually referred to using the terms *(adverse) reactions* or *side-effects*, both of which are prosodically more neutral than *damage*. There is evidence, too, of tension between concern for the public good and concern for the individual's own interests or the interests of their child. Since the purpose of immunization programmes is to ensure that herd immunity is achieved, there is a high occurrence of terms which refer to groups of people considered vulnerable, such as *(at) risk groups*. Because the imperative to make sure as many are vaccinated as possible in order to protect the entire population can conflict with the individual parent's interests, the NHS texts often appeal to the parents' concern for their child's health.

### 6.5.2 The collocational profiles of *risk* and *risks*

6.5.2.1 The 100 statistically most significant collocates of *risk* and *risks*

Although *risk* is not as statistically significant in the *JABS* corpus as it is in the *NHSvax* corpus, it nonetheless occurs in a relatively high position in the *JABS* keyword list (ranking at positon 60). Its raw frequency is somewhat high: there are 1,774 citations of *risk* in *JABS*. Because *risk* is so frequent and participates in a wide range of lexico-grammatical patterns, it is helpful to examine as long a list of collocates as is feasible. *Risks* occurs far less frequently than *risk*, but, since the two words are being considered together, an equal number of collocates has drawn up for *risk* and *risks*. The top 100 collocates of *risk* and *risks*, according to MI3 are shown in Appendix 11. The collocational profiles of *risk* and *risks* display a number of the features which were observed in Chapter Five, when the profiles of *risk* and *risks* in the *NHSvax* corpus were examined. The words *of*, *the*, and *at* are the three most statistically significant collocates of *risk*, suggesting that the expressions *the risk of* and *at risk of* are likely to recur with frequency. As one might expect, there are also words which indicate the assessment of degrees of risk, such as *increased* (position 6), *increase* (14), *high* (18), *greater* (24), *higher* (25), *reduce* (39), *increases* (64), *reduces* (75) and *low* (90). There are other words here, which were not highly significant in the *NHSvax* corpus data, which indicate the assessment of degrees of risk in the technical sense. These are f*actors* (position 12), *relative* (33), *factor* (34), *ratio* (58), and *assessment* (71). The words *outweighs* (35) and *worth* (51) also suggest assessments of degrees of risk.

In contrast to the *NHSvax* corpus data, the only word which denotes a group of people is *children* (position 16). There is, understandably, less concern in the *JABS* corpus data with identifying vulnerable groups deemed to be in particular need of vaccination. The effects of

vaccines on children are, of course, a key concern. Parents' concerns for their own children are central to much discussion on the JABS site, so *child* (28) is a significant collocate. There are also collocates which reflect the concern with potential vaccine damage and the alleged MMR-autism link: *cancer* (position 17); *vaccine* (27); *disease* (29); *catching* (37); *cervical* (41); *complications* (43); *measles* (45); *autism* (46); *infection* (53); *death* (56); *asthma* (62); *vaccines* (68); *vaccination* (77); *adverse* (82); *vaccinating* (88); *reaction* (89). In marked contrast to the profile of *risk* in the *NHSvax* corpus, though, are the words *put* (position 20), *putting* (21), *take* (42), and *run* (66). *Put, putting* and *take*, as collocates of *risk* in the *JABS* corpus, occur in the expressions *PUT (someone) at risk* and *TAKE the/a risk*. When used this way, they realize the HARM sub-frame of the RISK frame (Fillmore and Atkins, 1992). *Run* is used in the expression *run the risk*, an expression which, in general usage, realizes the CHANCE sub-frame (ibid.).

At the top of the collocate list for *risks* we see the words *outweigh* (position 2), *benefits* (3), *the* (4), and *of* (5), which, together, bring to mind the expression which recurs a number of times in the *NHSvax* corpus: *the benefits outweigh the risks*. We can also find other words related to the assessment of degrees of risk, for example, *balance* (10), *outweighed* (11), *greater* (30), *weigh* (31), *weighing* (34), *outweighs* (51) and *evaluate* (60). What is particularly striking, though, is the large number of words which can be used to express a writer's evaluative stance. These include adjectives such as *real* (16), *serious* (32), *favourable* (37), s*ignificant* (72), *small* (78), and so on; adverbs such as *properly* (32); intensifiers, such as *far* (15), and *overwhelmingly* (22); and hedging devices such as *potential* (35) and *possible* (42).

6.5.2.2 Clusters with *risk* and *risks*

Appendix 12 shows all of the two- to six-word clusters associated with *risk* and *risks* in the *JABS* corpus. As is predictable, they reflect many of the patterns highlighted in the examination of the collocates. What is clearer is the mix of discourses that are apparent. Expressions which are typical of the discourse of the domain of healthcare (Hamilton, Adolphs and Nerlich, 2007), such as *at risk* <497>, *high risk* <177>, *increased risk* <161>, *risk factors* <98>, and *health risks* <42> occur in the *JABS* corpus more frequently than any other recurrent string related to *risk*. Interestingly, technical expressions which were not identified by Hamilton, Adolphs and Nerlich (ibid.) and which do not occur as statistically significant feature of the *NHSvax* corpus, such as *risk benefit* <91> and *relative risk* <43>, are highly frequent in *JABS*. Further down the list are other technical terms from the field of risk management, such as *risk benefit ratio* <20>, *risk assessment* <16>, *risk benefit analysis* <14>, and *absolute risk* <12>. Although not as frequent as the occurrences of expressions from the field of healthcare, expressions of a non-technical nature are also relatively frequent, for example, *take the risk* <36>, *worth the risk* <34>, *put at risk* <28>, *run the risk* <24>, and so on. The data therefore shows a rich interdiscursive mix. One further interesting feature is that, while the *NHSvax* data had several occurrences of the expression *at risk children*, in the *JABS* data, we find *children at risk* and *child at risk*. These expressions warrant closer attention. As was seen in the *NHSvax* data, clusters involving *outweigh* and *benefits* are frequent, for example, *outweigh the risks* <46>, *(the) benefits outweigh the risks* <12>, and *far outweigh the risks* <9>. We also see *risks and benefits* <37>, and *benefits and risks* <10>.

### 6.5.3 Technical uses of *risk* and *risks*

As was observed in Section 6.4, when uses of *immune* and *immunity* were examined in the *JABS* corpus which corresponded to the patterns typical of medical-scientific discourse, technical meanings of *risk* and *risks* in the *JABS* corpus typically occur in articles and documents which have been uploaded onto the website or imported into the discussion forum. This is the case for expressions such as *at high risk, increased risk, risk factors, reduce the risk*, and so on, as the following examples illustrate:

> [6.65] A woman found to be positive for the same strain (genotype) of HPV on repeated testing is highly likely suffering from a persistent HPV infection and is considered to be **at high risk of developing precancerous intraepithelial lesions in the cervix**
>
> [6.66] MMR is **not associated with an increased risk of pervasive development disorders** such as autism
>
> [6.67] There is no credible epidemiological evidence to support the view that measles vaccination is **a risk factor for Crohn's disease or any other inflammatory bowel disorder**
>
> [6.68] After adjustment for potential confounders, **the relative risk of autistic disorder in the group of vaccinated children**, as compared with the unvaccinated group, was 0.92 (95 percent confidence interval, 0.68 to 1.24), and **the relative risk of another autistic-spectrum disorder** was 0.83 (95 percent confidence interval, 0.65 to 1.07).

However, there are cases in which forum posters deliberately draw on risk discourse from a technical register in order to perform expertise. Technical register terms of this kind are typically used, when they are used at all, only by high-frequency posters. The following is one such example:

> [6.69] The trouble is that since we are all being ignored (and even you admit that getting listened to would be an uphill struggle) **there is no basis for assessing the risk/benefit ratio.**

As is apparent in Example 6.69, in the second clause, the writer uses the technical expression *there is no basis for assessing the risk/benefit ratio* in order to challenge the science.

Furthermore, the first clause expresses mistrust in authority (*we are all being ignored*) and the two clauses are linked with the causal conjunction *since*. On occasions, once a technical term has been used by one poster, it may get repeated and reframed by subsequent members posting in the same thread. In a thread titled *The "anti-meningitis" vaccine: another turkey?*, for example, the term *risk/benefit ratio* is used for the first time by the high-frequency vaccine-critical poster, Minority View:

> [6.70] My research into vaccines has turned up a lot of problems. **When you look at the complete heap, the risk/benefit ratio doesn't look all that great, frankly**.

In the example above, the technical expression *risk/benefit ratio* is embedded in a context in which a number of negatively evaluative expressions are used. The *risk/benefit ratio* itself is evaluated as not looking 'all that great'. Use of the term *the complete heap* also encodes a negative evaluation. Investigation of a random sample of 100 concordance lines for heap in the *ukWaC* corpus shows that among the meanings of *heap* are notions of unpleasantness and uselessness. The most frequent left-hand collocates of *heap* connote meanings of unpleasantness and large quantities. They are *compost* <10>, *manure* <4>, *whole* <3>, *big* <2>, *dung* <2>, and *scrap* <2>. *Heap* occurs either in clause-final position, or as part of a complex nominal group in which it governs a prepositional phrase beginning with *of*. *Heap of* is followed by mass nouns or countable nouns. It may be followed by a noun which denotes something pleasant, for example, *fun* <3>, but more commonly, it is followed by nouns which denote undifferentiated collections of objects or materials, the semantics of some of which indicate that they are unwanted, for example, *stones* <4>, *junk* <2>, *rubble* <2>, *rubbish* <2>, *sand* <2>, and so on. Finally, the adverb *frankly* emphasizes the writer's commitment to the truth of the proposition.

In a later post in the thread, the high-frequency, pro-science poster, Deepika, responds to Minority View's post.

> [6.71] Minority, I think you are forgetting that we are discussing meningococcal vaccine. You have been saying: "My research into vaccines has turned up a lot of problems. When you look at the complete heap, the risk/benefit ratio doesn't look all that great, frankly." Presumably you think meningo C vaccine has **a poor risk/benefit ratio, as it is part of the "heap"** □ **too?**

As the example above illustrates, she not only repeats the term *risk/benefit ratio,* she reproduces three entire clauses from his post. The quoted stretch of text is framed by a preface, in which she challenges the propositional content of his assertion, on the grounds of relevance. She offers justification for her challenge in the clause complex which follows the quoted segment. She attributes to him the proposition that the meningococcal C vaccine has a poor risk/benefit ratio. She distances herself from the proposition by framing it with the reporting clause *presumably you think.* She justifies her decision to attribute the proposition to him by stating that the vaccine belongs to the class of vaccines whose efficacy he dismisses. She distances herself from this evaluation of vaccines by repeating his evaluation of vaccines as a 'heap' and putting the term *heap* between quotation marks.

Deepika herself is later challenged by another vaccine-critical poster:

> [6.72] But are you able to answer your own question Deepika? **Do you know what the risk/benefit ratio is for say my children**?

We see in the *JABS* corpus data that, where technical terms from the medical-scientific field are used, they most frequently occur as parts of texts which have been incorporated in their entirety onto the site, frequently in discussion forum posts. High-frequency posters are more likely than other forum participants to use technical terms themselves, but the contexts in which they use these terms show extensive reformulation. There is far greater use of

evaluative expressions than one sees in the *NHSvax* data. Once a term is used in a thread, it may be picked up on by other posters and undergo further reformulation. However, instances such as these are relatively infrequent. Forum participants evidently have other lexico-grammatical resources on which they draw. These are examined in the next sub-section.

### 6.5.4 'Lay' expressions of risk

6.5.4.1 Risk discourse and the expert-lay dichotomy

As argued in Chapter Two, it is misleading to draw a clear distinction between lay discourse and expert discourse. The boundaries between expert and lay knowledge are extremely fuzzy (Myers, 2003). Nonetheless, if it is the case that non experts have their own discursive resources on which to draw (ibid.), then we can talk about 'lay' discourse, to a certain extent. The expressions related to risk which I identify as examples of 'lay' discourse, and which will be subjected to closer scrutiny in this section are *put at risk, take the risk, run the risk* and *worth the risk.* I identify them as such on the grounds that they are terms which Hamilton, Adolphs and Nerlich (2007) observed are used to talk about personal risk. I add to this group the expressions *children at risk* and *child at risk*, because concordance evidence shows that they have a strong collocational relationship in the *JABS* corpus with *PUT*.

6.5.4.2 Expressions with *risk* which realize the HARM and CHANCE sub-frames

When one examines the concordance evidence, one finds that the strings *children at risk, child at risk* and *put at risk* are all realizations of the semantic sequence PUT SOMEONE'S CHILD/CHILDREN AT RISK. The expression *PUT at risk*, as used in the *JABS* corpus, realizes the HARM sub-frame of the RISK frame. What characterizes the HARM sub-frame is that it involves an Actor and an undesirable outcome. In the HARM sub-frame, the individual is

faced with a choice of actions and harm may ensue as a result of the action. The choice involved in the contexts in which realizations of the semantic sequence PUT SOMEONE'S CHILD/CHILDREN AT RISK is used in the *JABS* corpus is obviously whether or not to vaccinate. What is interesting about the concordance evidence for *children at risk, child at risk* and *put at risk* is that different choices are made as to who is represented as Actor and what constitutes the action. The concordance evidence for *children at risk* is shown in Concordance 6.5 below.

| N | Concordance |
|---|---|
| 1 | ing to anecdotal evidence and the drivel they read in the press are putting all children at risk to three awful diseases. MMR is complusory in the US and in |
| 2 | ntly at least, voluntary. The Government claims that the database will identify children at risk of poverty, abuse or future criminality. But since when did fil |
| 3 | ll made his name by pioneering the use of covert video surveillance to identify children at risk of abuse. During an eight-year study at London's Royal |
| 4 | e mitochondria, " Terry Poling told the government panel. "We need to identify children at risk, and we need to learn how to immunize them safely. We need to |
| 5 | les either - so please can you tell me who it is actually putting non-immunised children at risk? I do not mind sharing with you that I have had singles for me |
| 6 | many doctors are concerned that a drop in vaccination levels could leave many children at risk. Dr Wakefield's supporters gathered outside the hearing The U |
| 7 | able to afford this, which I think is very unfair that I should have to put my children at risk. Good luck to you all. Message received on 07/02/02 from: aw |
| 8 | ment. I know there are risks in all areas of life, but I am not going to put my children at risk from a vaccine which doesn't even give much if any long term pr |
| 9 | study undertaken by the HPA didn't suggest there were a large number of children at risk from the disease.' "But Dr George Kassianos, RCGP |
| 10 | the un-scientific ramblings of a discredited man. They are putting the lives of children at risk for the sake of a circulation boost. It looks now as though we |
| 11 | irrhosis, which leads to liver failure or liver cancer. To improve coverage of children at risk, the Joint Committee on Vaccination and Immunisation is to |
| 12 | as not enjoyable. With people not immunising their children it is putting other children at risk ie babies, children with low immunity and leukimia etc. When I |
| 13 | deciding whether your child should be immunised or not you are putting other children at risk. I certainly will not be taking my son to play groups, sports c |
| 14 | ns. Not sure either is safe but dont want to leave him unprotected or put other children at risk. I work with children with learning disabilitites, some who als |
| 15 | helps a lot in the beginning. dont want to leave him unprotected or put other children at risk That is classic allopathic propaganda. Vaccines don't protect |
| 16 | ny "Refusing to have your child/children vaccinated and putting them and other children at risk because you have a personal conviction that vaccines are bad. |
| 17 | #1: Refusing to have your child/children vaccinated and putting them and other children at risk because you have a personal conviction that vaccines are bad. |
| 18 | cinating said "you know by not vaccinating your child you are putting all other children at risk". my wife came home very annoyed and upset about this and |
| 19 | r child cannot have a vaccine for medical reasons, do you want to put all other children at risk because you feel that will help your own child with a so called |
| 20 | ce From The Times September 27, 2008 Is our fear of vaccinations putting our children at risk? Despite evidence of their safety, we're still reluctant to ha |
| 21 | acy. I stand by my original view that trivailising diseases like this puts our children at risk. john United Kingdom |
| 22 | of mercury may have on infants and to developbetter measures to screen out children at risk of an adverse vaccine reaction. The NIH and CDC have been slo |
| 23 | instead the IoM has issued a flawed, incomplete reportthat continues to put....children at risk . PART M FLAWED UK REGULATORY AND MONITORING |
| 24 | vaccine, which is administered throughout the world, could put children at risk of autism or bowel disease. The finding, published in The Lance |
| 25 | calcitriol, which is important for brain development. 3. Staying indoors puts children at risk of increased exposure to household chemicals that might play a |
| 26 | is serious and in some cases it can be fatal. Delaying immunisation puts children at risk.â¬" PCTs in London PCTs will receive Â£60,000 each and PCTs |
| 27 | is serious and in some cases it can be fatal. Delaying immunisation puts children at risk.â¬" PCTs in London PCTs will receive Â£60,000 each and PCTs |
| 28 | I am just too impatient but this is so impotant as a human rights issue putting children at risk on a massive scale. It is child abuse & I was a victim. There |
| 29 | I am just too impatient but this is so impotant as a human rights issue putting children at risk on a massive scale. It is child abuse & I was a victim. There |
| 30 | within government agencies. Our government will never pay the price to screen children at risk for autism â¬ that would be too proactive and expensive on the |
| 31 | ght facilitate viral persistenceand immunopathology The key to defining the children at risk was the examination of the co-factors that mightinterfere with |
| 32 | problems and a family medical history of autoimmune disease are the children at risk of being vaccine damaged with a vaccine that contains three liv |
| 33 | g a national outcry among parents who feared the triple vaccine could put their children at risk of autism. The hearing, which began in July 2007 was initially |
| 34 | g a national outcry among parents who feared the triple vaccine could put their children at risk of autism. The Panel is to inquire into allegations that the t |
| 35 | ted pregnant women, there is no way to identify and selectively vaccinate those children at risk of infection (Margolis, 1991). Another reason we vaccinate inf |
| 36 | ficant. â¬SShe said that large-scale population studies would never find those children at risk. â¬SIt is interesting the studies that support the theory of a |
| 37 | the World Health Organization remain unconvinced that thimerosal puts young children at risk. There is no evidence that removal of thimerosal from vaccines |

**Concordance 6.5** Concordance lines for *children at risk* in the *JABS* corpus

In Concordance 6.5, we see a number of instances in which the vaccines are represented as the source of danger, vaccination is the action, and the parent is (potentially) the Actor. Example 6.73 comes from the message board sub-corpus and Example 6.74 from the discussion forum:

> [6.73] Unfortunately as I am on disability benefits I am not i a position to be able to afford this, which I think is very unfair **that I should have to put my children at risk.** (Line 7)

> [6.74] **I am not going to put my children at risk** from a vaccine which doesn't even give much if any long term protection. (Line 8)

What is clear in the above examples is evidence of the tension between the imperative to vaccinate and the parents' sense of responsibility for their child. The expression *that I should have to put my children at risk* implies that the parent feels under pressure from the authorities to vaccinate. *I am not going to put my children at risk* suggests the individual is aware of their responsibility as a parent. Predictably, when the causal hypothesis is referred to, the vaccine is the Actor as well as the source of danger, while autism is the potential outcome. We see this in the following example from a news text:

> [6.75] This was followed by the publication of a paper in The Lancet sparking a national outcry among parents who feared **the triple vaccine could put their children at risk of autism.** (Lines 33 and 34)

More frequently, it is the act of choosing not to allow children to be vaccinated which is represented simultaneously as the source of danger and the Actor. Examples 6.76 to 6.78 illustrate the voices of pro-science JABS members. Example 6.79 is a quotation from the then DoH director of immunisation policy.

> [6.76] **With people not immunising their children it is putting other children at risk** ie babies, children with low immunity and leukimia etc. (Line 12)

[6.77] deciding whether your child should be immunised or not you are **putting other children at risk** (Line 13)

[6.78] **Refusing to have your child/children vaccinated and putting them and other children at risk** because you have a personal conviction that vaccines are bad. (Lines 16 and 17)

[6.79] **Delaying immunisation puts children at risk** (Lines 26 and 27)

The concordance evidence for *child at risk* and *put at risk* is similar to the evidence for children at *risk* (see Concordance 6.6).

| N | Concordance |
|---|---|
| 1 | ional medicine, then it's inconceivable why any parent would consider putting a child at risk by treating eczema conventionally, when homeopathy is a safe |
| 2 | single vaccines, with some time lags between vaccinations, would leave a child at risk from mumps, measles and rubella which could result in death or |
| 3 | vaccines, with time lags of a month or more between the three, would leave a child at risk from mumps, measles or rubella. This also has an impact upon |
| 4 | loping Hib Later as a direct result of their prior vaccines strain? Why put any child at risk of any further vaccines adverse effects without answering a valid |
| 5 | your child from diseases" (said by a doctor) "You are putting everyone else's child at risk at this school by refusing your children to be vaccinated" (said b |
| 6 | should have their children excluded from school as they put everyone elses child at risk. Write in to the Society letters page if you have time. Its all to |
| 7 | from health visitors and doctor.Even telling me that i am seriously putting my child at risk by doing so unfortunatly forgetting the risk with mmr jab. I want |
| 8 | cally treating people like us like idiots because we are not willing to put our child at risk from the MMR! I have an autistic cousin and I am dyslexic myself |
| 9 | parents whose children have suffered because of it, and do not want to put our child at risk. Also, WE DONT LIKE BEING LIED TO ! The government(s) LIED |
| 10 | them for attempting to do this. No mother worth her salt would put her own child at risk (and I believe it is a very real one) for the sake of so-called he |
| 11 | for the avoidance of any doubt I'm not refering to the vaccination putting the child at risk!!!) Occam48 not given |
| 12 | for the avoidance of any doubt I'm not refering to the vaccination putting the child at risk!!!) " Of couse these children are NOT PROTECTED Many children |
| 13 | for the avoidance of any doubt I'm not refering to the vaccination putting the child at risk!!!) Occam48 Posted - Answers to your comments:- Ellen Bolte,s s |
| 14 | at for once their may be some truth to the fact that vaccine may be putting the child at risk of DEATH. Just so you know the following vaccines Package Insert |
| 15 | at for once their may be some truth to the fact that vaccine may be putting the child at risk of DEATH. Just so you know the following vaccines Package Insert |
| 16 | as one) and have 4-6wks between jabs. Doctors will tell you this places the child at risk because they go longer before they get immunisation. But having |
| 17 | ay become really anxious if they think my non vaccinated child is putting their child at risk! I hope too that greater awareness could lead to greater support |
| 18 | n to me, that Iâ¬"m ruining the herd immunity â¬¦ that my child would put their child at risk,â¬□ she says. While the vast majority of American parents |
| 19 | ly. All the doctors care about is the money they get. No one wants to put their child at risk, but you do have to decide for yourself. Message received on 06 |
| 20 | washed into thinking we are dangerous and somehow putting their vaccinated child at risk and utter tosh about "herd immunity" repeated like a mantra and |
| 21 | "I don't vaccinate my kids, you do, you are a terrible person for putting your child at risk". However, it is a matter of position. Us pro-vaccination people |
| 22 | nsated financially. It's not entirely clear what family history would put your child at risk for vaccine problems, but Frances Page Glascoe, a professor of |

**Concordance 6.6** Concordance lines for *child at risk* in the *JABS* corpus

The data shown in Concordance 6.6 clearly shows evidence of the voice of the medical-scientific community, as the example in Line 2 demonstrates:

[6.80] **Offering single vaccines**, **with some time lags between vaccinations**, **would leave a child at risk from mumps, measles and rubella** which could result in death or serious illness. (Line 2)

In this instance, it is the act of offering single vaccines which represents the Actor and the source of danger. Occurrences such as these are rare, though. What is found most frequently in the concordances for *child at risk* are instances in which vaccines, or the act of vaccinating, represent the source of danger. We see this in the three repeated citations of *the vaccination putting the child at risk* (Lines 11, 12, 13) and the two repeated instances of *vaccine may be putting the child at risk of death* (Lines 14 and 15).

With *put at risk*, on the other hand, instances in which the decision not to vaccinate is represented as the source of danger are more frequent (see Concordance 6.7).

| N | Concordance |
|---|---|
| 1 | ho are first put at risk as children for vaccine adverse events and then again put at risk as adults for a disease that the vaccine fails to protect against l |
| 2 | se of vaginal lesions and irritation. It's bad enough that we've been duped and put at risk, but worse than that is the fact that the hype surrounding nonoxynol |
| 3 | under 1 and those who have a GENUINE reason for not having MMR are put at risk by none vaccinators and those who leave gaps between separate |
| 4 | with the unspeakable username--no! We don't know how many people are put at risk because of not vaccinating for mumps. Please tell us. What was the |
| 5 | to the disease against which they have been immunised. So they are put at risk to no benefit. 9. Epidemics of foot and mouth disease always start |
| 6 | made a mistake in causing fever to arise during illness, our children will be put at risk. There is an urgency for us to re-examine our basic assumptions abo |
| 7 | the biggest ever investigation into links between MMR and autism - has been put at risk by the submission to the Medical Research Council funding body. |
| 8 | ne 16, 2006 that Children in the most deprived areas of the country are being put at risk because GPs are not providing vaccinations, figures suggest. One in |
| 9 | d at 00:02 on April 07 2008. New evidence has shown children's lives are being put at risk by a surge in the use of controversial tranquillising drugs which a |
| 10 | of the single MMR vaccines, even if this means that a few individuals are being put at risk by their parents' refusal to have the combined vaccination. Because |
| 11 | e-up of the MMR vaccine is a public health disaster. Children's lives are being put at risk daily because of the collapse in confidence over the safety of the t |
| 12 | hey had been getting lots of calls from concerned parents whose kids are being put at risk from sitting in classes with unvaxed kids! A lot of time spent plan |
| 13 | other members of the community. This is because new born babies are being put at risk. Mothers cant pass on the vaccination immunity in the same way as |
| 14 | ' reference ranges. It is the health of ALL hypothyroid patients that is being put at risk because of the GMC. Please help all hypothyroid patients by signing |
| 15 | Disease in our immediate family & are do not want our 1 year old son being put at risk by the MMR vaccine. If anyone knows of a doctors/clinic where we |
| 16 | requiring purchase of multiple doses. The losers are the people, who are first put at risk as children for vaccine adverse events and then again put at risk a |
| 17 | hing and they are trying to weather the storm "Not likely!" not when my baby is put at risk.....Anyone out there know where I can speak to someone with a |
| 18 | Boston Globe needs to know? How many US children's lives do you want to put at risk by publishing drug industry advertorials like the one by Paul Offit? |
| 19 | obscene profits from a crony deal with Governor Rick Perry, while children were put at risk. Last November we reported on a case in Prince George's County, |
| 20 | obscene profits from a crony deal with Governor Rick Perry, while children were put at risk. Last November we reported on a case in Prince George's County, |
| 21 | y lives you put at risk by not being vaccinated? What? (1)How many lives were put at risk? (2)How were they at risk? Edited by - Wanda on 12/16/2008 17:56:4 |
| 22 | y lives you put at risk by not being vaccinated? What? (1)How many lives were put at risk? (2)How were they at risk? I think he/she has gone back to the pen |
| 23 | because you weren't vaccinated? Do you have any idea how many lives you put at risk by not being vaccinated?** What are you ranting on about?? lives at |
| 24 | because you weren't vaccinated? Do you have any idea how many lives you put at risk by not being vaccinated? John Stone |
| 25 | because you weren't vaccinated? Do you have any idea how many lives you put at risk by not being vaccinated? Me thinks MKF1MKF must be a male |
| 26 | because you weren't vaccinated? Do you have any idea how many lives you put at risk by not being vaccinated? What? (1)How many lives were put at risk? |
| 27 | because you weren't vaccinated? Do you have any idea how many lives you put at risk by not being vaccinated? MinorityView |
| 28 | because you weren't vaccinated? Do you have any idea how many lives you put at risk by not being vaccinated? What? (1)How many lives were put at risk? |

**Concordance 6.7** Concordance lines for *put at risk* in the *JABS* corpus

We see, for example, the six repeated instances of *do you have any idea how many lives you put at risk by not being vaccinated?* (Lines 23-28). There is also the citation for *children's*

*lives are being put at risk daily because of the collapse in confidence over the safety of the triple jab* (Line 11).

In terms of the social implications of what is evidenced in the concordance lines, it is clear to see expression of the tension between concern for the public good and concern for the interests of the parent and the individual child. Some of the contexts of use indicate that parents feel they are put under pressure by the authorities to accept vaccination, for example:

> [6.81] Unfortunately as I am on disability benefits I am not i a position to be able to afford this, which **I think is very unfair that I should have to put my children at risk**.

The expression of deontic modality encoded in the phrase *I should have to* in the example, implies that the writer feels a sense of obligation and that the obligation is imposed on her by others. Resistance to the perceived coercion is expression in her evaluation of the situation as *very unfair*. A similar sentiment is expressed in the following two examples, in which the words of health professionals are reformulated and represented in the posts:

> [6.82] The health visitor, after reciting everything she has been taught about the dangers of not vaccinating said "you know **by not vaccinating your child you are putting all other children at risk"**. **my wife came home very annoyed and upset about this**
>
> [6.83] I have a 19 month old son who i have flatly refused to give the mmr jab to, despite serious objections from health visitors and doctor. **Even telling me that i am seriously putting my child at risk by doing so**

In Example 6.82, the health visitor's interactional behaviour is given a negative evaluation through use of the word *reciting*, which implies the health visitor uncritically repeated standard medical-scientific dogma. In Example 6.83, the writer expresses resistance to the

coercion she perceives by stating that she refused to accept the vaccine, despite the health professionals' objections.

Few people on the forum explicitly refer to feeling torn between competing interests. This is one exception:

> [6.84] Not sure either is safe but **dont want to leave him unprotected or put other children at risk.**

The phraseological environment in which *take the risk* occurs indicates that the expression encodes a slightly different meaning from *put (someone) at risk* (see Concordance 6.8). Although, like *put at risk*, *take the risk* is also used to realize the HARM sub-frame of the RISK frame, and so involves an individual making a choice, the left-hand collocates indicate that *take the risk* encodes modal meaning. There are also explicitly evaluative items among the left-hand collocates. Several of the examples in Concordance 6.8 encode a meaning of willingness or unwillingness to do something, for example: *cannot/can't take the risk; (not) prepared to take the risk; unwilling to take the risk; didn't/don't want to take the risk.* Deontic meaning is expressed, too: *should take the risk; better to take the risk; forced to take the risk.* *Take the risk* is therefore typically used in order to justify the decision not to vaccinate.

| N | Concordance |
|---|---|
| 1 | ess and disability rather than blaming parents who refuse to salute smartly and take the risk of watching the child they love become one more victim of vaccine |
| 2 | I may have to take him for something that would regress him worse. I cannot take the risk of losing him to measles anyone wishing to email me can do so at |
| 3 | ccinated after reading all the stories about vaccinations in general but i cant take the risk of loosing him to measles. To say I am distrought is an understate |
| 4 | now 15 months old and I have decided after weighing up evidence that I cannot take the risk of MMR..Ive heard too much, and I hardly think Wakefield would |
| 5 | I may have to take him for something that would regress him worse. I cannot take the risk of losing him to measles anyone wishing to email me can do so at |
| 6 | and I am not prepared to gamble further with my son's health. A doctor doesn't take the risk - they only write up the prescription. This reminds me of one youn |
| 7 | have also been vaccine damaged, so of course we will say - don't do it - don't take the risk. There are also a few posters who evidently do not have vaccine da |
| 8 | have also been vaccine damaged, so of course we will say - don't do it - don't take the risk. There are also a few posters who evidently do not have vaccine da |
| 9 | onsequence of family members not making it to the toilet in time. Please don't take the risk with e.coli. Please don't take this the wrong way, but "may be con |
| 10 | the MMR JAB. Our youngest is at risk from measles etc and we will not take the risk of injecting him with this unsafe vacine. seperate jabs should now |
| 11 | plexed as to whether it is actually better to take the risk of vaccinations, or take the risk of the baby contracting a disease. I am having so much difficulty |
| 12 | you know which vits activate which enzymes? And do you really think I should take the risk of taking foreign, synthetic and potentially dangermouse |
| 13 | ly requires vaccinations, I am perplexed as to whether it is actually better to take the risk of vaccinations, or take the risk of the baby contracting a diseas |
| 14 | looked into the research behind the jab and were so alarmed they decided to take the risk and let child go vaccine free. 1 set of parents who didnt get chil |
| 15 | you. But what I won\'t do is put up with accusations and abuse for deciding to take the risk and vaccinate my child. Help! Cybertiger |
| 16 | of 3 But what I won\'t do is put up with accusations and abuse for deciding to take the risk and vaccinate my child. Help! You're getting your fun schmooz |
| 17 | of 3 But what I won\'t do is put up with accusations and abuse for deciding to take the risk and vaccinate my child. Help! You're getting your fun schmooz |
| 18 | but as when my daughter was immunised as a parent it is so difficult to take the risk of a major infection Baarbara Tompkinson jen |
| 19 | would be so unethical and amoral as to suggest one child should be forced to take the risk of a vaccine in order to protect another for the theory of herd im |
| 20 | y fuss about then...my kid is developing normal right now, I sure would hate to take the risk of regression into the abyss"! http://ije.oxfordjournals.org/cgi/ |
| 21 | y fuss about then...my kid is developing normal right now, I sure would hate to take the risk of regression into the abyss"! http://ije.oxfordjournals.org/cgi/ |
| 22 | vaccines for others and be cavalier with safety when you dont have to take the risk of one yourself Vaccinations cause chronic diseases is as fully a |
| 23 | ther is that as good citizens and "scientific believers" we must be prepared to take the risk for the greater good. Of course, part one is never really demonst |
| 24 | ! I have an autistic cousin and I am dyslexic myself and we are not prepared to take the risk of MMR. Message received on 22/02/02 from: [redacted] |
| 25 | from: [redacted] : Like so many of you, I am not prepared to take the risk of the MMR jab for my 22 months-old daughter. Can anybody tell |
| 26 | and hypocrites who are scared of your own medicine and who are unwilling to take the risk you expect every child to run. Talk is cheap. Let's see some actio |
| 27 | and hypocrites who are scared of your own medicine and who are unwilling to take the risk you expect every child to run. Talk is cheap. Let's see some actio |
| 28 | new bug which my immune system wasn't dealing with - and I didn't want to take the risk. I really do despair about the state of our healthcare system. 'H |
| 29 | great hardship and poverty, one can appreciate why many don't want to take the risk, simply on the say so of foreigners. From the Telegraph 18/04/200 |
| 30 | ay). She had indivdual ones done privately as a little one as we didn't want to take the risk with the combined jab. Our other daughter had her rubella jab las |
| 31 | are diagnosed after MMR with autism, Crohns disease, ADHD, but why take the risk? I have a friend whose son was perfectly normal and healthy |
| 32 | isease is rare and early treatment for this rare disease is very effective. Why take the risk of a vaccine for a rare disease that you are highly unlikely to ev |
| 33 | d decided to do about this chicken leucosis virus (ALV) contamination. It would take the risk of quietly allowing MMR to continue to be contaminated. It would p |
| 34 | d decided to do about this chicken leucosis virus (ALV) contamination. It would take the risk of quietly allowing MMR to continue to be contaminated. It would p |
| 35 | d decided to do about this chicken leucosis virus (ALV) contamination. It would take the risk of quietly allowing MMR vaccine production to continue in retrovir |
| 36 | ne hospital either, there were several involved. Nah, I'm sure doctors wouldn't take the risk of doing anything unless it was totally above board. |

**Concordance 6.8** Concordance lines for *take the risk* in the *JABS* corpus

*Run the risk* is far less frequent than *PUT (someone) at risk* or *take the risk* (see Concordance 6.9). Although *run the risk* realizes the CHANCE sub-frame, in which, in theory, the state of being at risk does not result from an action (Fillmore and Atkins, 1992), some of the contexts in which it is used in the JABS corpus indicate that there is overlap between *run the risk* and *take the risk*. For example, Line 2 (*will you dare run the risk of vaccinating*), Lines 14 and 15 (*does x really think I am going to run the risk of vaccinating*), Lines 16 and 17 (*decide not to run the risk*), and Line 21 (*seen one child descend into regressive autism following MMR and*

*won't run the risk of a second child following suit*) actually realize the HARM frame, in that they refer to undertaking an action when faced with a choice.

| N | Concordance |
|---|---|
| 1 | ic resistance, such as the hospital bacteria. All animals, plants and bacteria run the risk of being infected by specific viruses. For humans, such viruses inc |
| 2 | tinue to the present day. If you have one autistic child already will you dare run the risk of vaccinating child 2 or child 3? laura_c_a |
| 3 | h other data, revealed a danger that children could suffer convulsions and even run the risk of brain damage. This is devastating news for parents. Some health |
| 4 | h other data, revealed a danger that children could suffer convulsions and even run the risk of brain damage. This is devastating news for parents. Some health |
| 5 | h other data, revealed a danger that children could suffer convulsions and even run the risk of brain damage. This is devastating news for parents. Some health |
| 6 | nital rubella syndrome that it can cause. As an added bonus, she will also not run the risk of transmitting it to any children who have been as unfortunate as |
| 7 | nital rubella syndrome that it can cause. As an added bonus, she will also not run the risk of transmitting it to any children who have been as unfortunate as |
| 8 | enital rubella syndrome that it can cause. As an added bonus, she will also not run the risk of transmitting it to any children who have been as unfortunate as |
| 9 | all. It may be safer for healthy children to catch these illnesses rather than run the risk of immunisation. It's important that girls have either had rubella |
| 10 | 00 doses imported in that time as worried parents prefer to payout rather than run the risk of leaving their children unvaccinated - or let them undergo the M |
| 11 | e United States, Thailand, Belgium and France.' The current polio vaccines thus run the risk of having oncogenes in them. Again this was news to me. I had no id |
| 12 | e United States, Thailand, Belgium and France.' The current polio vaccines thus run the risk of having oncogenes in them. Again this was news to me. I had no id |
| 13 | autism and who, according to you, make the "wrong" conclusions, aren't about to run the risk of autism, in order to fully vaccinate their currently healthy chil |
| 14 | osted by Elizabeth Does this pea-brained nitwit really think that I'm going to run the risk of vaccinating my beautiful daughter with the so-called booster dos |
| 15 | for all concerned. Does this pea-brained nitwit really think that I'm going to run the risk of vaccinating my beautiful daughter with the so-called booster dos |
| 16 | nd risks and decide to give you the chance of a few more years or decide not to run the risk of your death messing up his/her statistics? Where will he/she draw |
| 17 | nd risks and decide to give you the chance of a few more years or decide not to run the risk of your death messing up his/her statistics? Where will he/she draw |
| 18 | next consultation. Whilst Sulphur 6C might have been adequate I did not want to run the risk of it failing to go the distance. His mother was instructed to give |
| 19 | rveillance Centrestated: We are below the critical threshold at which point we run the risk of getting alarge number of cases. We will have to reverse that tre |
| 20 | ven hundreds of thousand of people follow Ms. O'Connell's lack of reasoning, we run the risk of an epidemic like the one in the late 1980s. Refusing vaccination |
| 21 | ts who've seen one child descend into regressive autism following MMR and won't run the risk of a second child following suit. Come to think of it, why not make |
| 22 | ed, laboratory confirmation of measles will not have been sought, as this would run the risk of discovering that the children did have measles, despite vaccinat |
| 23 | from one species and put it into a another species or even another animal, you run the risk of unwanted pathogens that you didnâ¬"t know were there; thatâ¬"s |
| 24 | from one species and put it into a another species or even another animal, you run the risk of unwanted pathogens that you didnâ¬"t know were there; thatâ¬"s |

**Concordance 6.9** Concordance lines for *run the risk* in the *JABS* corpus

The expression *worth the risk* is typically used to justify the argument that vaccines are unsafe, or to justify a decision not to vaccinate, for example:

> [6.85] We now know too late that the MMR vaccination is **just not worth the risk**.
>
> [6.86] Ive seen the damage that the DTP can do, brain damage and phsyical problems too, **its just not worth the risk** IMO.

We see, in the above examples, that the force of the proposition is intensified through use of *just*.

6.5.4.3 Expressions with *risks*

It has already been observed that the most frequent two- to six-word cluster with *risks* after *the risks* is *outweigh the risks* <46>. This string was found to be included in longer strings, such as *(the) benefits outweigh the risks* <12>, and *far outweigh the risks* <9>. As might be predicted, instances of *(the) benefits (far) outweigh the risks* are usually attributed to official sources, such as spokespeople for the DoH or for pharmaceutical companies, for example:

> [6.87] Nicholas Kitchin, medical director of Sanofi Pasteur, insisted: "There are fewer side-effects with the five-in-one than were reported with the older four-in-one vaccine. "All drugs and vaccines have potential side-effects, which parents should be aware of and these are listed in the information leaflet included with every dose of the vaccine. **Severe reactions are rare and the benefits of vaccination far outweigh the risks**."

Often, intertextual references involving assertions that the benefits of vaccination outweigh the risks are framed in such a way that mistrust of the medical-scientific establishment is encoded in the meaning. In Example 6.88, commercial scientists are the cited source and their assurances are evaluated as untrustworthy because they overlook vital information:

> [6.88] Surely only these calculations can only be based on life long risk rather than than immediate, very short term or using a 30 day limit in many cases. Medium and long term side effects dont even come into the equation for **the commercial scientists who assure us the benefits outweigh the risks.**

Other expressions of mistrust are encoded in the contexts in which *real risks* <27> is used. When the terms *health risks* and *real risks* are used, the wider context reveals an implied or explicitly expressed belief that the authorities do not provide full information on the risks of vaccines:

[6.89] JABS is trying to support free choice and **full information on the real risks of vaccination and childhood diseases.**

[6.90] Loss of confidence in the dtp vaccine in the 80s and 90s lead to aggressive measures by the authorities to "restore public confidence" in vaccines. This has lead to **a complete denial that the mmr could be associated with any real risks**

[6.91] Jackie Fletcher, of the pressure group Jabs, which is trying to highlight the potential dangers, said: **"The Government should be giving people full and accurate information about health risks."**

[6.92] The discovery that the **Government ignored warnings that the MMR vaccine carried serious health risks** is shocking.

Expressions with *risks* in the *JABS* corpus trigger the presupposition that there are risks associated with vaccination. The contexts in which *risks* typically occurs in the *JABS* corpus indicate a preoccupation with assessing the relative risks and benefits of vaccination. This is evident not only in the expressions *the benefits (far) outweigh the risks*, but in the many examples of *risks and benefits* <37>. A typical example is:

[6.93] Mr Alison, Under-Secretary, said no immunisation procedure was entirely free from risk of ill-effects. **The balance of risks and benefits was kept continually under review.**

There is, however, uncertainty associated with the risks of vaccines, as is suggested by the relatively high frequency of expressions such as potential risks <16>, *possible risks* <14> and *known risks* <9>.

### 6.5.5 Summary of Section 6.5

There are noticeable differences between the ways in which risk is expressed in the *NHSvax* corpus and the *JABS* corpus. The sorts of expressions of degrees of assessment of risk typical of healthcare discourse are frequent in the *JABS* corpus data, but they tend to occur in texts

which have been imported onto the JABS site from external sources. Some technical terms which do not occur in the *NHSvax* data are found in the *JABS* corpus. These, too, are typically found in externally produced texts which are reproduced on the site, although sometimes high-frequency posters incorporate these terms into their own texts, usually as a way of performing expertise. The technical discourse of risk is used by high-frequency vaccine-critical posters in order to challenge the fundamental basis of the science. Challenges to science are often linked with expressions of mistrust of the medical-scientific authorities. Use of these terms by JABS posters is accompanied by a relatively high occurrence of signals of evaluation in the wider context. More frequently, JABS members express their concerns about the risk using terms such as *put (children/my child) at risk, take the risk, run the risk* and *worth the risk*. The contexts in which these terms are used hint at the tensions between the public good and the interests of the individual. In many contexts in which the terms such as *put (children/my child) at risk* and *take the risk* are used, the ideational content expressed can be interpreted as representing a form of resistance to state control. The expression *(not) worth the risk* is used to justify decisions not to vaccinate. *Risks* is most frequently used in the contexts of assessments of risks and benefits. Propositions expressed using the term *risks* are usually attributed to other speakers and are frequently framed in such a way as to suggest authorial distance.

## 6.6 Conclusion

This chapter has examined the ways in which claims about the causal hypothesis, health and the immune system, and risk are expressed in the *JABS* corpus, in order to test the claim that 'lay' people express beliefs about these issues in different ways from the medical-scientific community. Examination of the corpus found many similarities between the ways in which

propositions relating to the causal hypothesis and to risk are expressed in both corpora. For example, the expression most frequently used to make claims about the causal hypothesis is *link between mmr and autism*, and expressions used to rebut the hypothesis are particularly frequent, for example, *no evidence of a link between mmr and autism*. The sorts of expressions of risk which are typical of the domain of healthcare, such as *increased risk of*, and so on, occur with greater frequency than other expressions related to risk. However, all of these expressions typically occur in intertextual segments of text, most commonly, in news articles uploaded onto the site or copied into forum posts. When the individual voices of forum participants are identified, it is found that they are more likely to use explicit markers of causality, in particular, *CAUSE*, to talk about the causal hypothesis and to use expressions such as *PUT at risk*, *take the risk*, *run the risk*, and *worth the risk*, to express propositions about risk.

Propositions relating to the immune system have been found to be expressed in very similar ways to those which are typical of healthcare discourse. For example, in both the *JABS* and the *NHSvax* data, a distinction is drawn between natural immunity and vaccine-acquired immunity. The only difference here is that, in the vaccine-critical discourse that characterizes the *JABS* data, natural immunity is considered safer than acquiring immunity through vaccination. A notable difference between the two sets of data, though, is that, in the *NHSvax* data, there is a very narrow focus on one aspect of immune system function. In some of the *JABS* data, on the other hand, a distinction is drawn between humoral immunity and cellular immunity. What is evident in occurrences of *humoral immunity* and *cellular immunity* is that the contributors to the JABS site draw on particular discursive resources from the field of medical-science which are not drawn on in the *NHSvax* texts. These discursive resources are

exploited in order to support the argument that vaccination is potentially harmful. Similarly, references to past vaccination scares are used to justify mistrusting the authorities. This adds evidence to the theory that contemporary public discourse about health draws on and contributes to an interdiscursive memory bank.

As well as identifying similarities and differences between the discourse patterns of the *JABS* and *NHSvax* corpora, this chapter has uncovered some of the ways in which discourse from the medical-scientific domain, or from the domain of risk assessment, is reformulated and reframed in the JABS discussion forum. This aspect of interactivity and of the exploitation of intertextuality is further explored in Chapter Seven.

# CHAPTER SEVEN

# INTERACTIVITY AND INTERTEXTUALITY IN THE JABS CORPUS

## 7.1 Introduction

The overarching aim of this thesis is to help further understanding of the nature of scientific debate in the public sphere. The MMR debate is used as a case study in order to shed light on specific features of public debate about science and health. The main factor which inspired this project was the observation that an argument, which had very little support among the medical-scientific community, namely that there might be a causal connection between the MMR vaccine and a form of regressive autism and pervasive bowel disorder, had exerted a disproportionate influence in the public sphere. Research in some fields suggested that lay people hold different beliefs from health professionals about health, immunity and risk. One of the objectives of the project, therefore, was to discover the extent to which this is true. Another objective was to discover whether, and how, participants on the JABS site make strategic use of different discourses. Chapter Six contributed to achieving this objective by analysing the extent to which the lexico-grammatical patterns evident in the *JABS* corpus, which are used to express notions of health, immunity, and risk, are similar to and different from those observed in the *NHSvax* corpus. The purpose of the current chapter is to discover the ways in which contributors to the JABS website express their warrant for expertise. These issues were touched upon in the last chapter, in which the expression of

epistemological claims was examined. However, these phenomena form the main focus of this chapter.

Presenting knowledge is, of course, one way in which expertise may be performed. This chapter is therefore concerned with examining the ways in which claims become 'facts' and are reformulated and reframed as they travel from their source domain to another. However, warrants for expertise can also be seen in the way writers respond to other people's assertions. Being largely comprised of forum interaction, the *JABS* corpus displays many features which arise from the particular affordances of the forum platform. Forum participants interact with each other directly as well as recycling snatches of discourse from other sources. The discourse of the *JABS* corpus therefore shows the different types of manifest intertextuality characteristic of online forum interaction: the verbatim reproduction of texts or parts of texts from outside the corpus, the indirect reporting of utterances from outside the corpus, the verbatim reproduction of snatches of text produced by other posters, the reformulation and reframing of the words of other posters, and, at times, the reformulation of posters' own words (Richardson, 2001).

The discussion in the last chapter was informed by examination of keywords from the semantic/functional categories of 'vaccination and pharmaceutical products', 'health, diseases and medical conditions', 'likelihood, causation and temporal relations' and 'expressions with modal meaning' (see Appendix 8). The keywords which are subjected to close examination in this chapter are from the categories of 'people, places and institutions', 'science, medicine and healthcare', 'expressions of evaluation', 'nouns, verbs and phrases referring to events, actions and states of affairs' and 'grammatical items'. These keywords are shown in Table 7.1.

| | | |
|---|---|---|
| **people, places and institutions** | healthcare practitioners and scientific and medical experts | *researchers, scientists, experts, expert* |
| | public and private institutions | *government, nhs, pharma, pharmaceutical* |
| **science, medicine and healthcare** | | *science* |
| **expressions of evaluation** | Evaluative expressions used to refer to vaccines | *safety, safe, protect, protection, prevention, prevent, effective, benefits* |
| | status nouns: sign | *evidence, proof* |
| | status nouns: factivity | *truth, problems, facts, issue, opinion, conclusion, fact, problem, issues* |
| | status nouns: affect | *concerns, fear, concern, hope* |
| **other nouns, verbs, and phrases referring to events, actions, and states of affairs** | verbal processes | *reported, comments, reporting, quote, comment, claims, stated, according (to), states, mention, debate, suggest, statement, conclusion, claim, concluded, agree, explain, saying, mentioned, say, told, asking, suggests, ask* |
| | recounting | *story, stories* |
| | perception and cognition | *believe, wonder, known, note, trust, understand, know, noted, knows, (BE) aware, considered, hear, consider, doubt, watch, believed* |
| | seeming | *seems, appears, seem* |
| | showing | *prove, showed, revealed, showing, shows* |
| | discovering | *found, find, finding, identified, discovered* |
| | occurring | *occur, occurred, happened, happen* |
| | death | *died, die* |
| **grammatical items** | | *i, my, we, our, that, it, what* |

**Table 7.1** *JABS* corpus keywords identified for closer examination in Chapter Seven

In some instances, keywords which were examined in the previous chapter are of relevance to the current discussion and aspects of their use inform the discussion. These are *vaccine, damaged, autistic* and *healthy*, since expressions such as *vaccine-damaged child* or *previously healthy child* often occur in contexts in which writers express a warrant for expertise. Furthermore, in order to gain a fuller picture of the statistically most significant uses of status nouns and reporting verbs and nouns in the *JABS* corpus a list was drawn up of the collocates of *that* with a z-score of at least 4.0. Z-score was used as the measure of significance, since, when high frequency words are examined, it is more effective than log-likelihood or MI3 in suppressing grammatical collocates and bringing salient lexical collocates to the fore. All the

collocates of *that* which correspond to verbs and nouns which can be used to frame reporting clauses or can be used as status nouns were identified. These are shown in Appendix 13. A similar list of collocates of *that* in the *NHSvax* corpus was drawn up in order to provide a comparison. These are also shown in Appendix 13.

As with Chapters Five and Six, the discussion which follows does not focus on all of the keywords above separately, instead, the chapter is divided according to themes. The rest of the chapter is divided as follows. Section 7.2 examines the use of attribution. Section 7.3 investigates asserting identity as a form of warranting. Section 7.4 looks at warranting by presenting information, focussing on information structure, in particular on the use of expressions such as *the fact that*, cleft and pseudo-cleft constructions. The use of *evidence that* and *proof that* is also examined. Section 7.5 concludes the chapter.

## 7.2 Attribution

### 7.2.1 Attribution through use of terms characteristic of scientific discourse

#### 7.2.1.1 Features of scientific discourse

The *JABS* data shows a rich mix of lexical features typical of spoken discourse, journalistic discourse and (popular) scientific discourse. A number of words which are shown in Table 7.1 above or which were identified among the collocates of *that* (see Appendix 13) are terms which might be considered characteristic of scientific discourse. These are words such as *conclusion, FIND, REPORT, SHOW* and *SUGGEST*. In order to find out the extent to which this is true, we can compare the relative frequencies of these terms in the *JABS* corpus with those in the *New Scientist* sub-corpus of the *Bank of English*, which represents popularized scientific discourse, in the *Times* sub-corpus, which represents journalistic discourse, and the

*British Spoken* sub-corpus, which will highlight the extent to which particular features are common, or not, in spoken English. The relative frequencies per million words of text of the terms *conclusion*, *FIND, REPORT, SHOW* and *SUGGEST* in the *JABS* corpus and the *New Scientist, Times* and *British Spoken* sub-corpora of the *Bank of English* are shown in Table 7.2 below. Relative frequencies of the terms with and without *that* are shown.

| **Word** | **JABS** | **New Sci** | **Times** | **Br Spok** | **Word + *that*** | **JABS** | **New Sci** | **Times** | **Br Spok** |
|---|---|---|---|---|---|---|---|---|---|
| *conclusion* | 123.4 | 45.5 | 32.9 | 9.5 | *conclusion that* | 19.3 | 10.1 | 7.2 | 2.2 |
| *find* | 609.1 | 486.1 | 369.7 | 486.4 | *find that* | 35.1 | 34.2 | 21.4 | 63.8 |
| *finding* | 131.3 | 130.6 | 64.8 | 42.5 | *finding that* | 10.3 | 8.9 | 3.2 | 2.9 |
| *found* | 924.2 | 909.1 | 401.8 | 211.2 | *found that* | 154.5 | 290.8 | 46.7 | 31.7 |
| *report* | 575.3 | 357.4 | 319.0 | 68.6 | *report that* | 19.8 | 15.2 | 9.5 | 2.4 |
| *reported* | 601.2 | 137.4 | 120.7 | 11.9 | *reported that* | 57.3 | 24.3 | 11.9 | 0.6 |
| *reporting* | 162.7 | 18.5 | 33.6 | 8.0 | *reporting that* | 4.2 | 1.1 | 0.8 | 0.4 |
| *reports* | 601.2 | 131.7 | 114.4 | 24.6 | *reports that* | 22.7 | 15.6 | 17.4 | 1.3 |
| *show* | 310.4 | 287.8 | 327.5 | 171.2 | *show that* | 61.8 | 104.1 | 37.4 | 8.7 |
| *showed* | 180.1 | 152.9 | 110.8 | 25.4 | *showed that* | 50.2 | 81.4 | 27.1 | 1.8 |
| *showing* | 111.5 | 67.9 | 61.0 | 24.4 | *showing that* | 20.6 | 19.9 | 5.9 | 1.2 |
| *shows* | 177.8 | 146.7 | 128.8 | 32.4 | *shows that* | 54.1 | 56.0 | 26.1 | 3.8 |
| *suggest* | 202.1 | 160.1 | 74.0 | 32.2 | *suggest that* | 103.0 | 104.9 | 36.4 | 10.7 |
| *suggests* | 104.1 | 209.1 | 65.1 | 7.2 | *suggests that* | 53.6 | 130.7 | 29.9 | 3.5 |

**Table 7.2** Frequencies per million words of text of reporting verbs and nouns (with and without *that*) characteristic of (popular) scientific discourse in *JABS* and in the *New Scientist*, *Times* and *British Spoken* sub-corpora of the *Bank of English*

The relative frequencies of many of the expressions shown in Table 7.2 indicate, firstly, that most of the terms are used significantly more frequently in (popularized) scientific discourse than in journalism or spoken English. This is true particularly when the expressions are followed by *that*. The two exceptions are *conclusion that* and *reports that*, whose frequency of use in the *New Scientist* and *Times* sub-corpora is relatively similar. Secondly, it is apparent that there are marked similarities between the frequencies of use of most of the terms

above in the *JABS* and *New Scientist* corpora, especially when followed by *that*. For example, the relative frequencies for *find that* are 35.1/million and 34.2/million, respectively, in *JABS* and *New Scientist*. The relative scores for *report that* are similarly close: 19.8/million in *JABS* and 8.9/million in *New Scientist*; as are the relative scores for *suggest that*: 103.0/million in *JABS* and 104.9/million in *New Scientist*. Interestingly, the expressions *conclusion that*, *reported that,* and *reports that*, are significantly more frequent in the *JABS* corpus than in the *New Scientist*.

#### 7.2.1.2 The uses of *REPORT that,* and *conclusion that* in the *JABS* corpus

The high frequency of *reported that* and *reports that* in the *JABS* corpus is due to the mix in the corpus of scientific discourse, news discourse, and the vaccine-critical discourse. Several instances come from a review of scientific articles on the MMR vaccine written by a prominent vaccine-critical campaigner, which is uploaded onto the Guest Writer section of the JABS website. The following two examples are from this review. Reflecting Myers' (2003) observation that lay people often acquire extensive technical expertise by learning about subjects in which they are interested, the author is clearly conversant with the discursive patterns of scientific texts. The stretches of text represented here are part of a summary of the study by Wakefield et al. (1998).

> [7.1] **Following reports that measles virus might be present in the intestines of children with Crohn's Disease**, a new syndrome was reported in children with autism who exhibited developmental regression and gastrointestinal symptoms (autistic enterocolitis), in some cases after MMR vaccine, was reported (see papers by Wakefield et al). It was not known whether the virus, if confirmed as present in these patients, derived from wild strain or vaccine strain.
>
> [7.2] **The paper reported that**: Three children with regressive autism (autistic encepalopathy) underwent cerebrospinal fluid assessment, including studies for measles virus. All three children hadconcomitant onset of gastrointestinal symptoms and had

> already had measles virusgenomic RNA detected in biopsies of ileal-lymphoid nodular hyperplasia

Many other instances of *reported that* and *reports that* are similar, in that they represent either reproductions or reformulations of scientific reports. The following example is interesting. It comes from a stretch of text authored by Andrew Wakefield in response to a scientific report claiming to have found no evidence to support his hypothesis. Wakefield's response is included in an article in the *Spectator* magazine. The article has been copied in its entirety into a JABS forum post:

> [7.3] There is no evidence in the Baird et al. paper that these crucial factors were taken into account. This study's inappropriate symptom criteria would explain the discordance with **other reports that have revealed a high prevalence of significant gastrointestinal symptoms in general autism populations**

Examples such as those above illustrate how scientific claims circulate in the public sphere and how they are reframed and reformulated. They also illustrate the complexity of intertextual relations evident in the *JABS* data. Examples 7.1 and 7.2 show reformulations of parts of the paper by Wakefield *et al.* (1998). This is done through summary. Example 7.3 is a little more complex in terms of the intertextual relations it exhibits. The author makes explicit reference to a text which, in turn, has made explicit reference to a previous text the current writer co-authored. He reformulates the ideational content expressed in the paper by Baird et al. in an interesting way. Wakefield evaluates the paper negatively, on the grounds that it has overlooked factors which he evaluates as 'crucial'. He then justifies his evaluation on the grounds that he believes the paper's symptom criteria are inappropriate. However, he does not express his opinion as if it were an opinion, but as fact. By using the expression *this study's*

*inappropriate symptom criteria*, he presupposes that the criteria are inappropriate and thus closes down the negotiability of the proposition. He then expresses the notion that the paper by Baird et al. has reached different conclusions from other papers. However, he joins the two clauses with an expression which signals a cause-effect relation: *would explain*. The proposition expressed in the second clause which purportedly offers an explanation for the proposition expressed in the first is also realized through use of a presupposition: the expression *the discordance*. He signals his alignment with the truth of the findings in the 'other papers' by using the expression *reported that*. As mentioned above, Wakefield's response is reproduced in a journalistic article, which is then reproduced in full in a forum post. The forum participant thus uses the article as a discursive resource with which to present his views.

Concordance evidence for *conclusion that* reveals that it most frequently occurs in the expression *COME to the conclusion that*. It is pre-modified in only eighteen of its seventy-three occurrences. When it is pre-modified, the only expressions which recur are the *inevitable conclusion* <3>, *the unproven conclusion* <2>, *the obvious conclusion* <2>. The expression *conclusion that* is also typically used in reports about science stories. The following example originates in an online news publication and the expression *conclusion that* occurs in a stretch of cited text:

> [7.4] ROSEMEAD, Calif., Oct. 15 (A Scribe Newswire) -- The Centers for Disease Control and Prevention (CDC) was recently notified of an individual infected with a monkey cytomegalovirus. Previous reports to CDC of infections with viruses that originated from African green monkeys have largely gone ignored, according to Dr. W. John Martin, M.D., Ph.D., founder of the privately funded Centers for Complex Infectious Diseases in Rosemead, California. "Public health officials have been very resistant, since **any airing of this topic will lead to the inevitable conclusion that the government erred in accepting cytomegalovirus contamination of the African green**

**monkeys used in polio vaccine production**."

*Conclusion that* also often occurs in texts authored by prominent vaccine-critical writers. The following example comes from one such a text and, as we saw in Example 7.1 and 7.2 above, it summarizes an aspect of a scientific report.

> [7.5] **The conclusion that Fombonne drew** was that these data provide no support for the hypothesis of an association between IBD and autism.

Occurrences of *REPORT that* and *conclusion that* in the *JABS* corpus, therefore, result largely from intertextual references in the text. However, evaluative language can be seen in the extended contexts of use.

### 7.2.2 Attribution using *CLAIM*

Predictably, *CLAIM* is used to express authorial distance:

> [7.6] **To ease parental fears and counter claims that** children are being over-vaccinated, the number of jabs against meningitis C, which is already part of the childhood vaccination programme, is to be reduced.
>
> [7.7] **the drug companies claim that** the MMR vaccine is safe and does not cause Autism, IBD, Eplipsy, Deafness, artritis and worst of all death, yet Mercks own material info sheets list all of these reactions.

In the following example, *CLAIM* is used twice. Firstly, it is used to reformulate and challenge an argument which a previous poster made. Secondly, it is used to attribute a hypothetical argument to un-named others. The two arguments are represented as equivalent.

The question *would anyone claim that* invites the answer 'no,' thus generating the implicature that the first claim is wrong-headed.

> [7.8] **Someone in the debate claimed that** diphtheria has nothing to do with poor sanitation. This is silly. **Would anyone claim that** washing hands doesn't reduce the spread of influenza?

The use of *CLAIM* in the *JABS* corpus is thus largely unremarkable.

### 7.2.3 Attribution using language of a formal register

One particularly striking feature of the *JABS* corpus is the relatively high frequency of forms of the lemma *STATE*. The use of *STATE* is often thought of in terms of a register choice. *STATE* is associated with formal contexts, especially with written language. The explanation of the use of the verb *STATE* in the *Collins COBUILD English Dictionary* (Sinclair *et al.*, 1995) is as follows:

> If you **state** something, you say or write it in a formal or definite way.

If we look at Table 7.3 below, we see that it is certainly associated more with written than spoken modes, in that it is relatively far less frequent in the *British Spoken* sub-corpus than in the *Times* or *New Scientist*.

| **Word** | **JABS** | **New Sci** | **Times** | **Br Spok** | **Word + *that*** | **JABS** | **New Sci** | **Times** | **Br Spok** |
|---|---|---|---|---|---|---|---|---|---|
| *state* | 367.1 | 332.1 | 271.3 | 68.5 | *state that* | 23.0 | 7.9 | 5.4 | 2.0 |
| *stated* | 178.8 | 21.3 | 22.1 | 5.0 | *stated that* | 67.4 | 6.2 | 8.1 | 0.9 |
| *statement* | 204.7 | 62.6 | 75.9 | 27.7 | *statement that* | 19.5 | 8.1 | 4.1 | 1.8 |
| *states* | 346.8 | 145.8 | 140.7 | 30.5 | *states that* | 42.8 | 21.5 | 7.2 | 0.9 |
| *stating* | 43.3 | 7.6 | 6.0 | 1.6 | *stating that* | 18.5 | 4.2 | 3.0 | 0.8 |

**Table 7.3** Frequencies per million words of text of verb and noun forms of the lemma *STATE* (with and without *that*) in the *JABS* corpus, and in the *New Scientist, Times* and *British Spoken* sub-corpora of the *Bank of English*

It is surprising, though, to what extent the relative frequencies of the various forms of *STATE* (including statement) and *STATE* that in *JABS* outrank those of the *New Scientist, Times* or *British Spoken* corpora. The frequency of *state that* in the *JABS* corpus, for example, is 23.0/million, as opposed to 7.9/million in the *New Scientist* and 5.4/million in the *Times*. The difference between the relative frequencies of *stated that* in *JABS* and in the *New Scientist* and *Times* is even greater: 67.4/million in *JABS* and 6.2/million and 8.1/million respectively in the other two. A major factor accounting for the particularly high frequency of *STATE* is to do with the judicial, or quasi-judicial, nature of some of the discourse found in texts in the *JABS* corpus relating to vaccine damage claims or to Dr Andrew Wakefield's disciplinary hearing. The following example comes from an article reproduced in a forum post entitled 'Why Vaccine-Injured Kids are Rarely Compensated' from the 'alternative' medical and vaccine-critical website Mercola:

> [7.9] Twenty-two years later, on Nov. 18, 2008, **I made a statement to the Advisory Commission on Childhood Vaccines (ACCV)** and questioned whether the compensation program is fatally flawed and so broken that it should be repealed.

The following example comes from a report by an investigating officer on a complaint to the ombudsman against the pharmaceutical company, Glaxo Smith Kline:

> [7.10] In his complaint to the Ombudsman, Mr XXXX highlighted his first concern as relating to a perceived inconsistency between a press release issued by the Judicial Communications Office in May 2007 in which it was said that "at the date of the hearing before Davis J, the possibility of any conflict of interest arising from his brother's position was not raised with him and did not occur to him" and the OJC's dismissal letter of 16 October (AF3), which said that **Davis J had stated "categorically that he was not aware of this at the time of the hearing in February 2004**".

Obviously, a lot of the instances of *STATE* in the *JABS* corpus originate in media texts in which the words of, for example, spokespeople for public institutions and companies are reported:

> [7.11] **The company said in a statement that** an adverse event report "does not mean that a causal relationship between an event and vaccination has been established -- just that the event occurred after vaccination."
>
> [7.12] **Government scientists, in the Lancet of the 4th March, stated:** "The estimated absolute risk of 1 in 24,000 was 5 times that ... reported by clinicians"

The choice of the word *STATE*, though, may indicate more than a register-specific choice. Because of its connotations of formality and the fact that its use signals that the writer's stance is aligned with the proposition expressed in the cited text, *STATE* can be used to intensify the writer's expression of commitment to the truth. It may encode other aspects of the writer's attitude. In the following example, an extract from a book by a well-known vaccine-critical writer, use of the word *stated* rather than *said* expresses the idea that the manufacturer was somewhat brazen in its claims. This is, of course, reinforced by use of the expression *stated frankly* in the third sentence:

> [7.13] I went to the published reports of the MMR manufacturers and found these confirmed what the scientists at this workshop had reported. **A manufacturer stated in 2000 that** it made the MMR vaccine with "harvested virus fluids". **It stated frankly that** their "Measles vaccine bulk is an unpurified product whose potency was measured through a biological assay for the active substance rather than through evaluation of integrity of physical form. Degradation products are neither identified nor quantified."

Interestingly, *STATE* is also used in the JABS corpus by forum posters. It not only serves a cohesive function, introducing a reformulation of an earlier segment of text, it also carries evaluative meaning. The formality of the register, which is at odds with the relative

informality of much of the participant-to-participant interaction on the forum, flouts the maxim of manner and thus creates an implicature: irony can be inferred.

> [7.14] **Earlier today you stated that** 99.99% of doctors and scientists * believed that AIDS was caused by HIV. Do please provide the evidence.
>
> [7.15] **So please support your statement that** there are 1000 (or is it 40) simian viruses we need worry about with a reference to the scientific literature and **please just give me your view on how the epidemiological data fits with your statement** "There are no positive benefits to vaccination as far as I am concerned, especially not with polio vaccination"

The markedly high occurrence of *STATE* in the *JABS* corpus thus results mainly from the presence of texts from external sources on the website. There is evidence, though, of forum participants using *STATE* when reformulating prior sections of text authored by other participants in order to express a negative evaluation of the proposition expressed by the addressee.

### 7.2.4 Attribution using indirect means

#### 7.2.4.1 *CONCERN that* and *FEAR that*

Although terms such as *CONCERN that* and *FEAR that* can be used to attribute beliefs to people without specifying the referent, the referent is often recoverable from the surrounding context. Concordance evidence in *JABS* reveals that, while propositions prefaced by *CONCERN that* and *FEAR that* can be attributed to either the medical-scientific authorities or to lay people, the terms are not equally distributed across the two groups. Authorities are more likely to be described as expressing *concern*. Both authorities and lay people are described as having *concerns*. Lay people are more likely than authorities to have *fears*. *Fear* is frequently used as a verb with *I* as the Subject. The examples below illustrate typical uses of *concern*, *concerns*, *fear* and *fears*.

[7.16] **The Nursing Times this week expressed concern that** the government's plans to introduce s more multiple vaccines to immunise for many illnesses with one injection.

[7.17] **The results of exploratory subgroup analyses for study 013 suggested a concern that** subjects who were seropositive and PCR-positive for the vaccine-relevant HPV types had a greater number of CIN 2/3 or worse cases

[7.18] I think it's also healthy to question things and we would want to **Encourage parents to discuss any concerns that they have.**

[7.19] There were **two important concerns that were identified during the course of the efficacy review of this BLA [biologics license application**].

[7.20] **I fear that** if Joe is not vaccinated he may catch Measles and suffer hearing loss.

[7.21] **I fear that** even if an MMR-autism link is established, the vaccine manufacturers can not be sued.

[7.22] **Some parents** have refused to allow their child to be vaccinated **because of fears that** MMR could cause a form of autism

[7.23] Some religious groups refuse vaccination but **many parents have fears that** vaccines are unsafe or may cause conditions such as autism

#### 7.2.4.2 Reference to science

Taylor (2010) reports an increase in media reports about science issues of attribution using terms such as the *science, scientists, and expert(s).* These words have been identified as salient words in the *JABS* corpus. When *the science* is examined in the *JABS* corpus however, it is found that it is not so much used as an attributed source, but as a field of activity, or as a pre-modifier in an expression used to identify groups of people. The term *the science* is never followed by a reporting verb, for instance. The most frequent meaningful three- to six-word clusters with *the science* are: *science media centre* <20>, *about the science* <11>, *the science lobby groups* <8>, *the science of vaccination* <7>, *all the science* <6>, *understand the science* <6>, and *to do with the science* <5>.

Some of the contexts of use of *about the science* express the notion that the science of vaccination is uncertain and perhaps flawed, for example:

> [7.24] However, since its appearance, the study has sparked debate and criticism in the child neurology community, **with some experts expressing concerns about the science itself.**

There are three repeated instances of the following example, which shows how expressing that one has knowledge about the science of vaccination can be used to claim expertise:

> [7.25] And yes, **I probably do understand more about the science behind vaccines** and the side-effects that medicines can cause then almost every judge in the world.

The expression *the science lobby groups* is used to express mistrust. The expression itself triggers a presupposition that such groups exist. In many contexts in which the term is used, *the science lobby groups* appears as part of a compound nominal group. In the following example, they are classed together with the government and the medical establishment. In the subsequent example, they are classed with the General Medical Council (GMC) and Brian Deer, a journalist who investigated Andrew Wakefield.

> [7.26] **The other tactic that the government, the medical establishment and the science lobby groups have used** is to suggest that Wakefield and his supporters have actually said: “Autism is caused by vaccination"

> [7.27] **The GMC and the science lobby groups together with Brian Deer who is the sole complainant in the GMC fitness-to-practice case**, have studiously avoided all reference to the parents and damaged children.

In three out of the seven instances in which the expression *the science of vaccination* occurs, it is in the wider context of *demonstrate weaknesses with the science of vaccination*:

> [7.28] Occam48 is a simpleton. He restricts the sphere of rationally demonstrable truth, **denying, for instance, that reason can demonstrate weaknesses with the science of vaccination**.

Meanwhile, four of the five occurrences of *to do with the science* come from the same forum thread and are repeats of part of one post. The context is:

> [7.29] <q> Big pharma murders 50,000 with Vioxx using fraudulent science</q> **Vioxx has nothing to do with the science of vaccination.**

The corpus tags <q> and </q> indicate the boundaries of the quoted piece in this post. The assertion *Vioxx has nothing to do with the science of vaccination* is made by a high-frequency pro-science poster. As one might expect, pro-science posters are particularly likely to make appeals to science. In all, the contexts of *the science* demonstrate that scientists and people who support science are mistrusted. However, vaccine-critical people are almost as likely as pro-science posters to appeal to *the science*. Instead, when appeal is made to science in the *JABS* corpus, the terms *scientists* <918> or *experts* <801 occurrences> are preferred.

### 7.2.5 Using attribution to challenge the basis of an argument

Attribution can be used not to report what someone has said but in order to delegitimize a person's argument, either on the basis that they have withheld information or by questioning their integrity. One particular kind of proposition which recurs in the data is the proposition that someone has failed to say something. The phraseologies most associated with this are:

(*there BE*) *no mention of* <64>
*DO not/ don't/ doesn't/ didn't* (ADVERB) *mention* <44>
*BE* (ADVERB) *not/never* (ADVERB) *mentioned* <35>
*FAIL/forgot to* (ADVERB) *mention/ MAKE mention* <33>
*(has) not/never mentioned* <20>
*MAKE no mention of* <9>

This phraseological pattern is used in two distinct ways. It is most often used to rebut an argument by pointing out a flaw in the reasoning. At the same time, the author introduces new information that he or she deems significant. By so doing, the writer signals that the source of the original argument is untrustworthy and establishes his or her own expertise as the provider of better quality information. We can see this in the examples below.

> [7.30] The paper stated that triple vaccines were desirable to simplify administration, reduce costs and minimise visits (my emphasis). **There was no mention of** greater effectiveness, or inherent drawbacks with single vaccines.
>
> [7.31] The NEJM essay **does not mention** the dubious aluminium and oil-based adjuvants which Rosemary has just highlighted
>
> [7.32] The British media **has singularly failed to make mention in any national coverage of note** the fact that the vaccine/autism link has now been conceded.
>
> [7.33] The CDC says that multiple simultaneous vaccines are safe, "for children with normal immune systems," but **makes no mention of** the risk for everyone else.

In a smaller but no less interesting proportion of instances, the proposition that someone has not mentioned something is used to counter an argument by questioning the motives or the integrity of the source. This reflects the findings of Vayreda and Antaki (2011), who observed that suggesting that a person has links to a pharmaceutical company is one way in which pro-vaccine arguments are countered on a Usenet thread about vaccination. The individual is delegitimized.

> [7.34] Funny though that in this rush to embrace the US system **there is no mention of** replication their multi-million dollar compensation scheme here.
>
> [7.35] His sponsorship by GSK **was certainly never mentioned** by him in the Guardian or the BMJ, and though he listed the ABSW award on his Bad Science website, the GSK connection **was not mentioned**.

### 7.2.6 Self-attribution

Forum posters often need to reiterate things they have said earlier in their posts. They do this to emphasize a point they have made, either to correct the misapprehension of another, or to repeat the point in the light of new evidence which supports the point. The most frequent expressions used to introduce self-reiteration are *as/like I* (*have*) *said* (*before/earlier*) and, to a lesser extent, *what I (actually) said* (*was*), *when/where I said* (*that*) and *I never said* (*that*). The expression *as/like I* (*have*) *said* (*before/earlier*) serves merely to highlight the point being made.

> [7.36] **As I said** in my reply if the child does not give an immune response, because they are not immunocompetent, the vaccination will have no effect
>
> [7.37] Check this bit out, **as i said** they determined to use non science emotional ploys to get vaccination compliance as they can no longer use open and honest science.
>
> [7.38] I do not object to facts and **have never said that I do**. **I have said several times** that I object to rudeness in the manner of stating those facts.
>
> [7.39] **As I said before** Cathy, talk to a health professional you trust - not a conspiracy theorist.
>
> [7.40] they'll never run out of excuses for avoiding looking at the unvaccinated although, **as I said earlier**, it should be very good for their side of the argument as the unvaccinated should be constantly ill, catching all of the diseases that the vaccinated children are safe from

The expressions *what/when/where I said* (*that*) and *I never said* (*that*) are used not only to reiterate a point but to imply that another poster has misinterpreted the writer's words. The reiteration may involve verbatim repetition (Example 7.41) or it may involve a reformulation in the form of reported speech (Examples 7.42 to 7.44)

> [7.41] John, I'm sorry, **this is what I said**: "It says nothing about what the implications are for the safety of vaccines and vaccinations. Any interpolations you make are subjective and unreliable"

[7.42] **What I said was that** If these researchers were really interested in finding out if children with autism have a 'leaky gut' then surely they would firstly need to test if the children's intestinal lining is damaged in order for these 'peptides' to cause a problem?

[7.43] I was being kind **when I said Myths** I should just refer to them as lies as I always do

[7.44] **I never said** vitamins cure all things, or that zappers would cure all things, zappers are useful for some kinds of pathogens thats all and i avoid synthetic vitamins in favour of natural vitamins from real food.

### 7.2.7 Summary of Section 7.2

Certain expressions which are characteristic of the ways in which attribution is performed in scientific discourse are frequent in the *JABS* data. These expressions tend to occur as reformulations of claims which have been made in a scientific genre. These may be summaries of arguments, or they may display more complex forms of intertextuality and reformulation. Often the claims are reformulated in media texts, or in texts by prolific vaccine-critical writers. Reformulations of this kind which make extensive use of scientific discourse are typically authored by writers with a considerable degree of technical expertise. These texts are imported in their entirety onto the JABS site, often in forum posts, where they are used as a ready-made argument. Terms used for the purposes of attribution and which suggest a formal register, such as *STATE*, are frequent in the *JABS* corpus data on account of the heavy reliance on mass media texts, which is typical of online forums. However, these terms are sometimes used by forum posters as a way of intensifying the force of their propositions. Non-specific terms which signal attribution and which are significant in the corpus are *concerns* and *fears*. Typically, concerns are attributed to public bodies while fears are attributed to members of the public. Individuals who post on the forum may refer to themselves as having fears. People often refer to people in authority as having failed to mention something important or relevant. This is a not uncommon strategy for undermining

the credibility of the source. As is common on online forums, people may self-refer, usually to emphasize a point or to correct a misapprehension.

## 7.3 Expressing warrants for expertise

### 7.3.1 Warranting by asserting identity

On the JABS forum and message board, expressing a warrant for raising a topic or for claiming expertise is closely tied in with identity. The identity most frequently asserted is that of parent:

> [7.45] Hi. **My son is 11 months old** and I am concerned about the MMR jab.
>
> [7.46] **My daughter had the MMR jab a day before her 1st b/day**. A couple of days later she was covered in a rash, had a very high temperature and cried for hours.
>
> [7.47] From a slightly different point of view: **one of my children have had the Pediacel 5 in 1 vaccine with no ill effects.**

Some posters identify themselves as parents of an autistic or vaccine damaged child. This identity confers the poster an enhanced status on the forum. Asserting one's identity as the parent of a vaccine damaged child is sometimes used as a warrant for delegitimizing the claims of others:

> [7.48] **Is is so hard to understand that as a parent of a vaccine damaged child who uses a support forum for parents of vaccine damaged children that I take offense** at the sort of post in question.
>
> [7.49] **As a parent of a vaccine damaged child, I take issue with the idea** that this forum, and its users, are fair game to any troll, dick or harry who has nothing better to do than insult us by denying the reality of our painful experiences.

For posters who do not have an autistic or vaccine-damaged child, one way of claiming in-group identity is to claim sympathy with parents who do:

[7.50] **Show just one family with an autistic child** who believes Wakefield is guilty, for it is the damaged child who is the core and key to this issue not the well ones.

[7.51] **I think any parent with an autistic child** deserves a medal!

Meanwhile, an antagonist's argument may also be delegitimized by highlighting their status as an out-group member, belonging neither to the core in-group of parents of vaccine-damaged children nor to the wider in-group of sympathisers:

[7.52] **Dont forget this forum is for vaccine damaged kids getting truth and justice** not about bad science people rubbishing these parents claims who have not got vaccine damaged children and have no idea what that means.

[7.53] **Yes Becky just like thousands of parents whos children have been killed and maimed by vaccinations who you dont give a damn about on a forum that supports vaccine damaged children**. [...] You have shown your true colours on here Becky and **lack of compassion for vaccine damaged children** so you cant pretend you actually care now.

### 7.3.2 Warranting by recounting a health narrative

Presenting accounts of personal experience of vaccination, or of the experiences of others, constitutes an often used strategic device. As explained in Chapter Six, the justification for including *STORY, DIE, OCCUR* and *HAPPEN* among the candidate terms is that they are often used in the *JABS* corpus in contexts in which people recount their own experiences or the experiences of others. This is true to a certain extent. The word *story* is, of course, used in journalism. Many of the 1,226 citations of *story* are references to news stories or the titles of links to news articles. *Occurred* <438 occurrences> is frequently used to refer to events in general, rather than as part of a personal narrative, for example:

[7.54] **Several outbreaks of human monkeypox have occurred** since the virus was first isolated from humans in 1959 (Gipsen, 1976).

However, in all but a small proportion of occurrences, *died* <1,597 occurrences> and *happened* <785 occurrences> are used in contexts in which health narratives are recounted, and there are several instances in which *story* and *occurred* are used to introduce such accounts. More surprisingly, the word *healthy* (and, in some cases, *health*) is often used in this way.

Health narratives are used in three ways. A narrative about personal experience may be used as a warrant for raising a topic. For example:

> [7.55] **I am writing this to share my story** of the effects vaccinations have had on my children.

Alternatively, individuals may respond to another person's narrative as a warrant for joining a discussion. In the following example, the writer starts her post with an expression of sympathy after reading a previous poster's narrative. She then follows the expression of sympathy with a narrative about her own son.

> [7.56] Hi estherlou I hope oscar is now doing well **i read your story and it has worried me** you need to make sure his temp is kept down as it can lead to febrile convulsions ,**this is what happened to our son before he died at 18 mths** old there as been no answers yet but it seems the mmr lowered his immune system and then he had an acute convulsion in the night

More frequently, though, narratives about children who have died and whose stories have appeared in the press are used as a warrant for expressing the kind of expertise which comes from having access to knowledge. Typically, the narratives appear in news articles which are copied into forum posts. Typical examples are:

> [7.57] Christopher Coulter was 15 when **he suffered a fit and died** in his sleep 10 days after being vaccinated.

[7.58] **A girl of 14 died yesterday** hours after being given the cervical cancer vaccine at her school.

Frequently, in order to add force to the proposition that a vaccine has damaged or caused the death of a child, the child's state of health prior to the vaccination is referred to. The adjective *healthy* is typically used. For example:

[7.59] John and Faye Smith say **the jab transformed their healthy, intelligent son** into a child needing round-the-clock care

Often, *healthy* is pre-modified, often with the intensifier *perfectly* <31> or with the temporal adverb *previously* <23>, for example:

[7.60] **She was perfectly healthy prior to vaccination** and then became severly deaf following vaccination after an illness and measles type rash.

[7.61] Seven days before Christmas, **she found her previously healthy son dead**.

Recounting personal narratives or the narratives of others is a frequently used rhetorical strategy. They may be used as a warrant for starting or contributing to a thread, or they may serve to lend support to an argument that vaccines are dangerous.

### 7.3.3 Delegitimizing the expertise of others

Reflecting the findings of Vayreda and Antaki (2011), in their examination of a forum thread discussion about vaccination, suggesting that a person has links to the pharmaceutical industry is one strategy used on the JABS forum to delegitimize the arguments or others. As mentioned in Chapter Six, one of the *JABS* keywords is *pharma* <990>, which is a derogatory

term for the pharmaceutical industry. It often occurs as part of the expression *big pharma* <452>. The collocational environment of *pharma* shows clearly that it has a negative prosody. The top lexical collocates from among the top twenty collocates of *pharma*, according to MI3, are *big, trolls, puppets, troll,*[71] *shares, shill,*[72] *occam,*[73] *monopoly*. Referring to links between the pharmaceutical industry and public bodies is one way in which mistrust of medical-scientific authorities is suggested, for example:

> [7.62] **Big Pharma and their puppets in Government** are the biggest threat to your families health and well being while they pretend to be the opposite.
>
> [7.63] Interesting that **the system is so corrupted by Big Pharma/Big Science** that the only ones that dare to speak the whole truth are the citizen-scientists with no agendas
>
> [7.64] **The FDA now effectively protects and nurtures the monopoly of big pharma** in the name of regulating its activities.

Frequently, the suggestion that someone has links with the pharmaceutical industry or is endorsing it is used to delegitimize an argument. Forum participants who adopt a pro-science stance are sometimes accused of having links with 'pharma' of being a 'pharma troll' or 'pharma shill':

> [7.65] Your lies don't stack up even with back up from Believe Steve they're more transparent than ever, **worried about your PHARMA shares** and your free lunches for the talks you do Occam Ye!
>
> [7.66] **ignore Becky and Quail they are Pharma Trolls pushing PHARMA business** and your baby is the bread and butter of the PHARMA industry how sick are they?
>
> [7.67] Please don't flatter yourself that I hoped to influence you. **I can recognize a pharma shill when I see one.**

[71] The term *troll* is used to refer to someone who posts deliberately inflammatory comments on the internet.
[72] The term *shill* is used to refer to someone who endorses the products or services of a company without disclosing that they have close links with the company.
[73] *Occam* is the nickname of one of the pro-science posters on the JABS forum.

The term *pharma* therefore frequently occurs in the context of ad-hominem attacks, often in the expressions *pharma troll(s)* or *pharma shill*. The occurrence of the term *pharma* is a clear signal of an expression of mistrust.

## 7.4 The construction of arguments in the *JABS* corpus

### 7.4.1 Introducing information using *the fact that*

The main way in which expertise is performed is by presenting knowledge and incorporating it into a persuasive argument. There are a number of ways in which this is achieved, but the discussion here will focus on the use of expressions of factivity. In this section, the expression *the fact that* is examined. As argued in Chapter Three, because it acts as a presupposition trigger, the expression *the fact that* can be used to introduce new information into a text whilst presenting it as shared knowledge. At the same time, the proposition introduced by *the fact that* is given the status of 'fact'. *The fact that* acts as a vehicle for propositions to become facts and to travel from their source domain to new ones (Hunston, 2011). The phenomenon of introducing propositions using *the fact that* is markedly frequent in the *JABS* corpus data. We see this if we compare it with its occurrence in other written genres. Although it use in the JABS corpus is not as frequent as in spoken British English, it is relatively more frequent than in the *New Scientist* and *Times* sub-corpora, as Table 7.4 below shows.

| **Corpus** | ***JABS*** | ***New Scientist*** | ***Times*** | ***British Spoken*** |
|---|---|---|---|---|
| Relative frequency | 170.1/m | 97.8/m | 112.1/m | 192.7/m |

**Table 7.4** Frequencies of *the fact that* per million words in *JABS* corpus and in the *New Scientist*, *Times* and *British Spoken* sub-corpora of the *Bank of English*

The semantic sequences involving the expression *the fact that* are shown in 7.5 below. Hunston's (2011) categorization of semantic sequences and 'motifs' has been adapted to suit the JABS findings.

| **SEMANTIC SEQUENCES WITH *the fact that* IN THE JABS CORPUS** | | | |
|---|---|---|---|
| **MOTIF** | **SEMANTIC SEQUENCE** | **PHRASEOLOGY** | **Freq.** |
| HUMAN RESPONSE | AFFECTIVE REACTION TO A FACT | *BE/FEEL concerned/impressed etc. about/by the fact that/have reservations about the fact that/take heart from the fact that/find comfort in the fact that* | 25 |
| | | *the fact that + PREDICATE (e.g. makes one wonder; doesn't help; lent an air of surreality etc.)* | 24 |
| | | *the fact that …is (not)/should be seen as interesting/surprising/a threat etc.* | 23 |
| | | *what is/what I found interesting/important/a shock etc. is the fact that* | 19 |
| | | *AGREE with/OBJECT to/ REGRET/RESPECT/LIKE etc./ stand by the fact that* | 12 |
| | | *BE against the fact that* | 4 |
| | | *CONSIDER the fact that* | 3 |
| | SOMEONE TALKS ABOUT/ IS AWARE OF A FACT | *DISCUSS/MENTION/EXPLAIN/ CHALLENGE/QUERY etc. the fact that* | 23 |
| | | *DRAW/CALL (s.o.'s) attention to/alert s.o. to/POINT to/highlight/endorse the fact that* | 15 |
| | | *WAKE up/CATCH on/get wise to the fact that* | 6 |
| | | *BE aware of the fact that* | 4 |
| | | *BE ignorant of/oblivious to the fact that* | 4 |
| ORIENTATION/ HUMAN RESPONSE | SOMEONE FAILS TO MENTION/TAKE ACCOUNT OF A FACT | *despite/in spite of the fact that* | 40 |
| | | *IGNORE the fact that* | 10 |
| | | *OVERLOOK/MISS the fact that* | 6 |
| | | *BE/MAKE no/does not make mention of the fact that* | 6 |
| | | *not accept/acknowledge/talk about the fact that* | 5 |
| | | *FAIL/FORGET to/does not mention the fact that* | 4 |
| | | *COVER up the fact that* | 4 |

| | | | |
|---|---|---|---|
| | | *walking away from the fact that* | 3 |
| | | *hush hush on the fact that* | 2 |
| | | *not take account of the fact that/without regard for the fact that* | 2 |
| CAUSE | FACT EXPLAINS/INDICATES SOMETHING (or not) | *the fact that ... suggests/shows/indicates/proves/means that /does not mean that/tells us that* | 36 |
| | | *POINT to the fact that* | 5 |
| | | *as evidenced by/lends evidence to/is (not/no) evidence/proof of the fact that* | 5 |
| | | *X is shown/can be explained by the fact that / X can/could be attributed to the fact that* | 4 |
| | FACT IS THE BASIS FOR REASONING | *the fact that ... SUPPORT/reflects/bolsters/reinforces/strengthens/ corroborates/gives endorsement/credibility/ shouldn't make your point more valid* | 11 |
| | | *based/built on the fact that/no basis to say ... other than the fact that* | 9 |
| | | *strengthened/supported/justified by the fact that* | 3 |
| | FACT HAS A CAUSAL LINK WITH SOMETHING | *due/down/thanks to the fact that/the fact that BE due to* | 15 |
| | | *COME/STEM from/LIE/REST in the fact that* | 11 |
| | | *to do with/RELATE to the fact that* | 10 |
| | FACT IS THE CAUSE OF A PROBLEM | *[problem/difficulty is] compounded/complicated/caused by the fact that* | 7 |
| | | *problem/reason/cause BE (at heart of) the fact that* | 6 |
| | | *the fact that ... made X worse/safer/more boring* | 3 |
| ORIENTATION | SOMEONE ASSUMES A FACT | *in light of/ in the face of/ given the fact that* | 10 |
| | | *apart/aside from/besides the fact that* | 7 |
| | | *were it not for the fact that* | 3 |

**Table 7.5** Semantic sequences and 'motifs' with *the fact that* in the *JABS* corpus

The motif which most frequently occurs when *the fact that* is used in the *JABS* corpus is the 'human response', in particular the semantic sequence AFFECTIVE REACTION TO A FACT. Therefore the most common strategy used to align the reader with a point of view is through an appeal to emotions. The linguistic devices used to realize the semantic sequence AFFECTIVE REACTION TO A FACT can be divided into those where the writer directs the reader's attention to his/her evaluation of an aspect of the 'fact', usually by placing *the fact that* in clause initial position or by using a pseudo-cleft construction, and those where an individual's emotional response (typically the writer's or that of someone referred to by the writer) is foregrounded. Instances of the former, are as described in the literature (see Hunston, 2011) and warrant no further attention here. Instances of the latter are a little more interesting in that, when used by an individual poster on the forum or message board (as opposed to when used in uploaded texts), examples of AFFECTIVE REACTION TO A FACT may fulfill specific functions. For example, it is sometimes used as a warrant for raising the topic:

> [7.68] My son is 15 months old and at the stage where he is due for the MMR jab. Im in a real quandry about what to do. Firstly, I cant find anywhere that will do the vaccinations separately (I live in Stockport, Cheshire) and secondly, even if I could locate somewhere, **I still have reservations about the fact that** the single vaccines are unlicensed in this country.
>
> [7.69] In light of the fact that measles outbreaks seem to be a greater threat now we are thinking of the single vaccines - either NHS if we can or paying for them privately but **we are worried about the fact that** they seem to be unlicenced - does that mean they are now considered safe??

Sometimes it is used as a warrant for responding to a topic raised by another poster:

> [7.70] Hi John,
> **I am impressed by the fact that** Marc Girard is not afraid to speak the truth.
> Could you please let us know about any responses that he receives, I am certain that they will be interesting. Thank you. Rosemary

But in some instances, it serves not only as a warrant for contributing to a topic raised in a thread but as an apparent source of moral authority. These are cases where a poster offers an account of their emotional response to something together with a personal anecdote. Examples 7.71 and 7.72 show the final two paragraphs of a very long post. The paragraphs reproduced below constitute a response to the question, posed in a prior post, 'Why are you all so against all vaccines, even those that are not thought to play a part in things like autism?' In these paragraphs, the writer has used the rhetorical device of parallelism (to what might be considered an exaggerated degree) to provide a list of things to do with vaccination policy or the authorities that she objects to. Each sentence begins with *I am/I'm against*. In the final paragraph, the expression *the fact that* appears on three (consecutive) occasions for no other reason than that the grammar requires it. However, it is interesting that these sentences come together in a clump at the end of the final paragraph and the propositions expressed are not generalizations about the state of vaccination policy but are details of her and her child's experience. This foregrounds the propositions expressed in these sentences by giving them end-focus and implies that the writer possibly considers these propositions more persuasive than the preceding ones. The fact that the poster to whom the comments are addressed does not participate again on this thread perhaps suggests that an emotional argument, drawing on personal experience as this does, effectively closes down debate.

> [7.71] You ask why we are against all vaccines. There is no short answer to this. I am not against all vaccines as it happens. I am against the current system which allows badly tested vaccines which contain toxic ingredients be be administered to children on a one size fits all basis. I am against the lying and the cover ups. I am against the inadequate and dangerous reporting system for adverse events. I am against the shameful manner in which vaccine damaged people (particularly those damaged by MMR) are treated, or rather not treated. I am against the scaremongering when it comes to the risk of disease. I am against the exaggerating of vaccine efficacy. I am against the contamination of vaccines with monkey viruses like SV40. I am against the indemnity of vaccine manufacturers which relieves them of the responsibility of the damage they knowingly do whilst allowing them to make a ton of money. I am against the witch-hunting of honourable people like Dr Wakefield, Dr Walker-Smith and Dr Murch.

[7.72] **I am against the fact that** I trusted my doctor to care for my child, do her no harm and respect our right to informed consent and then he, and the system behind him, let us down. **I'm against the fact that** my child suffered terribly for years with a distressing condition that she never should have had. **I'm against the fact that** my child and an unknown number of other children will never regain their full health potential and will carry this burden for the rest of their lives. I'm against the people who know this happens but lie about it to protect their own interests. If you think about it a little you will see that the above is rather different to 'being against all vaccines' which is actually a pretty meaningless phrase IMO.

Example 7.73 comes from a thread about a news story concerning the then Chancellor of the Exchequer Gordon Brown's involvement in a finance initiative to promote vaccination in poor countries. In Example 7.74, Janet picks up on Barefoot 1's reference to Brown's son's medical condition, cystic fibrosis, and volunteers the fact that her nephew has the condition. This gives her a warrant for entering the thread and for claiming expertise on the subject of cystic fibrosis. She later provides a warrant for the sort of expertise that is valued on the JABS site by saying that she has a vaccine damaged child. The two pieces of information are linked by the sequence *I regret the fact that … because*.

[7.73] <barefoot 1> Be warned if you think Blair was bad wait for Brown to get in power. While Brown and his wife are rightfully getting some sympathy for his son's condition Brown wants to poison and commit genocide on millions of african children on behalf of his masters.
Then they will just blame the phony disease aids for the deaths as part of the population control programmes. They have not ever isolated the HIV virus and aids is one the biggest scams of the 21C.
Brown like Blair is a utter puppet of big pharma but he is far more of a Bully and hard man than Blair.
SCAM Society's Common Accepted Model is a scam.

[7.74] <Janet> My twelve year old nephew suffers with Cystic Fibrosis and may have a short lifespan. **I regret the fact that** I don't get the time to see him as often as I would like, because I have to care for my (vaccine damaged) autistic child, 24/7.
Hopefully a cure for CF will be found in the near future, but it's difficult to sympathize with someone from a nasty government that has inflicted so much suffering and misery on others.

Often, other people's affective responses are evaluated. These tend to occur in texts which are uploaded onto the site or copied into forum posts from external sources:

> [7.75] **Temporao**, who has expended considerable energy to legalize abortion, **claims he is concerned about the fact that** 17 Brazilian children each year suffer birth defects from the disease, in a nation of more than 180 million people.
>
> [7.76] **It almost seems that Mr. Deer is less upset about what I wrote than about the fact that** some web site somewhere had picked it up.

Although the 'human response' motif occurs in a lot of the JABS corpus evidence for *the fact that*, it does not permeate all of the discourse. Propositions prefaced by *the fact that* are frequently deployed in the *JABS* corpus in the construction of rational argument. The main vehicles for this are semantic sequences expressing the 'orientation/human response' and the 'cause' motifs. The high frequency of the semantic sequence SOMEONE FAILS TO MENTION/TAKE ACCOUNT OF A FACT has been noted above. The most frequently occurring lexical realization of this string is *despite the fact that*. This tends to be used to criticise the actions of a person or group of people in a position of authority and occurs most frequently in stretches of text authored by JABS posters themselves:

> [7.77] You will not believe this, but despite these 2 reactions and the mother's suspicion that the vaccines might be involved (and **despite the fact that** her symptoms were starting to resolve but were still more than evident), on the 31st of August, she was given a third HPV vaccine and, at the same time, a DPaT!
>
> [7.78] Note, also that it was Legal Services Commission who blocked the Vioxx litigation in this country **despite the fact that** the case had already been conceded in the US, so effectively UK citizens now have no legal protection against the pharmaceutical industry.

More often, propositions introduced by realizations of the semantic sequence SOMEONE FAILS TO MENTION/TAKE ACCOUNT OF A FACT are used to counter particular claims by

highlighting a flaw in an argument. Some of these, such as examples 7.79 and 7.80 occur in texts authored by *JABS* posters although the majority occur in uploaded or copied texts:

> [7.79] ... of course it will look the vaccine was responsible for the eradication **if you ignore the fact that** the posion caused the disease in the first place and the vaccine was not responsible for its eradication.
>
> [7.80] Even if there had been a change in pattern of diagnosis it would have been beyond comprehension how more than 80 children at secondary level with a statementable level of autism could have gone undetected despite the crippling and disruptive nature of the disorder, and **despite the fact that** they were being monitored by the same services.
>
> [7.81] Several methodological problems marred the research and, **despite the fact that** there was a higher prevalence of autism among the children who had received the MMR vaccine, the authors asserted that there was no MMR-autism connection.
>
> [7.82] This **despite the fact that** ten thousand or so are part of one medical practice in Chicago that says it has virtually no autism or asthma among its never-vaccinated, home-birthed children. **Despite the fact that** homeschooled kids have a significantly lower vaccination rate and, according to one doctor who treats them, almost no autism. **Despite the fact that** the Amish ... oh, never mind.
>
> [7.83] The WHO claims that "antivaccine" organizations have spread such rumors in such countries as Argentina and Nicaragua, and denounces such activity as "defamation". However, it **does not mention the fact that** laboratory tests in Argentina and previous campaigns proved the existence of sterilizing HCG.

As for realizations of the 'cause' motif, when this motif is used, the proposition prefaced by *the fact that* is used as the basis for a conclusion.

> [7.84] **The fact that** Fitzpatrick's original strategy was to attack the middle classes **suggests** that overall record vaccination record of the case load did not strongly support the case.
>
> [7.85] Some autistic children have measurable abnormalities in relation to heavy metals. Blanket dismissal of the issue is neither helpful or well-informed. **The fact that** people are even opposed to researching it **suggests** bias.
>
> [7.86] **The fact that** Government's have bought into this stupid vaccine **shows** how utterly hoodwinked they are by pharmaceutical interests.
>
> [7.87] The fact remains that Vit C does stop AIDS infection - **the fact that** it's only in cell cultures **means** that when scaled up to the human body (MILLIONS OF CELLS!!!!!!) then it's imense health giving properties will be realised and free will the human population be from the chains of drugs and allopathy.

[7.88] **The fact that** 'the postmortem studies show distinct immunological features different from what is typical of ALS' **suggest** an association between vaccination and ALS, he says.

There are examples in which *the fact that* is used to express causality, but these are rarer than the other 'motifs'. Two examples are:

[7.89] The Schick test was very inaccurate which was why it was abandoned in the 1930s in America and Britain, among other places. **The inaccuracy is due to the fact that** immunity to the disease does NOT correlate with protection from the toxin.

[7.90] When the data of study was reviewed, it was found that the sampling was flawed. **The low incidence of autism during the use of thimerosal can be attributed to the fact that** the database that was used only tracked inpatient cases of autism at the time.

In the majority of cases in which *the fact that* occurs in the *JABS* data, therefore, it realises a semantic sequence which, in turn, realizes the 'human response' motif. The semantic sequence most frequently realized by *the fact that* is AFFECTIVE REACTION TO A FACT, while the next most frequently occurring sequence is SOMEONE FAILS TO MENTION/TAKE ACCOUNT OF A FACT.

### 7.4.2 Introducing information using introductory *–it* constructions and pseudo-cleft constructions

#### 7.4.2.1 Introductory-*it* constructions and cleft sentences with *it*

Like *the fact that*, clauses with introductory-*it* and clefts also act as presupposition triggers. The most frequently occurring adjectives and nouns express an evaluation of the proposition encoded in the *that*-clause in terms of certainty. But other well represented expressions encode evaluation in terms of clarity and importance. The adjectives, nouns and phrases, and their frequencies when they occur in the patterns *it BE (to*-inf/ v-*ing)* ADJECTIVE *that* and *it BE* + NOMINAL GROUP *that* are shown in Table 7.6, ordered according to their semantic

categories.[74] As was seen with *the fact that*, expressions which denote an affective reaction to the propositional content expressed in the projected *that*-clause make up a substantial proportion of all occurrences of clefts and constructions with introductory-*it*. However, they do not make up the most frequently occurring category. Expressions belonging to the categories of LIKELIHOOD and CLARITY are more frequent. With the exception of the, somewhat infrequent, expressions *safe to say that* and *predictable that*, the expressions in the LIKELIHOOD category indicate a hedged commitment to the truth of the proposition.

| Category | Expression | Total |
|---|---|---|
| LIKELIHOOD | *possible* <115>, *likely* <73>, *expected* <15>, *plausible* <13>, *probable* <12>, *safe to say* <10>, *doubtful* <7>, *fair to say* <5>, *predictable* <3> | 253 |
| CLARITY | *clear* <95>, *obvious* <50>, *apparent* <18>, (*self-*)*evident* <21>, *commonsense* <3> | 187 |
| AFFECTIVE REACTION | *interesting* (*to note*) <56>, *a*/*any*/*no* (*small/mere*) *coincidence* <26>, (*not*) *surprising* <21>, *unfortunate* <13>, *hard to believe* <9>, *understandable* <7>, *nice to know* <7>, *no surprise* <6>, *no secret* <5>, *worrying* <5>, *curious* <4>, *disturbing* <4>, *troubling* <4>, *amazing* <3>, *difficult to believe* <3>, *little wonder* <3>, *regrettable* <3>, *extraordinary* <2> | 176 |
| SHARED KNOWLEDGE | *(well) known* <71>, *estimated* <35>, *understood* <19>, *(widely/generally) accepted* <15>, *recognised* <6>, *acknowledged* <5>, *agreed* <5>, *proven* <5>, *common knowledge* <4>, *presumed* <2> | 174 |
| IMPORTANCE | *important* (*to remember*) <60>, *worth noting* <15>, *vital* <13>, *crucial* <11>, *essential* <8>, *noteworthy* <7>, *imperative* <6>, *worth mentioning* <6>, *worth pointing out* <6>, *critical* <4> | 136 |
| FACTIVITY | *(not) the case <20>, fact <18>, my belief* <8>, *myth* <5> | 51 |
| JUDGEMENT | *true <29>, right* <8>, *bad enough* <4>, *wrong* <4> | 45 |

**Table 7.6** Frequencies of expressions used with introductory-*it* or in cleft constructions with *it* in the *JABS* corpus

[74] The categories are based on Hunston's (2011) description of the semantic sequences associated with *the fact that*.

Concordance evidence shows, though, that these expressions are frequently intensified. As much as 25% of the citations of *it is possible that* and 45 % of instances of *it is likely that* are intensified, for example:

> [7.91] He said "**It is entirely possible that** the immune systems of a small minority simply cannot cope with the challenge of the three live viruses in the MMR jab, and the ever-increasing vaccine load in general."
>
> [7.92] Now, I concede that with regard to the morality of inanimate objects, **it's perfectly possible that** you're too stupid to grasp the difference between (a) the different meanings of the word "bad" (i.e. morally wrong vs substandard or inferior) and (b) the morality of an action undertaken by a human and the result of that action/the tools used to commit that action (so, for example, a kitchen knife isn't immoral, but stabbing someone with one in order to steal their ipod is).
>
> [7.93] **It is quite possible that** those families with ahistory of autism went on to avoid MMR, undermining the study findings.
>
> [7.94] ... this now provides evidence that **it is highly likely that** MMR vaccine is the source of the measles virus that is in turn linked via significant evidence with ileal lymphoid nodular hyperplasia, which in turn is strongly and convincingly linked with regressive autism.
>
> [7.95] It is therefore **very likely that** many other infants' deaths that followed the administration of the influenza vaccine were never reported to VAERS.

Expressions in the CLARITY category imply a high degree of certainty in the truth of the proposition in the projected that-*clause*. One would expect, therefore, that expressions in this category would require no intensification. There is, it transpires, less intensification in this category than in the LIKELIHOOD category, although there is intensification all the same. 15% of citations of *it is clear that* are intensified, as are 30% of citations of *it is obvious that*:

> [7.96] **It is quite clear that** the rate of autism in this country is directly time-associated with the increased numbers of vaccines given to children.

[7.97] There would be no need for this site to exist, were it not for the fact that **it is increasingly obvious that** a small group of eminent and influential mainstream scientists are willing to countenance only one version of events about how AIDS began

[7.98] So Lee **it is quite obvious that** there are far better & more effective choices for proper treatment of viruses & diseases caused by viruses therefore drugs cannot be even considered an option.

[7.99] **It is perfectly obvious that** the integrity of the manufacturer of a vaccine is relevant

Expressions in the AFFECTIVE REACTION category can be divided into those which express an evaluation of:

- how interesting or striking the proposition is (or is not), for example *interesting (to note), curious, amazing, extraordinary*;
- how expected or unexpected the proposition is, for example, *a*/*any*/*no* (*small/mere*) *coincidence*, (*not*) *surprising, hard to believe, no surprise, difficult to believe, little wonder;*
- how desirable or undesirable the proposition is, for example, *unfortunate, regrettable, nice to know*;
- how unsettling the proposition is, for example, *disturbing, troubling, worrying;*
- and how comprehensible it is: *understandable*.

The SHARED KNOWLEDGE category is similar to the LIKELIHOOD category in that the expressions exist on a cline of degrees of certainty. Expressions in this category act as a form of attribution, in that they encode an appeal to (unspecified) people's knowledge or beliefs. By far the most frequently occurring expression in this category, though, is one which expresses a very high degree of certainty: *it is* (*well*) *known that*. Most of the other expressions in this category also encode a high degree of certainty. These are *it is understood that, it is recognised that, it is acknowledged that, it is agreed that, it is proven that* and *it is*

*common knowledge that*. Expressions such as these encode the notion that there is consensus about the status of the proposition, such that it is almost certain to be true.

Another very frequently occurring set of expressions are those which evaluate the proposition in terms of its importance. Expressions in the FACTIVITY category function in a similar way to the expression *the fact that*, in that they imply that the proposition in the projected-*that* clause aligns with reality. These are relatively infrequent, however. The smallest category contains those expressions which encode a moral judgement.

#### 7.4.2.2 Pseudo-cleft constructions with *what*

There are 210 occurrences of pseudo-cleft constructions with *what* in *JABS*. The greatest proportion is used to express the affective stance of the writer. These frequently encode the writer's affective reaction to something:

> [7.100] **what annoys me is that** these so called medical professionals give our children drugs without properly testing them 1st!
>
> [7.101] **What concerns me is that** the government is saying that the MMr is safe but as no health worker has recorded my son as having an adverse reaction to this vaccination
>
> [7.102] **What puzzles me is why** the American researchers needed to use children who were deprived and in another country?
>
> [7.103] **What worries me is that** will he have the same side effects as the first jab and can this second one lead to autisim.
>
> [7.104] **What I can't understand is why** anyone would get a child a mumps vaccine.
>
> [7.105] **What I don't get is why** they are doing all this; they must know the risks?

Another group express the writer's evaluation of the relevance or importance of the proposition:

[7.106] Perhaps **what matters most is that** millions of units of the vaccine are sold.

[7.107] Above all, surely **what matters is** the quality of the science in these papers.

These expressions may also be used to highlight the proposition the writer is putting forward (Example 7.108) or to introduce a reformulation of something they have previously said (Examples 7.109 and 7.110):

[7.108] **What I do know is that** the 'facts' of the HPV vaccine can already be classified in this clever guy's latter category of unknowns.

[7.109] but **what i am asking is** not about the mmr but about having the single measles jab and help in the questions i asked !

[7.110] **what I am saying is that** a celebrity passing off anecdote as medical evidence does neither side of the discussion any favours

Another relatively large group are used to report what a third party is saying whilst foregrounding, and thus evaluating, a particular aspect:

[7.111] **What the data is saying, is that** 15 plus years of 90% coverage (95 - 99.5% for 10 years) and these massive outbreaks still occur?

[7.112] **What they found was that** the rare adverse events that you may have heard about in the news don't seem to be any more common in the people who got the vaccine than the people who didn't get the vaccine," he said.

Often, the actions or words of a third party are evaluated and, once again, an accusation which recurs is that particular people in authority have failed to say something or take something into account:

[7.113] **What they didn't say is that** no one else had looked.'

[7.114] But **what they don't say is that** these complications are all derived from acute blood toxemia established by the very treatments used by allopathic physicians.

[7.115] **What they don't realise is that** every time these experts make predictions that don't materialise

The evidence for the use of introductory- *it* constructions and cleft and pseudo-cleft constructions shows, once again, a high degree of evaluative language. These constructions are used strategically to highlight a point in an argument, to evaluate a proposition, or to reformulate a proposition.

### 7.4.3 Negotiating factual status

#### 7.4.3.1 *Evidence that*

It was found in Chapters Five and Six that one of the expressions which occurred most frequently in the *JABS* and *NHSvax* corpora and was used to rebut the causal hypothesis was (*there is) no evidence of a link between mmr and autism*. It is therefore hardly surprising that the frequency of the word *evidence* is markedly high in both corpora. The frequencies of the word *evidence* per million words of text in the *JABS* and *NHSvax* corpora and in the *New Scientist, Times,* and *British Spoken* sub-corpora of the *Bank of English* are shown in Table 7.7 below.

| ***JABS*** | ***NHSvax*** | ***New Scientist*** | ***Times*** | ***British Spoken*** |
|---|---|---|---|---|
| 1223.2/m | 1622.9/m | 395.8/m | 169.2/m | 45.1/m |

**Table 7.7** Relative frequencies of *evidence* in *JABS* and *NHS*vax corpora, and *New Scientist, Times,* and *British Spoken* sub-corpora of the *Bank of English*

Evidence is, of course, fundamental in the epistemology of the hard sciences (Hunston, 2008). The relatively high frequency of *evidence* in the *New Scientist* sub-corpus compared with the *Times* and *British Spoken* sub-corpora testifies to this. As explained above, the markedly high frequency in the *JABS* and *NHSvax* corpora is due, mainly, to the centrality of the notion of evidence in the MMR debate. Use of the word *evidence* is not confined to discussion of the MMR issue, however. Also, whether or not used in contexts in which the causal hypothesis is rebutted, examination of the *JABS* corpus reveals interesting uses of the word. For instance, there is a lot of negotiation apparent in the *JABS* corpus data around the nature and role of evidence.

One source of contention among the contributors to the *JABS* texts concerns the so-called 'evidence-based medicine' policy of the Department of Health and the NHS. There are 79 citations of *evidence-based* in the *JABS* corpus. Concordance data shows that notions regarding the nature, reliability and usefulness of evidence-based medicine are frequently negotiated. In Example 7.116, evidence-based is not only modified with the adjective *credible*, it, in turn, is used to modify the word *opinion*, rather than *medicine*. In Example 7.117, the writer expresses the proposition that any justification for evidence-based medicine has been undermined once and for all. In Example 7.118, the writer questions another's belief in evidence-based medicine.

> [7.116] you need to widen your sources of information or you may not do justice to your daughter, you know the duty of care you owe her to be informed from any source with **credible evidence-based opinion**.
>
> [7.117] **PSRM has blown evidence-based medicine clear out of the water** - beyond any reasonable doubt, the evidence base has been rendered utterly meaningless.
>
> [7.118] Do you **not believe in evidence based medicine**?

In the following example, the vaccine-critical, 'alternative' medical writer, Dr Edward Yazbak, expresses his lack of faith in evidence-based medicine:

> [7.119] **I strongly objected to the fact that** the committee had decided that "further research to find the cause of autism should be directed toward other lines of inquiry that are **supported by current knowledge and evidence** and offer more promise for providing an answer." By doing so, the committee had effectively shut down all MMR and thimerosal autism research.

This ambivalence about evidence colours the use of *evidence* in the JABS corpus. A high degree of negotiation is apparent. For example, *evidence* is often pre-modified. The adjectives can be divided into the following categories:

- adjectives denoting a scientific or medical field: *scientific* <220>, *medical* <51>, *epidemiological* <35>, *clinical* <30>, *biological* <6>;
- adjectives denoting quality: *anecdotal* <54>, *documentary* <14>, *hard* <32>, *expert* <14>, *statistical* <14>, *real* <13>, *actual* <11>, *direct* <9>, *factual* <8>, *indirect* <8>, *physical* <8>, *solid* <8>, *vital* <7>, *circumstantial* <7>, *proper* <6>, *written* <6>;
- adjectives denoting reliability: *strong* <49>, *credible* <36>, *good* <36>, *convincing* <35>, *compelling* <25>, *clear* <20>, *reliable* <15>, *best* <9>, *conclusive* <8>;
- adjectives denoting novelty and availability: *new* <41>, *available* <23>, *further* <23>, *published* <15>, *current* <11>, *fresh* <7>, *existing* <7>;
- adjectives denoting quantity: *little* <26>, *enough* <25>, *overwhelming* <23>, *insufficient* <17>, *growing* <16>, *sufficient* <12>, *limited* <9>, *mounting* <8>, *increasing* <7>.

On the forum and message board, it is most often, but not exclusively, the pro-science posters who appeal to *evidence*, and strengthen their arguments by asserting that there is *no evidence*.

> [7.120] **There's absolutely no evidence that** homeopathy has anything actual medical effect.
>
> [7.121] **There is no evidence of** a link between MMR and autism. Yes there are side effects but this is the case for all vaccines and the small number of people who fall ill is miniscule compared to the amount of people who are saved.

While one might expect pro-science posters to appeal to *scientific evidence* more than their vaccine-critical counterparts, this is not the case. Vaccine-critical posters are just as likely to use the term. However, sometimes they use it to support their argument, as in the three examples below, especially when expressing mistrust of the medical-scientific community:

> [7.122] Vaccines are never tested on ill children, so **there is no scientific evidence of safety for children with unusual health problems.**
>
> [7.123] Herd immunity - pure scare tactics **No scientific evidence that vaccines work**.
>
> [7.124] Almost everything doctors do is based on a conjecture, a guess, a clinical impression, a whim, a hope, a wish, an opinion or a belief. In short, **everything they do is based on anything but solid scientific evidence.**

The reliability of anecdotal evidence is particularly contested. The anti-vaccination voices evaluate *anecdotal evidence* positively. In the following example, *anecdotal evidence* is represented as equivalent to *experimental evidence*, although *anecdotal evidence* is foregrounded by being placed before *experimental evidence* and the writer emphasises the strength of the anecdotal evidence by claiming that there is a *huge pile* of it. The positive evaluation is further encoded in the fact that the writer claims the evidence supports his argument:

> [7.125] What I am saying though is that there's **a huge pile of anecdotal evidence** and some experimental evidence too which supports the idea that MMR vaccination might cause autism.

The difference in attitude between the medical-scientific community and many parents towards the status of evidence classed as anecdotal is illustrated in the following example, from a post by a vaccine-critical poster:

> [7.126] **Anecdotal evidence** is no longer respected.

Meanwhile, the pro-science posters offer a negative evaluation of anecdotal evidence. The following poster equates *anecdotal evidence* with *the drivel* [parents] *read in the press*:

> [7.127] The reality is that the press are simply using the MMR debate as an opportunity to try and damage the government, while **fretful**, **paranoid parents listening to anecdotal evidence and the drivel they read in the press** are putting all children at risk to three awful diseases.

One significant difference between the attitudes of vaccine-critical and pro-science writers is in the extent to which they take the lack of evidence that MMR causes autism as providing proof that the vaccine is safe. The forum exchange below illustrates this. The first poster supports the 'evidence-based medicine' approach so says that choice should be *based on the evidence*. He evaluates the lack of evidence that MMR causes autism as a 'fact'. The word fact, in the second clause, presumably refers to the proposition that 'there is no evidence that MMR causes autism'. However, in the third clause, he reformulates the status of the proposition as a 'lie'. Presumably, here he intends the word *lie* to refer not to the proposition that there is no evidence, but to the proposition that MMR causes autism. Tomgirl accepts that there is not, or may not be, evidence to support the causal hypothesis, but she interprets the claim that there is no evidence as entailing that there might be evidence which no research has yet uncovered. She concludes that the possibility that MMR might cause autism, despite the lack of evidence, presents too great a risk.

> [7.128] Iggyfishtank:
> I agree that you should have a choice...**based on the evidence. There is no evidence that MMR causes Autism**. **FACT**. **The lie** is propegated by the daily mail and websites like

this. It is irresponsible for people to be peddling the claims of a discredited study 8 years after it was published.

[7.129] Tomgirl:
To Iggyfish tank
**There may not be evidence to show that Mmr definetly causes autism**, **but there is also no research to show that it definetly does not.** Not been a risk taker by nature, **the fact there is no evidence to show MMr does not cause autism** , is not good enough for me and is not worth the risk, as any possible damage cannot be then undone and the brian is rather important to functioning. I am not panicking i am making an educated decision. I can always change my mind **if further evidence comes up to show Mmr definetly does not cause autism** but i cannot change my mind to undo any damage if it does. Incidently my Fully mmr'd teenager got mumps, which my unvaccinated son did not catch.

#### 7.4.3.2 *Proof that*

Instead of using the word *evidence*, factivity is sometimes expressed using the word *proof.*

There are 470 occurrences of *proof* in the *JABS* corpus. Examples are:

[7.130] Melanie **No proof of a link** between MMR and autism has ever been established We're told that vaccination programmes eliminate diseases. **There is NO proof of this** in any country of the world.
Firstly **there is no scientific proof that** the MMR causes Autism as there is also **no scientific proof it doesnt**

The relationship between *proof* and *evidence* is sometimes negotiated:

[7.131] So, you can see why (in my opinion) its so difficult to prove, and **anecdotal evidence is not PROOF!** despite the growing numbers.

The term *proof* is taken as denoting a stronger degree of factivity than *evidence*.

### 7.4.5 Summary of Section 7.4

This section has focused on the use of the cohesive devices *the fact that*, *evidence that*, and *proof that*, as well as constructions with introductory–*it* and cleft and pseudo-cleft constructions. It has highlighted the fact that contributors to the *JABS* texts frequently frame propositions by expressing their affective reaction. The expression *the fact that*, may be used by forum participants as a warrant for raising a topic, or responding to a topic another poster

has raised. It is sometimes used as an appeal to moral authority. It is used to construct rational arguments. Propositions introduced with *the fact that* may be used to undermine an argument, by stating that someone has failed to mention a fact, or by pointing out a flaw in an argument. It may be used as the basis for a conclusion. Propositions prefaced by an introductory-*it* construction are typically modified with evaluative language. They are used to express the degree of commitment to the truth of a proposition, or to intensify the force of a proposition. Pseudo-clefts are used to highlight a proposition or to introduce a reformulation.

## 7.5 Conclusion

The aims of this chapter were to discover the ways in which contributors to the JABS forum express their warrant for expertise and to shed light on the ways in which they use discursive resources to construct persuasive arguments. Diverse methods for expressing a warrant for expertise were brought to light. Firstly, contributors to the *JABS* texts appeal to authoritative sources in diverse ways. There is evidence of extensive use of the modes of attribution which are characteristic of scientific discourse. Reporting terms with connotations of formality may be used to intensify the force of a proposition. However, few contributors to the forum employ terms from the scientific register to perform expertise, since most, presumably, lack such expertise. More frequent forms of warranting by forum contributors include referring to one's identity as a parent, or recounting a narrative. Narratives of vaccine damage, or of the death of a child following vaccination, are frequent, in particular, narratives in which the harmful effects of vaccines are emphasized by prefacing the account of the child's deterioration in health with the expression *previously healthy child*. A frequently used strategy for delegitimizing the arguments of pro-science posters on the JABS forum is to explicitly state or to imply that they have links with the pharmaceutical industry. Extensive use is made

of expressions which act as cohesive devices at the same time as giving a proposition a status label. In the environment of status labels which signal factivity, we find many expressions which signal an affective response to the proposition. There is negotiation around the notion of evidence, in particular in relation to evidence-based medicine, anecdotal versus scientific evidence, and evidence versus proof. This chapter also uncovered complexities in the ways in which intertextual references were exploited. As has been observed in prior studies, forum posters make intratextual references, repeating or reformulating their own words or the words of other posters. They repeat their own words in order to emphasize a point they have made in an argument or reformulate it in order to clarify their intended meaning. They reformulate the words of others in order to challenge an argument. Very frequently, texts are imported in their entirety into forum posts. The presenting of texts in this way serves to present a ready-made argument for the forum poster.

# CHAPTER EIGHT

# CONCLUSION

## 8.1 Introduction

The aims of this chapter are to assess the contribution this thesis makes to knowledge in the field of discourse analysis, with particular focus on the study of debates about science in the public sphere. It then assesses the extent to which the challenges of using a CADS approach to analyse discourse in text gathered from an online discussion forum have been overcome and considers the limitations of the research. It finally considers the implications for future research.

## 8.2 Contribution to knowledge

### 8.2.1 Introduction to Section 8.2

By undertaking an analysis of public debate about health and science in an online discussion forum, this thesis has responded to Myers' (2003) call for research into public understanding about science to move beyond the boundaries of the academic article into under-researched areas. This thesis has made a significant contribution to that goal. By examining the discourse in an online discussion forum, it has made important discoveries about the ways in which scientific issues are debated in one of the increasingly significant locations in which public debate takes place today. In particular, it has added to knowledge about the ways in which health risks are discussed online, thus extending the work of scholars such as Richardson (2001; 2003; 2005). It has also made a contribution to the body of knowledge gathered by studies which have focussed on vaccine-critical discourse (e.g. Hobson-West, 2005) and

public discourse about the risks of MMR (e.g. Boyce, 2007; Rundblad, Chilton and Hunter, 2006). Its contribution is unique in that it focuses on the ways in which lexico-grammatical resources are exploited by discussion forum participants to fulfil their interactional goals, and it succeeds in examining a large body of texts because it employs corpus linguistic methodology. As a study of vaccine-critical discourse, it differs from other studies in this field in that it not only illustrates the ways in which vaccine-critical individuals express their beliefs, but, since a number of pro-science posts are represented in the *JABS* corpus data, it offers insight into the ways in which vaccine-critical individuals and pro-science individuals interact with each other. With its mix of uploaded media articles, vaccine-critical texts, and forum posts by vaccine-critical and pro-science JABS members, the *JABS* corpus data affords access to a range of voices.

### 8.2.2 Contribution to knowledge about scientific popularization

#### 8.2.2.1 The reformulation of claims related to the causal hypothesis

The thesis has highlighted aspects of the complex ways in which claims about science and health are reformulated as they travel from one domain to another. Firstly, it was noted that, in the *NHSvax* corpus, when rebutting the MMR-autism causal hypothesis, the authors tend to used nominalized forms, such as *link, association, connection*, rather than explicit markers of causation, such as forms of the verb *CAUSE* (Xuelan and Kennedy, 1992). It was pointed out that the use of nominalization is typical of the tentative, or hedged, way in which claims are usually made in scientific discourse (Fahnestock, 1986) and that nominalizations are among the features which often carry over into popularization discourse (Jones, 2013). The authors of the *NHSvax* texts also tend to appeal to scientific evidence to support their claims.

Expressions which include the string *no evidence of a link* were found to be prevalent. The force of a claim might be intensified through use of pre-modification with a term such as *credible* or *scientific*. The most frequently found type of expression is *(there is) no (credible/scientific) evidence of a (causal) link between mmr and autism*.

Expressions of this kind were found to be frequent in the *JABS* corpus: nominalizations are used to refer to the causal hypothesis in just over 50% of cases. However, expressions such as *link/association between mmr and autism* and *no evidence of a link between mmr and autism* tend to occur in media texts, or extracts of media texts, which are copied into forum posts or uploaded onto the JABS website.[75] It was found that when JABS forum contributors express their beliefs using their own words, there is a greater tendency to use the more explicit signal of causation *CAUSE*. Expressions involving the use of *CAUSE* occur in just over 28% of cases in which the causal hypothesis is referred to in the *JABS* corpus. However, marked differences were found here between frequent and infrequent posters. Infrequent posters are more likely to express their ideas in a tentative way, hedging the verb with the modal *can* or *may*, for example, *does this mean that the single vaccine may also cause autism?* Frequent posters were found to be more likely to make stronger, unhedged claims, for example, *mmr kills and causes autism*. However, it proved to be the pro-science writers among the frequent posters who are most likely to make strong claims, for example, *mmr does not cause autism*.

[75] It was also noted that expressions such as *there is no evidence of a link between mmr and autism* in the *JABS* corpus tend to occur in segments of text in which authority figures are quoted. Expressions using these nominalized forms have thus already undergone some form of reformulation before they are reproduced on the JABS website: the reporting verb used signals how the proposition expressed in the quoted segment is to be evaluated, as does the message encoded in the headline and opening paragraph which frame the text as a whole (White, 1997).

8.2.2.2 Risk discourse

The collocational profiles of *risk* in the *JABS* and *NHSvax* corpora were found to show many similarities with the uses of *risk* identified by Hamilton, Adolphs and Nerlich (2007) as typical of the discourse of healthcare. Many of the significant collocates of *risk* in both the *NHSvax* and *JABS* corpora, therefore, are terms which denote assessment of degrees of risk, such as *INCREASE, REDUCE, high, greater*, and so on. The profiles of risks in the two corpora also showed marked similarities. The expression *the benefits (of vaccination) (far) outweigh the risks* was frequent in both, although, in the *JABS* corpus, the expression typically occurs in instances in arguments supporting vaccination are rebutted. When concordance evidence for *risk* in the *JABS* corpus was examined more closely, uses came to light which are not evident in the *NHSvax* corpus data. It was found that individual posters frequently use the expressions *PUT (someone) at risk, take the risk, run the risk* and *worth the risk,* which are typically used in contexts in which individuals refer to personal risk (Hamilton, Adolphs and Nerlich, 2007). These occur in the *JABS* corpus in contexts in which writers express resistance to the imperative to vaccinate. In such cases, there is evidence of the tension between government healthcare policy and parents' concerns for the well-being of their children. For example, in the citations shown below, we see expressions of the parents' sense of responsibility for their children and a sense that they feel the healthcare authorities are coercing them into accepting the vaccination.

> [8.1] Unfortunately as I am on disability benefits I am not i a position to be able to afford this, which I think is very unfair **that I should have to put my children at risk.**
>
> [8.2] **I am not going to put my children at risk** from a vaccine which doesn't even give much if any long term protection.

Scholars such as (Beacco *et al.*, 2002) and (Moirand, 2003) refer to the mix of discourses which characterize debates about science and health in the media. The relatively high frequency in the *JABS* corpus of expressions related to risk which are frequent in the *NHSvax* corpus alongside expressions which are more typical of assessments of personal risk is evidence of the sort of mix of discourses which characterize debates about vaccination in the *JABS* corpus.

#### 8.2.2.3 Expressing warrants for expertise

The findings of this thesis also support the observation of Myers (2003) and others that it is misleading to conceive of a strictly defined boundary between the expert and the lay person. Participants on the JABS forum were found to display diverse forms of expertise. Some display extensive knowledge of science, using discourse patterns typical of the scientific domain. The expressions *risk/benefit ratio* and *balance of risks and benefits*, for example, wer found to recure. These cases are relatively few, however, and displays of expertise of this kind are performed only by high-frequency posters, typically those who are high-profile campaigners. The more commonly used forms of warranting were found to consist of sharing news reports or recounting lived experience. This is evident in the examples of forum participants who appeal to their status as parents as a warrant for expertise. The experience of the parents of vaccine-damaged children was found to be particularly valued. This was evident in examples of the use of expressions such as *as a parent of a vaccine damaged child*. It was also found that the arguments of pro-science forum posters are often delegitimized on the grounds that they have no experience of being the parent of a vaccine-damaged child or have no sympathy with such parents. The discursive resources drawn on by participants also

include use of narratives of vaccine damage, thus there are examples such as *died …. after being vaccinated* or *found her previously healthy son dead*.

#### 8.2.2.4 Attribution

In the *NHSvax* corpus, appeals to authority consist mainly of appeals to scientific evidence, for example:

> [8.3] Update of **scientific evidence in published studies have continued not to find an increased risk of autistic spectrum disorder** associated with MMR

In the *JABS* corpus, by contrast, there is more evidence of attribution to specific individuals. Most of these, however, occur in uploaded news articles. We therefore find that the use of prosodically neutral reporting verbs, such as *SAY* or *REPORT*, is frequent. More interesting are instances in which indirect forms of attribution are used. It was found that expressions such as *concern(s) that* and *fear(s) that* are frequently used in attribution. Medical-scientific authorities and lay people alike are represented as having *concerns*, while authorities are more likely than lay people to be represented as expressing *concern* and lay people are more likely to be said to have *fears*. As is typical of online discussion forum interaction, segments of previous posts are often recycled (Richardson, 2001). This, too, contributes to the intertextual mix of the discourse. Forum participants may reiterate their own words in order to clarify or emphasize a point. Alternatively, they may do so in order to challenge the argument of another. Obviously, reporting verbs such as *CLAIM* may be used to challenge an argument by signalling that the writer does not concur with the reported proposition. However, it was also found that the verb *STATE* is sometimes used to imply disagreement. Because *STATE* is

typically associated with formal contexts, its use in the context of interaction between two discussion forum participants suggests that irony is intended.

#### 8.2.2.5 The use of *that*-clauses, introductory-*it* constructions, cleft constructions with *it*, and pseudo-cleft constructions with *what*

A lot of the discussion of the ways in which claims which originate in other domains are reformulated in *JABS* corpus texts focussed on the use of status nouns followed by a *that*-clause, introductory-*it* constructions, cleft constructions with *it*, and pseudo-cleft constructions with *what*. Unlike the *NHSvax* corpus, where little use was made of such constructions, the *JABS* corpus was found to show a relatively high frequency of use of these constructions. It was found that in many instances in which these constructions are used, there is marked use of evaluative language. It was found, for example, that the cohesive device, *the fact that*, which acts as a vehicle for reformulating propositions as facts (Hunston, 2011), is most frequently used in *JABS* to express the semantic sequence AFFECTIVE REACTION TO A FACT, typical realizations being *BE/FEEL concerned/impressed etc. about/by the fact that/have reservations about the fact that/take heart from the fact that/find comfort in the fact that*, and so on. Furthermore, in a relatively high proportion of expressions of the semantic sequence AFFECTIVE REACTION TO A FACT, the individual's affective reaction is foregrounded by being placed in the clause which precedes *the fact that*. Thus, not only is a proposition represented as a 'fact', but the reader's attention is directed towards the writer's reaction to the 'fact'. Interestingly, it was found that in many instances in which the writer's affective reaction is foregrounded, the expression of their reaction serves as an apparent source of moral authority.

When the use of introductory-*it* constructions and cleft sentences with *it* was examined, it was found that these constructions are most frequently used to evaluate a discursive object in terms of possibility, for example, *it is possible/likely that …*, or in terms of clarity, for example, *it is clear/obvious that….* Expressions of this kind are frequently intensified, for example, *it is highly likely that …*, *it is quite clear that …*, *it is increasingly obvious that ….* Introductory-*it* and cleft constructions were also found to be used frequently to express the writer's affective reaction to a proposition, for example, *it is interesting (to note) that …*, or to appeal to shared knowledge, for example, *it is well-known that ….* Pseudo-cleft constructions with *what* were found to be used most frequently to encode the writer's affective reaction to a proposition. There were several instances of use of the expressions *what annoys/concerns/puzzles/worries me is that ….* Constructions of this kind are also frequently used to evaluate what other people have said or done, for example, we find expressions such as *what they didn't say was that …*, *what they don't realise is that ….* The use of presupposition triggers, such as *the fact that*, introductory-*it* constructions, cleft sentences with *it*, and pseudo-cleft constructions with *what*, function to introduce new information as though it is given information, in other words, as though it is shared knowledge. Where these forms are used in the *JABS* corpus, they are frequently accompanied by explicit expressions of evaluation.

### 8.2.3 Contribution to knowledge about vaccine-critical discourse

#### 8.2.3.1 Science, risk and representations of the immune system

The findings of the thesis support Hobson-West's (2005) observation that vaccine-critical groups have an ambivalent attitude towards science, sometimes making appeals to science and scientific evidence to support their claims and sometimes challenging the fundamental precepts of medical science. We see this if we compare the ways in which beliefs about health, the immune system, and the risks associated with vaccination are represented in the *JABS* and *NHSvax* corpus data. Firstly, the risks of contracting diseases and the role of vaccines in protecting against disease are emphasized in the *NHSvax* corpus data. We therefore find numerous instances of realizations of the semantic sequence A BACTERIUM OR VIRUS CAUSES A DISEASE and repeated instances of *immunisation is the safest way to protect your child for life/your child's health.* Where possible undesirable effects of vaccines are referred to, euphemistic terms such as *adverse reactions/events* or *side effects* are used. In the *JABS* corpus, on the other hand, there are relatively few occurrences of the semantic sequence A BACTERIUM OR VIRUS CAUSES A DISEASE but several occurrences of the sequence A VACCINE CAUSES DAMAGE. Where expressions of the semantic sequence A BACTERIUM OR VIRUS CAUSES A DISEASE are apparent, the uses are interesting. Reflecting Hobson-West's (ibid.) findings, the pathogenic model, or 'germ' theory, of disease is sometimes represented in the *JABS* corpus as a valid explanation for the causes of some diseases but sometimes the very pretext of the pathogenic model is challenged. Where it is accepted as valid, it is often exploited in order to challenge scientific claims. For example, *JABS* corpus evidence for *hpv* (the cancer-causing human papilloma virus) shows that, while forum posters express resistance to the imperative to vaccinate against the HPV, the notion

that HPV causes cancer is typically not challenged. In some cases, though, the claim is negotiated, for example:

> [8.4] It's also **unclear if HPV "causes" the cancer** or, rather, is "associated" with it.

The ambivalence of certain vaccine-critical writers towards the pathogenic model of disease is clearly apparent, though, in the presence in the corpus of, on the one hand, claims that the live polio vaccine has been known to cause polio and, on the other hand, denials that HIV causes AIDS.

Secondly, as Hobson-West (2005) found in her data, in the *JABS* corpus, the uncertainty of the science is exploited for rhetorical purposes. It was found, for example, that vaccine-critical participants on the JABS discussion forum are as likely as pro-science posters to appeal to science or to scientific evidence. As one vaccine critical poster argues:

> [8.4] **No scientific evidence that vaccines work**.

However, vaccine-critical writers on the JABS discussion forum were found to exploit the uncertainty implicit in the use of the word *evidence* in order to counter the arguments put forward by the medical-scientific community (although sometimes the word *proof* is used in place of *evidence*). The belief is expressed that a lack of evidence of a causal connection does not mean that there is no connection.

> [8.5] **There may not be evidence to show that Mmr definetly causes autism**, but there is also **no research to show that it definetly does not**

> [8.6] Firstly **there is no scientific proof that** the MMR causes Autism as there is also **no scientific proof it doesnt**

The status of different kinds of evidence was found to be negotiated. For example, anecdotal evidence was represented by some vaccine-critical writers as equivalent to the types of scientific evidence valued by medical-science:

> [8.7] … **a huge pile of anecdotal evidence and some experimental evidence too** which supports the idea that MMR vaccination might cause autism.

Thirdly, although there are similarities in the ways in which the immune system is represented in the *JABS* and *NHSvax* corpora, some key differences were found. The collocational profiles of *immune* and *immunity* in the *JABS* corpus show some features which are characteristic of the profiles of the words in general English (as reflected in the *UkWaC* corpus) but are lacking from the *NHSvax* corpus. This suggests that the view of the immune system represented in the *JABS* corpus is more complex than that represented in the *NHSvax* corpus. The *NHSvax* texts focus on the type of immunity which functions through antibody response. A distinction is drawn between *active* and *passive immunity* (the former generated when the immune system is triggered to produce antibodies, the latter generated when proteins are introduced which mimic antibodies). In the *JABS* corpus, by contrast, a distinction is made, not between active and passive immunity, but between *cellular* and *humoral* immunity: two different aspects of immune system function. The term *natural immunity*, meanwhile, is frequent in both corpora. In the *NHSvax* corpus, the argument that it is safer to acquire immunity by catching a disease than through vaccination is rebutted, while in the environment of use of *natural immunity* in the *JABS* corpus indicates a positive evaluation of it.

Furthermore, the idea is sometimes proposed in the *JABS* corpus that healthy individuals who have a good diet and lifestyle, do not need vaccination. The concept of herd immunity is also challenged in the *JABS* corpus. There are examples of expressions such as the *myth of herd immunity*. Finally, one of the ways in which vaccination policy is resisted in the *JABS* corpus data is by putting forward the argument that multiple vaccines can overload a child's immune system. The word *OVERLOAD* occurs relatively more frequently in *NHSvax* than in *JABS*. Although the term *OVERLOAD* is used in *JABS* to put forward the proposition that multiple vaccines are potentially harmful, writers often draw scientific discourse and refer instead to *immune stimulation* or *immune activation*.

#### 8.2.3.2 Trust

A lot of literature on parents' attitudes to MMR has commented on the relationship between perceptions of vaccine-risk and trust (e.g. Casiday, 2007), while literature on vaccine-critical groups has highlighted the ways in which notions of trust are exploited (e.g. Hobson-West, 2005). One of the main ways in which a lack of trust in authority is expressed by writers represented in the *JABS* corpus is by asserting that people in positions of authority fail to admit to certain things, fail to notice certain things, or carry out actions regardless of the evidence. Terms such as *(the) real risks (of vaccination)*, for example, are used to imply that crucial information on vaccine risk has been withheld. More frequent, though, are expressions such as *FAIL to mention (the fact that)* and *despite the fact that*. Suggestions that a person, or institution, has links to pharmaceutical companies are used to express mistrust, and accusations that someone is a *(big) pharma troll* or *pharma shill* are sometimes used to deligitimize the arguments of pro-science posters. Finally, as Hobson-West (2005) finds in

her study, some of the precepts of the pathogenic model of disease are sometimes challenged, for example, the notion that HIV causes AIDS. Adding weight to the argument that debate about science in the public sphere nowadays both draws on and contributes to an interdiscursive memory bank (Beacco *et al.*, 2002; Moirand, 2003), we find references to past vaccine scares, such as the scare which followed the discovery that some batches of polio vaccine were contaminated, used in order to express mistrust of the medical-scientific authorities.

## 8.3 Methodological considerations and limitations of the study

The methodological objective which was articulated in Chapter One was to assess what opportunities and challenges are afforded by carrying out a corpus assisted discourse study of interactive web-based texts and to suggest how challenges might be addressed. The fact that a corpus-assisted approach to discourse analysis can enable the analysis of a large body of data has been demonstrated many times (e.g. Hardt-Mautner, 1995; Krishnamurthy, 1996; Stubbs, 1996; Baker and McEnery, 2005). This study has benefited from the advantages of corpus analysis. It has enabled me to undertake an analysis of a 4 million word corpus. The thesis has therefore highlighted frequent patterns of usage which are salient to the expression of beliefs about health, immunity, and risk in the *JABS* corpus, as well as uncovering patterns which typify the ways in which forum participants construct their arguments and express their evaluation of discursive objects and sources. The CADS approach, which is, in many ways, an eclectic approach, enables the analyst to draw on some of the strengths of corpus linguistics but to enhance her understanding of the data by familiarizing herself with the data. Thus, keyword analysis was used in order to approach the data with as open a mind as

possible, but prior knowledge of the contents of the *NHSvax* and *JABS* corpora and the context in which the texts were produced, together with consideration of the objectives of the thesis, guided the selection of keywords to subject to closer examination.

One of the aims of the thesis was to find out how 'lay' people on the JABS discussion forum express their beliefs and concerns. The main challenge here lay in the fact that a large proportion of the *JABS* corpus comprises uploaded media articles, or segments thereof. Corpus methods are principally designed for identifying what is 'central and typical' (Sinclair, 1991: 17). The main basic tool of corpus analysis, the concordance, is useful for bringing to light frequently occurring patterns. Given the high proportion of media texts in the *JABS* corpus, the majority of citations of many of the *JABS* keywords reflect media usage, therefore many of the frequently occurring linguistic patterns uncovered proved to be typical of media usage rather than the usage of JABS forum participants. In order to identify the voice of the individual JABS member, it was often necessary to examine the expanded context of a concordance line in order to identify the source. This was time-consuming and restricted the amount of data which could be analysed. On the positive side, consulting the expanded context of a concordance line affords one the opportunity of accessing much of the wider contextual information necessary to fully appreciate the meaning of an example of text (Widdowson, 2004).

A further challenge concerned the use of intratextual references on the JABS forum. Forum posters are able to recycle segments of text which occur in earlier posts in a given thread. It is useful to be able to distinguish the first occurrence of a recycled segment from the repeated instances. An attempt was made to address this need using automatic means. The *JABS*

*Cleanup* program was set up to mark the beginnings and ends of quoted sections into the tags <q> and </q>. As explained in Chapter Four, this was only partially successful. One challenge which was not addressed at all concerned aspects of the use of intertextual references. It would be useful for the *JABS* corpus texts to be marked up in such a way that one would easily be able to distinguish between sections of forum posts which consist of the forum poster's own words and sections which comprise uploaded texts. There was no way of doing this automatically, since no codes are used on the JABS site to mark the beginnings and ends of uploaded segments, and it would have been too time consuming to do it manually.

Finally, the project adopted a purely text-linguistic approach to analysing the corpora. As a result, certain features were overlooked which are salient to the ways in which messages are interpreted. The 'MMR The facts' and 'NHS immunisation' sites, from which much of the data for the *NHSvax* corpus was gathered, made a lot of use of graphic images. The JABS site made minimal use of graphic images at the time the *JABS* corpus was constructed and it makes little use of them today. The layout of the homepage of a website also influences the way in which messages are interpreted, while the structure of the website impacts on the way in which the user navigates the site. By basing the textual analysis solely on the contents of the text files which make up the *JABS* and *NHSvax* corpora, the meaning encoded in the graphic images was overlooked and important connections between the various pages on the websites, or between individual texts on specific pages, were lost. This is one of the limitations of using a purely text-linguistic approach rather than a multi-modal approach to textual analysis. However, by focussing solely on written text, I was able to conduct a wide-ranging analysis of a four million word corpus and a five thousand word corpus.

## 8.4 Implications for future research

The study reported here opens up avenues for future research. The *JABS* corpus represents a rich resource, as does the *NHSvax* corpus. Owing to constraints of time and space, a limited selection of keywords was subjected to full analysis. It would be desirable to examine some of the keywords which were overlooked, for example, the keywords which were classed as explicitly evaluative items. A thorough investigation of sources and attribution is desirable. Furthermore, while extended context of concordance lines was used in order to gain an idea of the wider contexts of use of specific examples, a thorough analysis of an entire text from the corpus, or several texts, would offer insight into the ways in which arguments are constructed across an extended stretch of discourse. Examination of an entire forum thread, for example, would shed light on the ways in which arguments are constructed in a multi-party conversation. As argued in the section above, carrying out a multi-modal discourse analysis of some of the website pages from which the corpus data was gathered would add a more rounded view of the ways in which meanings are encoded in the data. One further avenue of research concerns a technical issue mentioned in the section above. It would be useful to be able to mark up, accurately and thoroughly, the beginnings and ends of recycled segments of text and to examine what sort of utterances get repeated and how they are responded to.

## 8.5 Conclusion

This thesis has succeeded in its aims of discovering the ways in which claims about health, risk and immunity are expressed in the *JABS* and *NHSvax* corpora and to find out how participants on the JABS forum express their warrant for expertise. In so doing, it has contributed to enhancing knowledge of some of the ways in which debates about science are

enacted in the public sphere. It has shed light on some of the ways in which claims of a scientific nature are reformulated once they leave their source domain. In particular, it has highlighted the fact that talk about vaccine risk on an online discussion forum is characterised by a relatively high degree of evaluative language. It has also brought to light the fact that, in *JABS* data, writers frequently use *the fact that* or introductory-*it* constructions, cleft and pseudo-cleft constructions in order to introduce new information. These constructions serve as a means whereby new information is presented as given information. Use of these constructions was found frequently to be accompanied by use of evaluative language. Some of the findings of the thesis echo findings in previous work on vaccine-critical discourse and enhance such work by illustrating the discursive resources which writers draw on to construct their arguments.

# References


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

Alexander, R. J. (2010) *Framing Discourse on the Environment: A Critical Discourse Approach.* London: Routledge.

Ali, T. (2008) Where has all the rage gone? *The Guardian.* 22nd March 2008, cited in Hunston (2011) *Corpus Approaches to Evaluation: Phraseology and Evaluative Language* London: Routledge, p. 3.

Atkins, S. and Harvey, K. (2010) How to use corpus linguistics in the study of health communication. In: O'Keefe A. and McCarthy M. (eds) *The Routledge Handbook of Corpus Linguistics.* London: Routledge, 605-619.

Baker, P. (2004) 'Unnatural Acts': Discourses of homosexuality within the House of Lords debates on gay male law reform. *Journal of Sociolinguistics* 8 (1): 88-106.

Baker, P. (2006) *Using Corpora in Discourse Analysis.* London: Continuum.

Baker, P. (2012) Acceptable bias? Using corpus linguistics methods with critical discourse analysis. *Critical Discourse Studies* 9: 247-256.

Baker, P. and Ellece, S. (2011) *Key Terms in Discourse Analysis.* London: Continuum.

Baker, P., Gabrielatos, C., KhoshraviNik, M., Krzyzanowski, M., McEnery, T. and Wodak, R. (2008) A useful methodological synergy? Combining critical discourse analysis and corpus linguistics to examine discourses of refugees and asylum seekers in the UK press. *Discourse & Society* 19 (3).

Baker, P. and McEnery, T. (2005) A corpus-based approach to discourses of refugees and asylum seekers in UN and newspaper texts. *Journal of Language and Politics* 4 (2): 197-226.

Bakhtin, M. (1981) *The Dialogic Imagination: Four Essays by M. Bakhtin.* Austin: Austin University Press, reproduced in Toolan, M. (ed.) (2002) *Critical Discourse Analysis: critical concepts in linguistics Vol 1: Precursors*. London: Routledge, 40-75. .

Barnbrook, G., Mason, O. and Krishnamurthy, R. (2013) *Collocation: Applications and implications.* Basingstoke: Palgrave Macmillan.

Baroni, M., Bernardini, S., Ferraresi, A. and Zanchetta, E. (2009) The WaCky wide web: a collection of very large linguistically processed web-crawled corpora. *Language Resources and Evaluation* 43 (3): 209-226.

Barton, E. (1993) Evidentials, argumentation, and epistemological stance. *College English* 55 (7): 754-769.

Bastow, A. (2006) *Persuasion and consensus in U.S.defence speeches: a corpus study*. Unpublished PhD thesis, Birmingham: University of Birmingham.

Baxter, J. (2002) Competing discourses in the classroom: a Post-structuralist Discourse Analysis of girls' and boys' speech in public contexts. *Discourse and Society* 13 (6): 827-842.

Beacco, J.-C., Claudel, C., Doury, M., Petit, G. and Reboul-Toure, S. (2002) Science in media and social discourse: new channels of communication, new linguistic forms. *Discourse Studies* 4 (3): 277-300.

Beach, R. and Anson, C. M. (1992) Stance and Intertextuality in Written Discourse. *Linguistics and Education* 4 (3-4): 335-357.

Beck, U. (1992) *Risk Society: Towards a New Modernity.* London: Sage.

Begg, N., Ramsey, M., White, J. and Bozoky, Z. (1998) Media dents confidence in MMR vaccine. *British Medical Journal* 316: 561.

Bell, A. (1994) Climate of Opinion: Public Media and Discourse on the Global Environment. *Discourse and Society* 5 (1): 33-62.

Bellaby, P. (2003) Communication and miscommunication of risk: understanding UK parents' attitudes to combined MMR vaccination. *British Medical Journal* 327 (7417): 725–728.

Berber-Sardinha, T. (2000) Comparing corpora with WordSmith Tools: How large must the reference corpus be? *Proceedings of the Workshop on Comparing Corpora.* 7-13, available at http://ucrel.lancs.ac.uk/acl/W/W00/W00-0902.pdf, accessed 0924/0908/2011.
Berry, D. M. (2004) Internet research: privacy, ethics and alienation: an open source approach. *Internet Research* 14 (4): 323-332.
Betsch, C., Renkewitz, F., Betsch, T. and Ulshöfer, C. (2010) The Influence of Vaccine-critical Websites on Perceiving Vaccination Risks. *Journal of Health Psychology* 15 (3): 446-455.
Biber, D. (1993) Representativeness in corpus design. *Literary and Linguistic Computing* 8 (4): 243-257.
Biber, D., Connor, U. and Upton, T. A. (2007) *Discourse on the Move: Using Corpus Analysis to Describe Discourse Structure.* Amsterdam: John Benjamins.
Biber, D., Conrad, S. and Cortes, V. (2003) Lexical bundles in speech and writing: An initial taxonomy. In: Wilson A., Rayson P. and McEnery T. (eds) *Corpus Lingusitics by the Lune.* Frankfurth: Peter Lang, 71-92.
Biber, D., Conrad, S. and Cortes, V. (2004) *If you look at ...* : Lexical Bundles in University Teaching and Textbooks. *Applied Linguistics* 25 (3): 371-405.
Biber, D., Conrad, S. and Reppen, R. (1998) *Corpus Linguistics: Investigating Language Structure and Use.* Cambridge: Cambridge University Press.
Biber, D. and Finegan, E. (1988) Adverbial Stance Types in English. *Discourse Processes* 11: 1-34.
Biber, D. and Finegan, E. (1989) Styles of Stance in English: Lexical and Grammatical Marking of Evidentiality and Affect. *Text* 9 (1): 93-124.
Biber, D., Johansson, S., Leech, G., Conrad, S. and Finegan, E. (1999) *Longman grammar of spoken and written English.* London: Longman.
Blommaert, J. (2005) *Discourse.* Cambridge: Cambridge University Press.
Bondi, M. and Mauranen, A. (2003) Editorial: Evaluative Language Use in Academic Discourse. *Journal of English for Academic Purposes* 2 (4): 269-271.
Bowker, L. and Pearson, J. (2002) *Working with Specialized Language: A Practical Guide to Using Corpora.* London: Routledge.
Boyce, T. (2007) *Health, Risk and News: the MMR vaccine and the media.* New York: Peter Lang.
Brown, G. and Yule, G. (1983) *Discourse Analysis.* Cambridge: Cambridge University Press.
Burgess, D. C., Burgess, M. and Leask, J. (2006) The MMR vaccination and autism controversy in United Kingdom 1998–2005: Inevitable community outrage or a failure of risk communication? *Vaccine* 24: 3921–3928.
Burr, V. (1995) *An Introduction to Social Constructionism.* London: Routledge, cited in Baker, P. and McEnery, T. (2005) A corpus based approach to discourses of refugees and asylum seekers in UN and newspaper texts. *Journal of Language and Politics* 4(2): 197-226.
Caldas-Coulthard, C. R. and Moon, R. (2010) 'Curvy, hunky, kinky': Using corpora as tools for critical analysis. *Discourse & Society* 21 (2): 99-133.
Calsamiglia, H. and Lopez Ferrero, C. (2003) Role and position of scientific discourses: reported speech in the media. *Discourse Studies* 5 (2): 147-173.
Calsamiglia, H. and van Dijk, T. (2004) Popularization discourse and knowledge about the genome. *Discourse and Society* 15 (4): 369-389.
Candlin, C. (2005) Strategic interdiscursivity as expert practice. *38th Annual Meeting British Association for Applied Linguistics* University of Bristol, handout circulated 16 September 2005.
Candlin, C. and Candlin, S. (2003) Healthcare communication as a problematic site for applied linguistic research. *Annual Review of Applied Linguistics* 23: 134-154.
Carter, R. (1987) *Vocabulary: Applied Linguistic Perspectives.* London: Routledge.
Carter, R. (1998) *Vocabulary: Applied Linguistic Perspectives, 2nd Edition.* London: Routledge.
Carvalho, A. (2005) Representing the politics of the greenhouse effect: Discursive strategies in the British media. *Critical Discourse Studies* 2 (1): 1-29.
Casiday, R. (2005) Risk and trust in vaccine decision making. *Durham Anthropology Journal* 13 (1): http://dro.dur.ac.uk/6898/6891/6898.pdf.

Casiday, R. (2007) Children's health and the social theory of risk: insights from the British mealses, mumps and rubella (MMR) controversy. *Social Science and Medicine* 65 (5): 1059-1070.
Casiday, R., Cresswell, T., Wilson, D. and Panter-Brick, C. (2006) A survey of UK parental attitudes to the MMR vaccine and trust in medical authority. *Vaccine* 24 (2): 177-184.
Cassell, J., Leach, M., Poltorak, M., Mercer, C. H., Iversen, A. and Fairhead, J. R. (2006) Is the cultural context of MMR rejection a key to an effective public health discourse? *Public Health* 120 (9): 783-794.
Chafe, W. and Nichols, J.(eds) (1986) *Evidentiality: The Linguistic Encoding of Epistemology*. Norwood, NJ: Ablex.
Channell, J. (2000) Corpus-based Analysis of Evaluative Lexis. In: Hunston S. and Thompson G. (eds) *Evaluation in Text: Authorial Stance and the Construction of Discourse.* Oxford: Oxford University Press, 38-54.
Charles, M. (2006a) The Construction of Stance in Reporting Clauses: A Cross-disciplinary Study of Theses. *Applied Linguistics* 27 (3): 492-518.
Charles, M. (2006b) Phraseological patterns in reporting clauses used in citation: A corpus-based study of theses in two disciplines. *English For Specific Purposes* 25 (3): 310-331.
Cheng, W. (2012) *Exploring Corpus Linguistics: Language in Action.* London: Routledge.
Chiang, W. Y. and Duann, R. F. (2007) Conceptual metaphors for SARS: 'war' between whom? *Discourse and Society* 18 (5): 579-602.
Chouliaraki, L. and Fairclough, N. (1999) *Discourse in Late Modernity. Rethinking Critical Discourse Analysis.* Edinburgh: Edinburgh University Press.
Church, K. and Hanks, P. (1990) Word Association Norms, Mutual Information and Lexicography. *Computational Linguistics* 16 (1): 22-29.
Ciapuscio, G. (2003) Formulation and reformulation procedures in verbal ineractions between experts and (semi) laypersons. *Discourse Studies* 5 (2): 207-234.
Clark, C. (2010) Evidence of *evidentiality* in the quality press 1993 and 2005. *Corpora* 5 (2): 139-160.
Clarke, C. (2008) A Question of Balance The Autism-Vaccine Controversy in the British and American Elite Press. *Science Communication* 30 (1): 77-107.
Clarke, C. (2011) A case of conflicting norms? Mobilizing and accountability information in newspaper coverage of the autism-vaccine controversy. *Public Understanding of Science* 20 (5): 609-626.
Coffin, C. and O'Halloran, K. (2006) The role of appraisal and corpora in detecting covert evaluation. *Functions of Language* 13 (1): 77-110.
Conrad, S. and Biber, D. (1999) Adverbial Marking of Stance in Speech and Writing. In: Hunston S. and Thompson G. (eds) *Evaluation in Text: Authorial Stance and the Construction of Discourse.* Oxford: Oxford University Press, 56-73.
Cook, G. (2004) *Genetically Modified Language: The discourse of arguments for GM crops and food.* London: Routledge.
Cook, G., Pieri, E. and Robbins, P. (2004) 'The scientist thinks and the public feels': expert perceptions of the discourse of GM food. *Discourse and Society* 15 (4): 433-449.
Cook, G., Robbins, P. and Pieri, E. (2006) "Words of mass destruction": British newspaper coverage of the genetically modified food debate, expert and non-expert reactions. *Public Understanding of Science* 15: 5-29.
Coupland, J. and Williams, A. (2002) Conflicting discourses, shifting ideologies: pharmaceutical, 'alternative' and feminist emancipatory texts on the menopause. *Discourse & Society* 13 (4): 419-445.
Dahlberg, L. (2001) The Internet and Democratic Discourse: Exploring the prospects of online deliberative forums extending the public sphere. *Information, Communication & Society* 4 (4): 615-633.
Dahlgren, P. (2005) The Internet, Public Spheres, and Political Communication: Dispersion and Deliberation. *Political Communication* 22: 147-162.

Dales, L., Hammer, S. J. and Smith, N. J. (2001) Time Trends in Autims and in MMR Immunization Coverage in California. *JAMA: The Journal of the American Medical Association* 285 (9): 1183-1185.
Deignan, A. (2005) *Metaphor and Corpus Linguistics.* Amsterdam: John Benjamins.
DeStefano, F. and Chen, R. T. (1999) Negative association between MMR and autism. *Lancet* 353 (9169): 1987-1988.
Dixon, G. N. and Clarke, C. E. (2013) Heightening Uncertainty Around Certain Science: Media Coverage, False Balance, and the Autism-Vaccine Controversy. *Science Communication* 35 (3): 358-382.
Duguid, A. (2010a) Investigating *anti* and some reflections on Modern Diachronic Corpus-Assisted Discourse Studies (MD-CADS). *Corpora* 5 (2): 191-220.
Duguid, A. (2010b) Newspaper discourse informalisation: a diachronic comparison from keywords. *Corpora* 5 (2): 109-138.
Dunning, T. (1993) Accurate Methods for the Statistics of Surprise and Coincidence. *Computational Linguistics* 19 (1): 61-74.
Ekbom, A., Wakefield, A. J., Zack, M. and Adami, H. O. (1994) Crohn's disease following early measles exposure. *Lancet* 344: 508-510.
Ensslin, A. and Johnson, S. (2006) Language in the news: investigating representations of 'Englishness' using WordSmith Tools. *Corpora* 1 (2): 153-185.
Epstein, R. A. (2005) It Did Happen Here: Fear and Loathing On the Vaccine Trail. *Health Affairs* 24 (3): 740-743.
Evans, M., Stoddart, H., Condon, L., Freeman, E., Grizzell, M. and Mullen, R. (2001) Parents' perspectives on the MMR immunisation: a focus group study. *British Journal of General Practice* 51: 904-910.
Eysenbach, G. and Till, J. E. (2001) Ethical issues in qualitative research on internet communities. *BMJ* 323: 1103-1105.
Fahnestock, J. (1986) Accommodating Science: The Rhetorical Life of Scientific Facts. *Written Communication* 3 (3): 275-296.
Fahnestock, J. (2004) Preserving the Figure: Consistency in the Presentation of Scientific Arguments. *Written Communication* 21 (1): 6-31.
Fairclough, I. and Fairclough, N. (2011) Practical reasoning in political discourse: The UK government's response to the economic crisis in the 2008 Pre-Budget Report. *Discourse & Society* 22 (3): 243–268.
Fairclough, N. (1989) *Language and Power.* London: Longman.
Fairclough, N. (1992) *Discourse and Social Change.* Cambridge: Polity.
Fairclough, N. (1993) Critical Discourse Analysis and the Marketization of Public Discourse: The Universities. *Discourse & Society* 4 (2): 133-168.
Fairclough, N. (1995a) *Critical Discourse Analysis.* London: Longman.
Fairclough, N. (1995b) *Media Discourse.* London: Edward Arnold.
Fairclough, N. (2000) *New Labour, New Language?* London: Routledge.
Fairclough, N. (2003) *Analysing Discourse: Textual analysis for social research.* London: Routledge.
Fillmore, C. and Atkins, B. (1992) Towards a frame-based lexicon: The semantics of RISK and its neighbors. In: Lehrer A. and Kittay E. (eds) *Frames, Fields, and Contrast: New Essays in Semantics and Lexical Organization.* Hillsdale: Lawrence Erlbaum Associates, 75-102.
Firth, J. R. (1957) *Papers in Linguistics 1934-1951.* London: Oxford University Press.
Firth, J. R. (1968a) Linguistic analysis as a study of meaning. In: Palmer F. R. (ed) *Selected Papers of J. R. Firth 1952-59.* London and Harlow: Longmans, 12-26.
Firth, J. R. (1968b) A synopsis of linguistic theory 1930-55. In: Palmer F. R. (ed) *Selected Papers of J.R. Firth 1952-59.* London and Harlow: Longmans, 168-205.
Fletcher, J. (2012) *Italian MMR Case, JABS, http://www.jabs.org.uk/italian-mmr-case.html, accessed 23/04/2015.*

Fletcher, W. H. (2004) Making the web more useful as a source for linguistic corpora. In: Connor U. and Upton T. A. (eds) *Applied Corpus Linguistics: A multidimensional perspective.* Amsterdam: Rodopi, 191-206.
Fletcher, W. H. (2007) Concordancing the web: promise and problems, tools and techniques. In: Hundt M., Nesselhauf N. and Biewer C. (eds) *Corpus Linguistics and the Web.* Amsterdam: Rodopi, 25-46.
Fombonne, E. and Chakrabarti, S. (2001) No evidence for a new variant of measles-mumps-rubella-induced autism. *Pediatrics* 108 (4): e58.
Foucault, M. (2000) The Discourse on Language (1970). In: Burke L., Crowley T. and Girvin A. (eds) *The Routledge Langauge and Cultural Theory Reader*. London: Routledge.
Foucault, M. (2002) *The Arhaeology of Knowledge.* London: Routledge Classics.
Francis, G. (1986) *Anaphoric Nouns.* Birmingham: University of Birmingham English Language Research.
Francis, G. (1993) A Corpus-Driven Approach to Grammar: Principles, Methods and Examples. In: Baker M., Francis G. and Tognini-Bonelli E. (eds) *Text and Technology: In Honour of John Sinclair.* Amsterdan: John Benjamins, 137-156.
Francis, G., Hunston, S. and Manning, E. (1996) *Collins COBUILD Grammar Patterns 1: Verbs.* London: HarperCollins.
Francis, G., Hunston, S. and Manning, E. (1998) *Collins COBUILD Grammar Patterns 2: Nouns and Adjectives.* London: HarperCollins.
Freake, R. (2012) A cross-linguistic corpus-assisted discourse study of language ideologies in Canadian newspapers. *Proceedings from the Corpus Linguistics 2011 conference* Birmingham: Birmingham University Centre for Corpus Reseach, http://www.birmingham.ac.uk/research/activity/corpus/publications/conference-archives/2011-birmingham.aspx, accessed 28 August 2013.
Freake, R., Gentil, G. and Sheyholislami, J. (2011) A bilingual corpus-assisted discourse study of the construction of nationhood and belonging in Quebec. *Discourse & Society* 22 (1): 21-47.
Fudenberg, H. H. (1996) Dialysable lymphocyte extract (DLyE) in infantile onset autism: a pilot study. *Biotherapy* 9: 13-17.
Gales, T. (2009) 'Diversity' as enacted in US immigration politics and law: a corpus-based approach. *Discourse & Society* 20 (2): 223–240.
Georgakopoulou, A. and Goutsos, D. (1997) *Discourse Analysis: an introduction.* Edinburgh: Edinburgh University Press.
Giddens, A. (1991) *The Consequences of Modernity.* Cambridge: Polity Press.
Gries, S. T. (2010) Corpus Linguistics and Theoretical Linguistics: A love-hate relationship? Not necessarily ... *International Journal of Corpus Linguistics* 15 (3): 327-343.
Groom, N. (2007) *Phraseology and epistemology in humanities writing: a corpus-driven study.* Unpulished PhD Thesis. Birmingham: University of Birmingham.
Grundman, R. and Cavaillé, J. P. (2000) Simplicity in Science and its Publics. *Science as Culture* 9 (3): 353-389.
Grundmann, R. and Krishnamurthy, R. (2010) The Discourse of Climate Change: A Corpus Based Approach. *Critical Approaches to Discourse Analysis across Disciplines* 4 (2): 125-146.
Gűlich, E. (2003) Conversational Techniques Used in Transferring Knowledge between Medical Experts and Non-experts. *Discourse Studies* 5 (2): 235-263.
Habermas, J. (1989) *The Structural Transformation of the Public Sphere: An Inquiry into a Category of Bourgeois Society.* Cambridge: Polity, translated by Thomas Burger with the Assistance of Frederick Lawrence.
Halliday, M. A. K. (1985) *An Introduction to Functional Grammar.* London: Edward Arnold.
Halliday, M. A. K. (1994) *An Introduction to Functional Grammar 2nd Edition.* London: Edward Arnold.
Halliday, M. A. K. (1998) Things and relations: Regrammaticising experience as technical knowledge. In: Martin J. R. and Veel R. (eds) *Reading Science: Critical and Functional Perspectives on Discourses of Science.* London: Routledge, 185-235.

Halliday, M. A. K. and Hasan, R. (1976) *Cohesion in English.* London: Longman.
Hamilton, C., Adolphs, S. and Nerlich, B. (2007) The meanings of 'risk': a view from corpus linguistics. *Discourse & Society* 18 (2): 163-181.
Hanks, P. (2013) *Lexical Analysis: Norms and Exploitations.* Cambridge, Mass and London: The MIT Press.
Hardt-Mautner, G. (1995) Only Connect. Critical Discourse Analysis and Corpus Linguistics. *UCREL Technical Paper 6.* Lancaster: Lancaster University.
Harvey, K. (2012) Disclosures of depression: Using corpus linguistics methods to examine young people's online health concerns *International Journal of Corpus Linguistics* 17 (3): 349-379.
Harvey, K., Brown, B., Crawford, P., Macfarlane, A. and McPherson, A. (2007) 'Am I normal?' Teenagers, sexual health and the internet. *Social Science and Medicine* 65: 771–781.
Harvey, K., Locher, M. A. and Mullany, L. (2013) "Can I Be at Risk of Getting AIDS?" A linguistic analysis of two Internet advice columns on sexual health. *Linguistik online* 59 (2/13): 113-134.
Hasan, R. (2000) The uses of talk. In: Sarangi S. and Coulthard M. (eds) *Discourse and Social Life.* London: Longman, 28-47.
Heffernan, C., Misturelli, F. and Thomson, K. (2011) The representation of highly pathogenic avian influenza in the Chinese media. *Health, Risk and Society* 13 (7-8): 603–620.
Hellsten, I. (2003) Focus on Metaphors: The Case of "Frankenfood" on the Web. *Journal of Computer Mediated Communication* 8 (4): http://jcmc.indiana.edu/vo8/issue4/hellsten.html.
Henderson, A., Weaver, C. K. and Cheney, G. (2007) Talking 'facts': identity and rationality in industry perspectives on genetic modification. *Discourse Studies* 9 (1): 9-41.
Herring, S. (1997) Ethics in Cyberresearch: To cite, or not to cite? *The College* 1 (2): 18-23, available at http://ella.slis.indiana.edu/~herring/cola.1997.pdf, accessed 1920/1905/2010.
Hilgartner, S. (1990) The Dominant View of Popularization: Conceptual Problems, Political Uses. *Social Studies of Science* 20 (3): 519-539.
Hilton, S., Hunt, K., Langan, M., Bedford, H. and Petticrew, M. (2010) Newsprint media representations of the introduction of the HPV vaccination programme for cervical cancer prevention in the UK (2005-2008). *Social Science and Medicine* 70 (6): 942-950.
Hilton, S., Petticrew, M. and Hunt, K. (2006) 'Combined vaccines are like a sudden onslaught to the body's immune system': parental concerns about vaccine 'overload' and 'immune-vulnerability'. *Vaccine* 24 (20): 4321-4327.
Hinyard, L. and Kreuter, M. (2007) Using Narrative Communication as a Tool for Health Behavior Change: A Conceptual, Theoretical, and Empirical Overview. *Health Education & Behavior* 34 (5): 777-792.
Hobson-West, P. (2003) Understanding vaccination resistance: moving beyond risk. *Health, Risk and Society* 5 (3): 273-283.
Hobson-West, P. (2005) *Understanding resistance to childhood vaccination in the UK: Radicals, Reformists and the discourses of risk, trust and science.* Unpublished PhD Thesis. Nottingham: Nottingham University.
Hobson-West, P. (2007) 'Trusting blindly can be the biggest risk of all': organised resistance to childhood vaccination in the UK. *Sociology of Health and Illness* 29 (2): 198-215.
Hodson-Champeon, C. (2010) Conversations within Conversations: Intertextuality in Racially Antagonistic Online Discourse. *Language@Internet* 7, http://www.languageatinternet.de, accessed 12/11/2011.
Hoey, M. (1983) *On the Surface of Discourse.* London: Allen & Unwin.
Hoey, M. (1991) *Patterns of Lexis in Text.* Oxford: Oxford University Press.
Hoey, M. (1994) Signalling in discourse: a functional analysis of a common discourse pattern in written and spoken English. In: Coulthard M. (ed) *Advances in Written Text Analysis.* London: Routledge, 26-45.
Hoey, M. (2005) *Lexical Priming: A New Theory of Words and Language.* London: Routledge.
Holmes, J. (1988) Doubt and uncertainty in ESL textbooks. *Applied Linguistics* 9 (1): 20-44.

Holton, A., Weberling, B., Clarke, C. and Smith, M. (2012) The Blame Frame: Media Attribution of Culpability About the MMR–Autism Vaccination Scare. *Health Communication* 27 (7): 690-791.
Hunston, S. (1994) Evaluation and organization in a sample of written academic discourse. In: Coulthard M. (ed) *Advances in Written Text Analysis.* London: Routeledge, 191-218.
Hunston, S. (2000) Evaluation and the Planes of Discourse: Status and Value in Persuasive Texts. In: Hunston S. and Thompson G. (eds) *Evaluation in Text: Authorial Stance and the Construction of Discourse.* Oxford: Oxford University Press, 176-206.
Hunston, S. (2002) *Corpora in Applied Linguistics.* Cambridge: Cambridge University Press.
Hunston, S. (2007) Semantic prosody revisited. *International Journal of Corpus Linguistics* 12 (2): 249-268.
Hunston, S. (2008) Starting with the small words: Patterns, lexis and semantic sequences. *International Journal of Corpus Linguistics* 13 (3): 271-295.
Hunston, S. (2011) *Corpus Approaches to Evaluation: Phraseology and Evaluative Language.* London: Routledge.
Hunston, S. and Francis, G. (1999) *Pattern Grammar: A Corpus-driven Approach to the Lexical Grammar of English.* Amsterdam: John Benjamins.
Hunston, S. and Thompson, G.(eds) (2000) *Evaluation in Text: Authorial Stance and the Construction of Discourse.* Oxford: Oxford University Press.
Hyland, K. (1996a) Talking to the academy: Forms of hedging in science research articles. *Written Communication* 13 (2): 251-281.
Hyland, K. (1996b) Writing Without Conviction? Hedging in Science Research Articles. *Applied Linguistics* 17 (4): 433-454.
Hyland, K. (1998) 'Writing without conviction?' Hedging in scientific research articles. *Applied Linguistics* 17 (4): 433-454.
Hyland, K. (1999) Disciplinary Discourses: Writer Stance in Research Articles In: Candlin C. and Hyland K. (eds) *Writing Texts: Processes and Practices.* London: Longman, 99-121.
Hyland, K. (2004) Engagement and Disciplinarity: the other side of evaluation. In: del Lungo Camiciotti G. and Tognini-Bonelli E. (eds) *Academic Discourse: New Insights into Evaluation.* Amsterdam: Peter Lang, 13-30.
Hyland, K. (2005) Stance and engagement: a model of interaction in academic discourse. *Discourse Studies* 7 (2): 173-192.
Hyland, K. (2007) Stance and Engagement: a Model of Interaction in Academic Discourse. In: van Dijk T. (ed) *Benchmarks in Discourse Studies Vol. 3.* Thousand Oaks, CA: Sage, 102-121.
Hyland, K. (2010) Constructing proximity: Relating to readers in popular and professional science. *Journal of English for Academic Purposes* 9 (2): 116-127.
Hyland, K. and Tse, P. (2004) Metadiscourse in Academic Writing: A Reappraisal. *Applied Linguistics* 25 (2): 156-177.
Jaworska, S. and Krishnamurthy, R. (2012) On the F word: A corpus-based analysis of the media representation of feminism in British and German press discourse, 1990–2009. *Discourse & Society* 23 (2): 401–431.
Jones, R. (2013) *Health and Risk Communication: An Applied Linguistic Perspective*. London: Routledge.
Joye, S. (2010) News discourses on distant suffering: a Critical Discourse Analysis of the 2003 SARS outbreak. *Discourse and Society* 21 (5): 586-610.
Kaplan, S. and Garrick, B. J. (1981) On the Quantitative Definition of Risk. *Risk Analysis* 1 (1): 11-27.
Kennedy, G. (2007) An under-exploited resource: using the BNC for exploring the nature of language learning. In: Hundt M., Nesselhauf N. and Biewer C. (eds) *Corpus Linguistics and the Web.* Amsterdam: Rodopi, 152-166.
Kilgarriff, A. (1997) Using Word Frequency Lists to Measure Corpus Homogeneity and Similarity between Corpora. . *Proceedings of Workshop on Language Engineering for Document Analysis and Recognition. Proceedings, AISB Workshop, Falmer.* Ulm: Falmer, 231-245, available at http://ucrel.lancs.ac.uk/acl/W/W297/W297-0122.pdf, accessed 0125/0108/2011.

Kilgarriff, A. (2005) Language is never, ever, ever random. *Corpus Linguistics and Linguistic Theory* 1 (2): 265-275.
Kilgarriff, A. (2007) Last Word: Googleology is bad science. *Computational Linguistics* 33 (1): 147-151.
Kitzinger, J. (1999) Researching risk and the media. *Health, Risk and Society* 1 (1): 55-69.
Krishnamurthy, R. (1996) Ethnic, Racial and Tribal: the Langauge of Racism? In: Caldas-Coulthard C. R. and Coulthard M. (eds) *Texts and Practices: Readings in Critical Discourse Analysis.* London: Routledge, 129-149.
Labov, W. (1972) *Langauge in the Inner City.* Philadelphia: University of Philadelpia.
Labov, W. (1984) Intensity. In: Schiffrin D. (ed) *Meaning, Form, and Use in Context: Linguistic Applications.* Washington, DC: Georgetown University Press.
Leask, J., Chapman, S., Hawe, P. and Burgess, M. (2006) What maintains parental support for vaccination when challenged by anti-vaccination messages? A qualitative study. *Vaccine* 24 (49-50): 7238-7245.
Lee, D. (1992) *Competing Discourses: Perspective and Ideology in Language* London: Longman.
Leech, G. (2005) Adding Linguistic Annotation. In: Wynne M. (ed) *Corpus Lingusitics and the Web.* Amsterdam: Rodopi, 17-29.
Leech, G. (2007) New resources, or just better old ones? The Holy Grail of representativeness. In: Hundt M., Nesselhauf N. and Biewer C. (eds) *Corpus Linguistics and the Web.* Amsterdam: Rodopi, 133-150.
Leitch, S. and Davenport, S. (2007) Strategic ambiguity as a discourse practice: the role of keywords in the discourse on 'sustainable' biotechnology. *Discourse Studies* 9 (1): 43-61.
Lewis, J. and Speers, T. (2003) Misleading media reporting? The MMR story. *Nature Reviews: Immunology* 3 (11): 913-918.
Louw, B. (1993) Irony in the Text or Insincerity in the Author?: The Diagnostic Potential of Semantic Prosodies. In: Baker M., Francis G. and Tognini-Bonelli E. (eds) *Text and Technology: In Honour of John Sinclair.* Amsterdam: John Benjamins.
Lupton, D. (1999) *Risk.* London: Routledge.
Luzon, M. J. (2013) Public Communication of Science in Blogs: Recontextualizing Scientific Discourse for a Diverse Audience. *Written Communication* 30 (4): 248-457.
Maingueneau, D. (1999) Analysing self-consituting discourses. *Discourse Studies* 1 (2): 175-199.
Marchi, A. (2010) 'The moral in the story': a diachronic investigation of lexicalised morality in the UK press. *Corpora* 5 (2): 161-189.
Martin, J. R. (2000) Beyond exchange: APPRAISAL systems in English. In: Hunston S. and Thompson G. (eds) *Evaluation in Text: Authorial Stance and the Construction of Discourse.* Oxford: Oxford University Press, 142-175.
Martin, J. R. and White, P. R. R. (2005) *The Language of Evaluation: Appraisal in English.* Basingstoke: Palgrave Macmillan.
Martin Rojo, L. and Galibondo Pujol, A. (2000) Michel Foucault. In: Verschueren J., Östman J. O., Blommaert J., *et al.* (eds) *Handbook of Pragmatics 2000.* Amsterdam: John Benjamins, 1-24.
Mautner, G. (2005) The entrepreneurial university. A discursive profile of a higher education buzzword. *Critical Discourse Studies* 2 (2): 95-120.
Mautner, G. (2007) Mining large corpora for social information: The case of *elderly*. *Language in Society* 36 (1): 51-72.
Mautner, G. (2009) Corpora and Critical Discourse Analysis. In: Baker P. (ed) *Contemporary Corpus Linguistics.* London: Continuum.
McDonald, M. and Hunter, D. (2013) The discourse of Olympic security: London 2012. *Discourse & Society* 24 (1): 66-88.
McEnery, T. and Hardie, A. (2012) *Corpus Lingusitics: Method, Theory and Practice.* Cambridge: Cambridge University Press.
McEnery, T. and Wilson, A. (2001) *Corpus Linguistics: An Introduction.* Edinburgh: Edinburgh University Press.

McEnery, T., Xiao, R. and Tono, Y. (2006) *Corpus -based Language Studies: An Advanced Resource Book.* London: Routledge.
Meyer, C. F. (2002) *English Corpus Linguistics: An Introduction.* Cambridge: Cambridge University Press.
Millar, N. and Budgell, B. S. (2008) The language of public health: a corpus-based analysis. *Journal of Public Health* 16: 369-374.
Miller, D., Wadsworth, J., Diamond, J. and Ross, E. (1997) Measles vaccination and neurological events. *Lancet* 349: 730-731.
Moirand, S. (2003) Communicative and Cognitive Dimensions of Discourse on Science in the French Mass Media. *Discourse Studies* 5 (2): 175-206.
Morgan, M. (2007) How well do "facts" travel? Paper read at *Enquiry, Evidence and Facts: an interdisciplinary conference. In association with the Leverhulme Trust / ESRC funded research programme s on 'The Nature of Evidence' at UCL and LSE. The British Academy, December 2007.* Cited in Hunston, S. (2011) *Corpus Approaches to Evaluation: Phraseology and Evaluative Language* London: Routledge, pp. 116-117.
Motion, J. and Doolin, B. (2007) Out of the laboratory: scientists' discursive practices in their encounters with activists. *Discourse Studies* 9 (1): 63-85.
Myers, G. (1994) Narratives of science and nature in popularizing molecular genetics. In: Coulthard M. (ed) *Advances in Written Text Analysis.* London: Routledge, 179-190.
Myers, G. (2003) Discourse studies of scientific popularization: questioning the boundaries. *Discourse Studies* 5 (2): 265-279.
Myers, G. (2004) *Matters of Opinion: Talking about public issues.* Cambridge: Cambridge University Press.
Myers, G. (2007) Commonplaces in Risk Talk: Face Threats and Forms of Interaction. *Journal of Risk Research* 10 (3): 285-305.
Myers, G. (2010) Stance-taking and public discussion in blogs. *Critical Discourse Studies* 7 (4): 263-275.
Mythen, G. (2004) *Ulrich Beck: a critical introduction to the risk society.* London: Pluto Press.
O'Dell, L. and Brownlow, C. (2005) Media reports of links between MMR and autism: a discourse analysis. *British Journal of Learning Disabilities* 33: 194-199.
O'Halloran, K. (2007) Critical Discourse Analysis and the Corpus-informed Interpretation of Metaphor at the Register Level. *Applied Linguistics* 28 (1): 1-24.
O'Halloran, K. (2009) Inferencing and Cultural Reproduction: a corpus-based critical discourse analysis. *Text and Talk* 29 (1): 21-52.
Ochs, E. (1989) Introduction: The Pragmatics of Affect. *Special Issue of Text* 9 (1).
Orpin, D. (2005) Corpus Linguistics and Critical Discourse Analysis: Examining the ideology of sleaze. *International Journal of Corpus Linguistics* 10 (1): 37-61.
Pang, B. and Lee, L. (2008) Opinion mining and sentiment analysis. *Foundations and Trends in Information Retrieval* 2 (1-2): 1-135.
Pareek, M. and Pattison, H. M. (2000) The two-dose measles, mumps and rubella (MMR) immunisation schedule : factors affecting maternal intendtion to vaccinate. *British Journal of Medical Practice* 50 (461): 969-971.
Partington, A. (1998) *Patterns and Meanings: Using Corpora for English Language Research and Teaching.* Amsterdam: John Benjamins.
Partington, A. (2003) *The Linguistics of Political Argument: The spin-doctor and the wolf-pack at the White House.* London: Routledge.
Partington, A. (2004a) Corpora and discourse, a most congruous beast. In: Partington A., Morley J. and Haarman L. (eds) *Corpora and Discourse.* Bern: Peter Lang, 11-20.
Partington, A. (2004b) “Utterly content in each other’s company”: Semantic prosody and semantic preference. *International Journal of Corpus Linguistics* 9 (1): 131-156.
Partington, A. (2008) The Armchair and the Machine: Corpus-Assisted Discourse Research. In: Taylor-Torsello C., Ackerley K. and Castello E. (eds) *Corpora for University Language Teachers.* Bern: Peter Lang, 95-118.

Partington, A. (2010) Modern Diachronic Corpus-Assisted Discourse Studies (MD-CADS) on UK newspapers: an overview of the project. *Corpora* 5 (2): 83-108.
Partington, A., Duguid, A. and Taylor, C. (2013) *Patterns and Meanings in Discourse: Theory and practice in corpus-assisted discourse studies (CADS).* Amsterdam: John Benjamins.
Peltola, H., Patja, A., Leinikki, P., Valle, M., Davidkin, I. and Paunio, M. (1998) No evidence for measles, mumps, and rubella vaccine-associated inflammatory bowel disease or autism in a 14-year prospective study. *Lancet* 351 (9112): 1327-1328.
Petts, J. and Niemeyer, S. (2004) Health risk communication and amplification: learning from the MMR vaccination controversy. *Health, Risk and Society* 6 (1): 7-23.
Petts, J., Wheeley, S., Homan, J. and Niemeyer, S. (2003) Risk Literacy and the Public: MMR, Air Pollution and Mobile Phones - Final Report for the Deparment of Health. Birmingham: Birmingham University.
Poltorak, M., Leach, M., Fairhead, J. and Cassell, J. (2005) 'MMR talk' and vaccination choices: An ethnographic study in Brighton. *Social Science and Medicine* 61 (3): 709-719.
Potter, J. (1996) *Representing Reality: Discourse, Rhetoric and Social Construction.* London: Sage.
Prentice, S., Rayson, P. and Taylor, P. J. (2012) The language of Islamic extremism: Towards an automated identification of beliefs, motivations and justifications. *International Journal of Corpus Linguistics* 17 (2): 259-286.
Ramsay, M. E., Yarwood, J., Lewis, D., Campbell, H. and White, J. M. (2002) Parental confidence in measles, mumps and rubella vaccine: evidence from vaccine coverage and attitudinal surveys. *British Journal of General Practice* 52: 912-916.
Rayson, P. and Garside, R. (2000) Comparing Corpora using Frequency Profiling. *Proceedings of the workshop on Comparing Corpora, held in conjunction with the 38th annual meeting of the Association for Computational Linguistics.* Hong Kong, 1-6, available at http://ucrel.lancs.ac.uk/llwizard.html, accessed 16/06/2013.
Renouf, A., Kehoe, A. and Banerjee, J. (2005) The WebCorp Search Engine: a holistic approach to Web Text Serach. *Electronic Proceedings of CL2005.* Birmingham: University of Birmingham.
Renouf, A., Kehoe, A. and Banerjee, J. (2007) WebCorp: an integrated system for web text search. In: Hundt M., Nesselhauf N. and Biewer C. (eds) *Corpus Linguistics and the Web* Amsterdam: Rodopi, 47-68.
Richardson, K. (2001) Risk in the world of Internet newsgroups. *Journal of Sociolinguistics* 5 (1): 50-72.
Richardson, K. (2003) Health risks on the internet: establishing credibility on line. *Health, Risk and Society* 5 (2): 171-184.
Richardson, K. (2005) *Internet Discourse and Health Debates.* Basingstoke: Palgrave Macmillan.
Riesigl, M. and Wodak, R. (2001) *Discourse and Discrimination.* London: Routledge.
Rundblad, G., Chilton, P. and Hunter, P. (2006) An inquiry into scientific and media discourse in the MMR controversy. *Communication and Medicine* 3 (1): 69-80.
Sacks, H., Schegloff, E. A. and Jefferson, G. (1974) A simplest systematics for the organization of turn-taking for conversation. *Language* 50: 696-735.
Sarangi, S. and Roberts, C.(eds) (1999) *Talk, Work and Institutional Order: Discourse in Medical, Mediation and Management Settings* Berlin: Mouton de Gruyter.
Scott, M. (1997) PC Analysis of Key Words - and Key Key Words. *System* 25: 233-245.
Scott, M. (2004) *WordSmith Tools version 4.* Oxford: Oxford University Press.
Scott, M. (2012) *WordSmith Tools version 6.* Liverpool: Lexical Analysis Software.
Scott, M. and Tribble, C. (2006) *Textual Patterns: Key words and corpus analysis in language education.* Amsterdam: John Benjamins.
Seale, C. (2002) *Media and Health.* London: Sage.
Seale, C. (2006) Gender accommodation in online cancer support. *Health: An Interdisciplinary Journal for the Social Study of Health, Illness and Medicine* 10 (3): 345-360.

Seale, C., Charteris-Black, J., Macfarlane, A. and McPherson, A. (2010) Interviews and Internet Forums: A Comparison of Two Sources of Qualitative Data. *Qualitative Health Research* 20 (5): 595-606.
Sealey, A. (2010) Probabilities and Surprises: A Realist Approach to Identifying Linguistic and Social Patterns, with Reference to an Oral History Corpus. *Applied Linguistics* 31 (2): 215-235.
Sealey, A. (2012) 'I just couldn't do it': representations of constraint in an oral history corpus. *Critical Discourse Studies* 9 (3): 195-210.
Sealey, A. (2014) Cats and categories - reply to Teubert. *Language and Dialogue* 4 (2): 299-321.
Sealey, A. and Carter, A. (2004) *Applied Linguistics as Social Science.* London: Continuum.
Sealey, A. and Oakley, L. (2014) Why did the Canada goose cross the sea? Accounting for the behaviour of wildlife in the documentary series *Life*. *International Journal of Applied Linguistics* 24 (1): 19-37.
Shah, K. V. (2004) Siminan Virus 40 and Human Disease. *Journal of Infectious Diseases* 190 (12): 2061-2064.
Sharrock, W. W. and Anderson, D. C. (1981) Language, thought and reality again. *Sociology* 15: 287-293.
Simpson, P. (1993) *Language, Ideology and Point of View.* London: Routledge.
Sinclair, J. (1981) Planes of Discourse'. In: Rizvi S. N. A. (ed) *The Two-Fold Voice: Essays in Honour of Ramesh Mohan* Salzburg: University of Salzburg, 70-89.
Sinclair, J. (1986) Fictional Worlds. In: Coulthard M. (ed) *Talking About Text: Studies Presented to David Brazil on his Retirement.* Birmingham: Birmingham University English Language Research, 43-60.
Sinclair, J. (1991) *Corpus, Concordance, Collocation.* Oxford: Oxford University Press.
Sinclair, J. (2004) *Trust the Text: language, corpus and discourse.* London: Routledge.
Sinclair, J. (2005a) Appendix: How to build a corpus. In: Wynne M. (ed) *Developing Linguistics Corpora: A Guide to Good Practice.* Oxford: Oxbow Books, 79-83.
Sinclair, J. (2005b) Corpus and Text: Basic Principles. In: Wynne M. (ed) *Developing Linguistic Corpora: A Guide to Good Practice.* Oxford: Oxbw Books, 1-16.
Sinclair, J. and Coulthard, M. (1975) *Towards an Analysis of Discourse: the English Used by Teachers and Pupils.* Oxford: Oxford University Press.
Sinclair, J., Fox, G., Bullon, S. and Manning, E. (1995) *Collins COBUILD English Dictionary.* London: HarperCollins.
Sinclair, J., Jones, S. and Daley, R. (2004) *English Collocation Studies: The OSTI Report.* London: Contimuum.
Skea, Z. C., Entwistle, V., Watt, I. and Russell, E. (2008) 'Avoiding harm to others' considerations in relation to parental MMR vaccination discussions : an analysis of an online chat forum. *Social Science and Medicine* 67 (9): 1382-1390.
Solin, A. (2004) Intertextuality as mediation: On the analysis of intertextual relations in public discourse. *Text* 24 (2): 267-296.
Sotillo, S. M. and Wang-Gempp, J. (2004) Using Corpus Linguistics to Investigate Class, Ideology, and Discursive Practices in Online Political Discussions: Pedagogical Applications of Corpora. In: Connor U. and Upton T. A. (eds) *Applied Corpus Linguistics.* Amsterdam: Rodopi, 91-122.
Speers, T. and Lewis, J. (2004) Journalists and jabs: media coverage of the MMR vaccine. *Communication and Medicine* 1 (2): 171-181.
Stubbs, M. (1983) *Discourse Analysis: The Sociolinguistic Analysis of Natural Language.* Oxford: Blackwell.
Stubbs, M. (1996) *Text and Corpus Analysis: Computer-assisted Studies of Language and Culture.* Oxford: Blackwell.
Stubbs, M. (1997) Whorf's Children: Critical comments on Critical Discourse Analysis. In: Ryan A. and Wray A. (eds) *Evolving Models of Language.* Clevedon: BAAL in association with Multilingual Matters, 110-116.

Stubbs, M. (2001) Texts, corpora, and problems of interpretation: a response to Widdowson. *Applied Linguistics* 22 (2): 149-172.
Stubbs, M. (2002) *Words and Phrases: Corpus Studies of Lexical Semantics.* Oxford: Blackwell.
Stubbs, M. (2003) Two quantitative methods of studying phraseology in English. *International Journal of Corpus Linguistics* 7 (2): 215-244.
Stubbs, M. (2010) Three Concepts of Keyness. In: Bondi M. and Scott M. (eds) *Keyness in Texts.* Amsterdam: John Benjamins, 21-42.
Stubbs, M. and Gerbig, A. (1993) Human and Inhuman Geography: on the computer-assisted analysis of long texts. In: Hoey M. (ed) *Data, Description, Discourse: Papers on the English Lanaguage in Honour of John McH Sinclair.* London: HarperCollins, 64-85.
Tadros, A. (1985) *Prediction in Text.* Birmingham: Birmingham University English Language Research.
Taylor, B., Miller, E., Farrington, P., Petropoulos, M. C., Favot-Mayaud, I., Li, J. and Waight, P. A. (1999) Autism and measles, mumps, and rubella vaccine: no epidemiological evidence for a causal association. *Lancet* 353 (9169): 1987-1988.
Taylor, C. (2010) Science in the news: a diachronic perspective. *Corpora* 5 (2): 221-250.
Taylor, C. (2013) Searching for similarity using corpus-assisted discourse studies. *Corpora* 8 (1): 81-113.
Teubert, W. (2004) Language and Corpus Linguistics. In: Halliday M. A. K., Teubert W., Yallop C., *et al.* (eds) *Lexicology and Corpus Linguistics: An Introduction.* London: Continuum, 73-112.
Teubert, W. (2005) My version of corpus linguistics. *International Journal of Corpus Linguistics* 10 (1): 1-13.
Teubert, W. (2013) Was there a cat in the garden?: Knowledge between discourse and the monadic self. *Language and Dialogue* 3 (2): 273-297.
Thelwall, M. and Buckley, K. (2013) Topic-Based Sentiment Analysis for the Social Web: The Role of Mood and Issue Related Words. *Journal of the American Society for Information Science and Technology* 64 (8): 1608–1617.
Thelwall, M., Buckley, K. and Paltoglou, G. (2012) Sentiment strength detection for the social Web. *Journal of the American Society for Information Science and Technology* 63 (1): 163-173.
Thompson, N., Montgomery, S., Pounder, R. E. and Wakefield, A. J. (1995) Is measles vaccination a risk factor for inflammatory bowel diseases? *Lancet* 345 (8957): 1071-1074.
Tognini-Bonelli, E. (2001) *Corpus Linguistics at Work.* Amsterdam: John Benjamins.
van Dijk, T. (1991) *Racism and the Press.* London: Routledge.
van Dijk, T. (1993) Principles of Critical Discourse Analysis. *Discourse & Society* 4 (2): 249-283.
van Dijk, T.(ed) (1997a) *Discourse as Social Interaction.* London: Sage.
van Dijk, T.(ed) (1997b) *Discourse as Structure and Process.* London: Sage.
van Dijk, T. (2003) Critical Discourse Analysis. In: Schiffrin D., Tannen D. and Hamilton H. (eds) *The Handbook of Discourse Analysis 2nd Edition.* London: Blackwell, 352-371.
Vayreda, A. and Antaki, C. (2011) To vaccinate or not? The disqualification of commercial sources of health advice in an online forum. *Communication and Medicine* 8 (3): 273-288.
Veel, R. (1998) The greening of school science: Ecogenesis in secondary school classrooms. In: Martin J. R. and Veel R. (eds) *Reading Science: Critical and Functional Perspectives on Discourses of Science.* London: Routledge, 114-151.
Verschueren, A. (1999) *Understanding Pragmatics.* London: Arnold.
Wakefield, A. J., Murch, S. H., Anthony, A., Linnell, J., Casson, D. M., Malik, M., Berelowitz, M., Dhillon, A. P., Thomson, M. A., Harvey, P., Valentine, A., Davies, S. E. and Walker-Smith, J. A. (1998) Ileal-lymphoid-nodular hyperplasia, non-specific colitis, and pervasive developmental disorder in children. *Lancet* 351 (9103 ): 637-641.
Wallis, P. and Nerlich, B. (2005) Disease metaphors in new epidemics: the UK media framing of the 2003 SARS epidemic. *Social Science and Medicine* 60: 2629-2639.
Weberling McKeever, B. (2013) News Framing of Autism: Understanding Media Advocacy and the Combating Autism Act. *Science Communication* 35 (2): 213–240.

Wetherell, M., Taylor, S. and Yates, S.(eds) (2001) *Discourse as Data: A Guide to Analysis.* London: Sage.
White, P. R. R. (1997) Death, Disruption and the Moral Order: the Narrative Impulse in Mass-Media Hard News Reporting. In: Christie F. and Martin J. R. (eds) *Genres and Institutions: Social Processes in the Workplace and School.* London: Cassell, 101-133.
Whitsitt, S. (2005) A critque of the concept of semantic prosodies. *International Journal of Corpus Linguistics* 10 (3).
Widdowson, H. G. (1995a) Discourse Analysis: A critical view. *Language and Literature* 4 (3): 157-152.
Widdowson, H. G. (1995b) Review of Fairclough Discourse and Social Change. *Applied Linguistics* 16 (4): 510-516.
Widdowson, H. G. (1996) Discourse and Interpretation: Conjectures and refutations [Reply to Fairclough, 1996]. *Language and Literature* 5 (1): 57-69.
Widdowson, H. G. (2000) On the limitations of linguistics applied. *Applied Linguistics* 21 (1): 3-25.
Widdowson, H. G. (2004) *Text, Context, Pretext: Critical Issues in Discourse Analysis.* Oxford: Blackwell.
Wild, K., Church, A., McCarthy, D. and Burgess, J. (2013) Quantifying lexical usage: vocabulary pertaining to ecosystems and the environment. *Corpora* 8 (1): 53–79.
Williams, R. (1976) *Keywords: A Vocabulary of Culture and Society.* London: Fontana.
Winter, E. (1977) A clause relational approach to English texts: A study of some predictive lexical items in written discourse. *Instructional Science* 6 (1): 1-92.
Wodak, R. (1996) *Disorders of Discourse.* London: Longman.
Wodak, R., De Cilia, R., Riesigl, M. and Liebhart, K. (1999/2009) *The Discursive Construction of National Identity.* Edinburgh: Edinburgh University Press.
Wodak, R. and Riesigl, M. (1999) *Discourse and Discrimination. The Rhetorics of Racism and Antisemitism.* London: Routledge.
Wodak, R. and Riesigl, M. (2001) Discourse and Racism. In: Schiffrin D., Tannen D. and Hamilton H. (eds) *The Handbook of Discourse Analysis.* Oxford: Blackwell, 372-397.
Wooffitt, R. (2005) *Conversation Analysis and Discourse Analysis: A Comparative and Critical Introduction.* London: Sage.
Xuelan, F. and Kennedy, G. (1992) Expressing Causation in Written English. *RELC Journal* 23: 62-80.
Yamaguchi, T. (2007) Controversy over genetically modified crops in India: discursive strategies and social identities of farmers. *Discourse Studies* 9 (1): 87-107.
Zimmerman, R. K., Wolfe, R. M., Fox, D. E., Fox, J. R., Nowalk, M. P., Troy, J. A. and Sharp, L. K. (2005) Vaccine Criticism on the World Wide Web. *Journal of Medical Internet Research* 7 (2): e17.

**Appendix 1** The 200 most frequent words in the *NHSvax* corpus

| Rank | Word | Frequency | Percentage of corpus | No. of texts | Percentage of texts |
|---|---|---|---|---|---|
| 1 | THE | 31453 | 5.0669 | 290 | 99.315 |
| 2 | # | 28633 | 4.6126 | 284 | 97.26 |
| 3 | OF | 19112 | 3.0789 | 279 | 95.548 |
| 4 | AND | 14542 | 2.3426 | 282 | 96.575 |
| 5 | TO | 14104 | 2.2721 | 277 | 94.863 |
| 6 | IN | 11744 | 1.8919 | 272 | 93.151 |
| 7 | A | 9851 | 1.587 | 265 | 90.753 |
| 8 | IS | 7355 | 1.1849 | 249 | 85.274 |
| 9 | VACCINE | 7228 | 1.1644 | 246 | 84.247 |
| 10 | FOR | 5991 | 0.9651 | 260 | 89.041 |
| 11 | THAT | 5520 | 0.8892 | 239 | 81.849 |
| 12 | BE | 5209 | 0.8391 | 229 | 78.425 |
| 13 | ARE | 4220 | 0.6798 | 222 | 76.027 |
| 14 | OR | 4128 | 0.665 | 227 | 77.74 |
| 15 | THIS | 3905 | 0.6291 | 240 | 82.192 |
| 16 | MMR | 3677 | 0.5923 | 194 | 66.438 |
| 17 | WITH | 3664 | 0.5903 | 231 | 79.11 |
| 18 | IMMUNISATION | 3447 | 0.5553 | 210 | 71.918 |
| 19 | HAVE | 3337 | 0.5376 | 218 | 74.658 |
| 20 | ON | 3229 | 0.5202 | 243 | 83.219 |
| 21 | AS | 3133 | 0.5047 | 218 | 74.658 |
| 22 | CHILDREN | 3031 | 0.4883 | 203 | 69.521 |
| 23 | IT | 3010 | 0.4849 | 206 | 70.548 |
| 24 | BY | 2958 | 0.4765 | 238 | 81.507 |
| 25 | NOT | 2867 | 0.4619 | 216 | 73.973 |
| 26 | HEALTH | 2837 | 0.457 | 210 | 71.918 |
| 27 | AT | 2763 | 0.4451 | 230 | 78.767 |
| 28 | FROM | 2625 | 0.4229 | 231 | 79.11 |
| 29 | VACCINES | 2436 | 0.3924 | 173 | 59.247 |
| 30 | UK | 2365 | 0.381 | 223 | 76.37 |
| 31 | DISEASE | 2342 | 0.3773 | 174 | 59.589 |
| 32 | WILL | 2323 | 0.3742 | 168 | 57.534 |
| 33 | HAS | 2246 | 0.3618 | 209 | 71.575 |
| 34 | CAN | 2179 | 0.351 | 200 | 68.493 |
| 35 | WAS | 2097 | 0.3378 | 171 | 58.562 |
| 36 | YOU | 2088 | 0.3364 | 163 | 55.822 |
| 37 | BEEN | 1904 | 0.3067 | 201 | 68.836 |
| 38 | MEASLES | 1890 | 0.3045 | 147 | 50.342 |
| 39 | AN | 1839 | 0.2963 | 211 | 72.26 |
| 40 | IF | 1826 | 0.2942 | 186 | 63.699 |
| 41 | THERE | 1807 | 0.2911 | 197 | 67.466 |
| 42 | SHOULD | 1770 | 0.2851 | 167 | 57.192 |
| 43 | YOUR | 1730 | 0.2787 | 144 | 49.315 |
| 44 | VACCINATION | 1699 | 0.2737 | 169 | 57.877 |
| 45 | WHO | 1650 | 0.2658 | 181 | 61.986 |
| 46 | INFORMATION | 1612 | 0.2597 | 201 | 68.836 |
| 47 | THEY | 1610 | 0.2594 | 178 | 60.959 |
| 48 | THEIR | 1553 | 0.2502 | 191 | 65.411 |
| 49 | AUTISM | 1545 | 0.2489 | 97 | 33.219 |
| 50 | MAY | 1544 | 0.2487 | 174 | 59.589 |
| 51 | ALL | 1490 | 0.24 | 198 | 67.808 |

| | | | | | |
|---|---|---|---|---|---|
| 52 | CHILD | 1425 | 0.2296 | 129 | 44.178 |
| 53 | HIB | 1420 | 0.2288 | 75 | 25.685 |
| 54 | PROGRAMME | 1397 | 0.2251 | 140 | 47.945 |
| 55 | ANY | 1387 | 0.2234 | 179 | 61.301 |
| 56 | NO | 1384 | 0.223 | 203 | 69.521 |
| 57 | GIVEN | 1340 | 0.2159 | 153 | 52.397 |
| 58 | WHICH | 1330 | 0.2143 | 168 | 57.534 |
| 59 | AGAINST | 1293 | 0.2083 | 179 | 61.301 |
| 60 | YEARS | 1269 | 0.2044 | 154 | 52.74 |
| 61 | WWW | 1262 | 0.2033 | 179 | 61.301 |
| 62 | WERE | 1247 | 0.2009 | 142 | 48.63 |
| 63 | RUBELLA | 1243 | 0.2002 | 108 | 36.986 |
| 64 | OTHER | 1237 | 0.1993 | 184 | 63.014 |
| 65 | RISK | 1183 | 0.1906 | 155 | 53.082 |
| 66 | ABOUT | 1149 | 0.1851 | 182 | 62.329 |
| 67 | MORE | 1138 | 0.1833 | 184 | 63.014 |
| 68 | THESE | 1124 | 0.1811 | 174 | 59.589 |
| 69 | HAD | 1116 | 0.1798 | 139 | 47.603 |
| 70 | NHS | 1113 | 0.1793 | 165 | 56.507 |
| 71 | ONE | 1075 | 0.1732 | 155 | 53.082 |
| 72 | AGE | 1040 | 0.1675 | 142 | 48.63 |
| 73 | AFTER | 1033 | 0.1664 | 146 | 50 |
| 74 | ALSO | 1018 | 0.164 | 175 | 59.932 |
| 75 | YEAR | 1013 | 0.1632 | 143 | 48.973 |
| 76 | BUT | 1010 | 0.1627 | 164 | 56.164 |
| 77 | MUMPS | 1003 | 0.1616 | 108 | 36.986 |
| 78 | MONTHS | 995 | 0.1603 | 135 | 46.233 |
| 79 | UP | 978 | 0.1576 | 174 | 59.589 |
| 80 | BETWEEN | 962 | 0.155 | 169 | 57.877 |
| 81 | OVER | 948 | 0.1527 | 168 | 57.534 |
| 82 | NEW | 946 | 0.1524 | 162 | 55.479 |
| 83 | HPV | 936 | 0.1508 | 48 | 16.438 |
| 84 | DOSE | 930 | 0.1498 | 109 | 37.329 |
| 85 | CASES | 917 | 0.1477 | 134 | 45.89 |
| 86 | PEOPLE | 912 | 0.1469 | 124 | 42.466 |
| 87 | INFECTION | 906 | 0.146 | 124 | 42.466 |
| 88 | S | 900 | 0.145 | 143 | 48.973 |
| 89 | FIGURE | 898 | 0.1447 | 22 | 7.5342 |
| 90 | EVIDENCE | 893 | 0.1439 | 125 | 42.808 |
| 91 | WE | 885 | 0.1426 | 141 | 48.288 |
| 92 | PROTECTION | 871 | 0.1403 | 128 | 43.836 |
| 93 | WHEN | 861 | 0.1387 | 163 | 55.822 |
| 94 | PNEUMOCOCCAL | 856 | 0.1379 | 47 | 16.096 |
| 95 | I | 848 | 0.1366 | 101 | 34.589 |
| 96 | TWO | 821 | 0.1323 | 153 | 52.397 |
| 97 | FLU | 804 | 0.1295 | 69 | 23.63 |
| 98 | MENINGITIS | 801 | 0.129 | 96 | 32.877 |
| 99 | TIME | 782 | 0.126 | 148 | 50.685 |
| 100 | MOST | 779 | 0.1255 | 145 | 49.658 |
| 101 | DATA | 770 | 0.124 | 100 | 34.247 |
| 102 | USED | 756 | 0.1218 | 147 | 50.342 |
| 103 | CAUSE | 754 | 0.1215 | 131 | 44.863 |
| 104 | HOW | 750 | 0.1208 | 122 | 41.781 |
| 105 | VIRUS | 745 | 0.12 | 104 | 35.616 |
| 106 | STUDY | 744 | 0.1199 | 98 | 33.562 |

| | | | | | |
|---|---|---|---|---|---|
| 107 | THAN | 738 | 0.1189 | 139 | 47.603 |
| 108 | SCHOOL | 723 | 0.1165 | 102 | 34.932 |
| 109 | AVAILABLE | 721 | 0.1161 | 153 | 52.397 |
| 110 | DEPARTMENT | 721 | 0.1161 | 154 | 52.74 |
| 111 | DH | 721 | 0.1161 | 98 | 33.562 |
| 112 | MEDICAL | 721 | 0.1161 | 125 | 42.808 |
| 113 | UNDER | 720 | 0.116 | 120 | 41.096 |
| 114 | RESEARCH | 714 | 0.115 | 102 | 34.932 |
| 115 | THOSE | 714 | 0.115 | 143 | 48.973 |
| 116 | WHAT | 710 | 0.1144 | 115 | 39.384 |
| 117 | BEING | 701 | 0.1129 | 161 | 55.137 |
| 118 | BEFORE | 698 | 0.1124 | 133 | 45.548 |
| 119 | SUCH | 697 | 0.1123 | 147 | 50.342 |
| 120 | COMMITTEE | 696 | 0.1121 | 70 | 23.973 |
| 121 | NUMBER | 695 | 0.112 | 145 | 49.658 |
| 122 | SOME | 694 | 0.1118 | 151 | 51.712 |
| 123 | DISEASES | 691 | 0.1113 | 121 | 41.438 |
| 124 | DO | 687 | 0.1107 | 137 | 46.918 |
| 125 | INFLUENZA | 666 | 0.1073 | 50 | 17.123 |
| 126 | TB | 665 | 0.1071 | 20 | 6.8493 |
| 127 | FIRST | 661 | 0.1065 | 137 | 46.918 |
| 128 | WOULD | 648 | 0.1044 | 125 | 42.808 |
| 129 | VERY | 632 | 0.1018 | 122 | 41.781 |
| 130 | PARENTS | 627 | 0.101 | 122 | 41.781 |
| 131 | PCT | 616 | 0.0992 | 38 | 13.014 |
| 132 | THREE | 616 | 0.0992 | 121 | 41.438 |
| 133 | IPV | 611 | 0.0984 | 52 | 17.808 |
| 134 | ONLY | 611 | 0.0984 | 144 | 49.315 |
| 135 | POLIO | 608 | 0.0979 | 48 | 16.438 |
| 136 | NEED | 607 | 0.0978 | 127 | 43.493 |
| 137 | AL | 580 | 0.0934 | 48 | 16.438 |
| 138 | ET | 578 | 0.0931 | 46 | 15.753 |
| 139 | USE | 576 | 0.0928 | 136 | 46.575 |
| 140 | SERIOUS | 574 | 0.0925 | 105 | 35.959 |
| 141 | SAFETY | 561 | 0.0904 | 102 | 34.932 |
| 142 | SO | 561 | 0.0904 | 138 | 47.26 |
| 143 | FOLLOWING | 558 | 0.0899 | 121 | 41.438 |
| 144 | GET | 547 | 0.0881 | 128 | 43.836 |
| 145 | PRACTICE | 542 | 0.0873 | 102 | 34.932 |
| 146 | FURTHER | 539 | 0.0868 | 123 | 42.123 |
| 147 | PUBLISHED | 536 | 0.0863 | 128 | 43.836 |
| 148 | SEE | 534 | 0.086 | 133 | 45.548 |
| 149 | CHILDHOOD | 533 | 0.0859 | 109 | 37.329 |
| 150 | C | 531 | 0.0855 | 93 | 31.849 |
| 151 | DOES | 522 | 0.0841 | 118 | 40.411 |
| 152 | WHERE | 517 | 0.0833 | 124 | 42.466 |
| 153 | CARE | 510 | 0.0822 | 101 | 34.589 |
| 154 | GOV | 507 | 0.0817 | 89 | 30.479 |
| 155 | LINK | 504 | 0.0812 | 98 | 33.562 |
| 156 | ENGLAND | 503 | 0.081 | 92 | 31.507 |
| 157 | GROUP | 494 | 0.0796 | 104 | 35.616 |
| 158 | SINGLE | 494 | 0.0796 | 103 | 35.274 |
| 159 | PRIMARY | 489 | 0.0788 | 86 | 29.452 |
| 160 | AGED | 484 | 0.078 | 88 | 30.137 |
| 161 | IMMUNISED | 483 | 0.0778 | 97 | 33.219 |

| | | | | | |
|---|---|---|---|---|---|
| 162 | PLEASE | 483 | 0.0778 | 102 | 34.932 |
| 163 | UPTAKE | 483 | 0.0778 | 70 | 23.973 |
| 164 | B | 481 | 0.0775 | 78 | 26.712 |
| 165 | OUT | 480 | 0.0773 | 127 | 43.493 |
| 166 | INTO | 479 | 0.0772 | 131 | 44.863 |
| 167 | PROTECT | 476 | 0.0767 | 120 | 41.096 |
| 168 | INFECTIONS | 473 | 0.0762 | 75 | 25.685 |
| 169 | BECAUSE | 469 | 0.0756 | 116 | 39.726 |
| 170 | NONE | 468 | 0.0754 | 34 | 11.644 |
| 171 | ADVICE | 466 | 0.0751 | 99 | 33.904 |
| 172 | IMMUNISATIONS | 465 | 0.0749 | 79 | 27.055 |
| 173 | SYSTEM | 458 | 0.0738 | 100 | 34.247 |
| 174 | WEBSITE | 458 | 0.0738 | 117 | 40.068 |
| 175 | WORK | 458 | 0.0738 | 114 | 39.041 |
| 176 | CATCH | 457 | 0.0736 | 97 | 33.219 |
| 177 | NOW | 456 | 0.0735 | 128 | 43.836 |
| 178 | PCTS | 455 | 0.0733 | 47 | 16.096 |
| 179 | TETANUS | 455 | 0.0733 | 50 | 17.123 |
| 180 | CAMPAIGN | 454 | 0.0731 | 74 | 25.342 |
| 181 | CAUSED | 451 | 0.0727 | 93 | 31.849 |
| 182 | DOSES | 448 | 0.0722 | 96 | 32.877 |
| 183 | HIGH | 446 | 0.0718 | 104 | 35.616 |
| 184 | IMMUNE | 445 | 0.0717 | 79 | 27.055 |
| 185 | BOOSTER | 444 | 0.0715 | 57 | 19.521 |
| 186 | PROFESSIONALS | 442 | 0.0712 | 74 | 25.342 |
| 187 | FEVER | 439 | 0.0707 | 71 | 24.315 |
| 188 | EFFECTS | 438 | 0.0706 | 90 | 30.822 |
| 189 | SUPPLY | 437 | 0.0704 | 56 | 19.178 |
| 190 | SYMPTOMS | 435 | 0.0701 | 79 | 27.055 |
| 191 | HTTP | 431 | 0.0694 | 99 | 33.904 |
| 192 | YOUNG | 431 | 0.0694 | 100 | 34.247 |
| 193 | RECOMMENDED | 430 | 0.0693 | 96 | 32.877 |
| 194 | DIPHTHERIA | 429 | 0.0691 | 50 | 17.123 |
| 195 | MY | 426 | 0.0686 | 48 | 16.438 |
| 196 | PERTUSSIS | 424 | 0.0683 | 49 | 16.781 |
| 197 | OFFERED | 422 | 0.068 | 90 | 30.822 |
| 198 | PUBLIC | 422 | 0.068 | 97 | 33.219 |
| 199 | ROUTINE | 422 | 0.068 | 90 | 30.822 |
| 200 | GROUPS | 419 | 0.0675 | 69 | 23.63 |

**Appendix 2** Keywords in *NHSvax* corpus occurring in at least 10% of corpus files, with *BNC World* as reference corpus and log-likelihood as measure of significance

| Rank | Keyword | Freq. | LL | Rank | Keyword | Freq. | LL |
|---|---|---|---|---|---|---|---|
| 1 | VACCINE | 7188 | 70934.16 | 51 | INFECTIONS | 473 | 3281.65 |
| 2 | MMR | 3653 | 37508.30 | 52 | THIOMERSAL | 322 | 3265.59 |
| 3 | IMMUNISATION | 3425 | 34225.91 | 53 | JCVI | 309 | 3175.11 |
| 4 | VACCINES | 2431 | 23712.53 | 54 | DOSES | 447 | 3086.40 |
| 5 | # | 27361 | 22606.45 | 55 | CERVICAL | 377 | 2939.64 |
| 6 | MEASLES | 1784 | 16864.86 | 56 | ORG | 298 | 2929.95 |
| 7 | VACCINATION | 1686 | 15895.24 | 57 | VACCINATED | 304 | 2763.95 |
| 8 | AUTISM | 1521 | 15336.28 | 58 | IMMUNE | 437 | 2759.87 |
| 9 | HIB | 1368 | 13893.88 | 59 | FEVER | 439 | 2712.66 |
| 10 | WWW | 1262 | 12939.87 | 60 | INFECTIOUS | 368 | 2638.81 |
| 11 | DISEASE | 2324 | 12557.37 | 61 | WAKEFIELD | 348 | 2615.11 |
| 12 | RUBELLA | 1154 | 11463.29 | 62 | CHILDHOOD | 532 | 2586.30 |
| 13 | HEALTH | 2827 | 11182.71 | 63 | GP | 393 | 2452.15 |
| 14 | UK | 2347 | 10142.14 | 64 | AGE | 1024 | 2446.87 |
| 15 | CHILDREN | 3073 | 9732.91 | 65 | GIVEN | 1340 | 2352.83 |
| 16 | MUMPS | 923 | 9254.38 | 66 | MEDICAL | 715 | 2313.22 |
| 17 | HPV | 920 | 8819.88 | 67 | CASES | 917 | 2282.54 |
| 18 | PNEUMOCOCCAL | 856 | 8645.16 | 68 | PRE-SCHOOL | 294 | 2262.64 |
| 19 | NHS | 1111 | 7027.29 | 69 | LANCET | 274 | 2242.03 |
| 20 | MENINGITIS | 784 | 6967.57 | 70 | AUTISTIC | 251 | 2240.21 |
| 21 | DH | 721 | 6444.52 | 71 | MEDICINES | 336 | 2227.96 |
| 22 | FLU | 788 | 6217.82 | 72 | INJECTION | 372 | 2225.77 |
| 23 | INFLUENZA | 656 | 5981.73 | 73 | ET | 578 | 2217.38 |
| 24 | PCT | 612 | 5966.65 | 74 | VIRUSES | 314 | 2160.45 |
| 25 | DOSE | 886 | 5850.95 | 75 | AL | 578 | 2153.97 |
| 26 | IPV | 553 | 5348.42 | 76 | EMAIL | 232 | 2145.90 |
| 27 | POLIO | 593 | 5347.92 | 77 | PROFESSIONALS | 442 | 2139.14 |
| 28 | INFECTION | 905 | 5268.16 | 78 | MONTHS | 995 | 2113.67 |
| 29 | GOV | 507 | 5103.44 | 79 | PDF | 220 | 2053.07 |
| 30 | S | 1640 | 5016.52 | 80 | CAUSE | 753 | 2047.45 |
| 31 | WEBSITE | 457 | 4695.99 | 81 | EVIDENCE | 885 | 1917.64 |
| 32 | VIRUS | 728 | 4689.40 | 82 | GPS | 302 | 1917.54 |
| 33 | PCTS | 455 | 4675.44 | 83 | WHOOPING | 226 | 1876.97 |
| 34 | IMMUNISATIONS | 465 | 4673.70 | 84 | SYMPTOMS | 435 | 1864.76 |
| 35 | IMMUNISED | 474 | 4550.71 | 85 | LINK | 504 | 1825.44 |
| 36 | CHILD | 1478 | 4525.23 | 86 | HAEMOPHILUS | 186 | 1820.80 |
| 37 | HTTP | 431 | 4428.80 | 87 | AGED | 483 | 1811.71 |
| 38 | PROGRAMME | 1398 | 4413.15 | 88 | HPA | 181 | 1806.81 |
| 39 | DISEASES | 688 | 4138.75 | 89 | ROUTINE | 419 | 1768.80 |
| 40 | PERTUSSIS | 409 | 4029.22 | 90 | GSI | 184 | 1750.92 |
| 41 | TETANUS | 450 | 3973.43 | 91 | COUGH | 289 | 1731.65 |
| 42 | BOOSTER | 444 | 3946.81 | 92 | INFLUENZAE | 177 | 1718.76 |
| 43 | DTAP | 382 | 3925.27 | 93 | PROTECT | 476 | 1707.31 |
| 44 | DIPHTHERIA | 423 | 3857.44 | 94 | RECOMMENDED | 429 | 1693.27 |
| 45 | UPTAKE | 483 | 3816.78 | 95 | BOWEL | 315 | 1678.82 |
| 46 | MENC | 371 | 3812.23 | 96 | DATA | 764 | 1669.04 |
| 47 | RISK | 1055 | 3690.88 | 97 | HEALTHCARE | 226 | 1662.68 |
| 48 | INFORMATION | 1608 | 3493.32 | 98 | CONJUGATE | 188 | 1660.85 |
| 49 | BCG | 371 | 3374.59 | 99 | BABIES | 368 | 1644.47 |
| 50 | PROTECTION | 867 | 3331.21 | 100 | SAFETY | 558 | 1640.54 |

| 101 | AGAINST | 1293 | 1609.99 | 151 | DOCTOR | 395 | 862.58 |
|---|---|---|---|---|---|---|---|
| 102 | REACTIONS | 346 | 1609.07 | 152 | NURSE | 257 | 851.70 |
| 103 | DEPARTMENT | 725 | 1586.28 | 153 | GIRLS | 386 | 846.76 |
| 104 | UPDATE | 300 | 1559.36 | 154 | SCHEDULE | 234 | 838.56 |
| 105 | IMMUNITY | 254 | 1555.68 | 155 | SWELLING | 146 | 835.86 |
| 106 | VACCINATIONS | 167 | 1494.23 | 156 | YOUR | 1730 | 817.88 |
| 107 | ADVERSE | 284 | 1491.24 | 157 | SURVEILLANCE | 162 | 814.54 |
| 108 | COMMITTEE | 710 | 1452.10 | 158 | CONTACT | 405 | 812.42 |
| 109 | SHOULD | 1770 | 1403.02 | 159 | SYNDROME | 179 | 782.96 |
| 110 | BACTERIA | 275 | 1395.43 | 160 | COMPLICATIONS | 161 | 781.88 |
| 111 | CMO | 138 | 1374.67 | 161 | CAUSES | 278 | 781.40 |
| 112 | STUDY | 747 | 1370.90 | 162 | INFLAMMATION | 138 | 779.36 |
| 113 | SERIOUS | 573 | 1362.55 | 163 | INFECTED | 176 | 779.27 |
| 114 | OR | 4128 | 1359.60 | 164 | PROTECTED | 225 | 776.12 |
| 115 | INACTIVATED | 153 | 1348.85 | 165 | LICENSED | 172 | 758.21 |
| 116 | FACTSHEET | 156 | 1340.85 | 166 | DEATHS | 216 | 753.41 |
| 117 | PUBLISHED | 535 | 1325.02 | 167 | INFANTS | 156 | 744.63 |
| 118 | AGENCY | 417 | 1314.38 | 168 | ENCEPHALITIS | 85 | 744.42 |
| 119 | CANCER | 374 | 1301.14 | 169 | BABY | 340 | 743.43 |
| 120 | JAB | 164 | 1297.98 | 170 | YEARS | 1266 | 743.33 |
| 121 | PARENTS | 626 | 1288.78 | 171 | CONGENITAL | 117 | 742.67 |
| 122 | ANAPHYLACTIC | 129 | 1276.47 | 172 | IMMUNISE | 81 | 722.65 |
| 123 | INCIDENCE | 275 | 1246.44 | 173 | ARE | 4220 | 719.45 |
| 124 | PRIMARY | 489 | 1244.28 | 174 | SINGLE | 490 | 714.96 |
| 125 | PROTECTS | 190 | 1235.16 | 175 | ILLNESS | 227 | 696.18 |
| 126 | NONE | 468 | 1187.75 | 176 | OUTBREAKS | 108 | 693.50 |
| 127 | DISORDERS | 215 | 1154.90 | 177 | PROFESSOR | 264 | 690.32 |
| 128 | RASH | 181 | 1118.90 | 178 | SEASONAL | 151 | 689.08 |
| 129 | JOURNAL | 278 | 1113.58 | 179 | MILD | 171 | 658.27 |
| 130 | CAMPAIGN | 453 | 1092.10 | 180 | GREENBOOK | 64 | 657.60 |
| 131 | CAUSED | 450 | 1078.78 | 181 | INTRODUCTION | 291 | 657.32 |
| 132 | ADVICE | 466 | 1070.13 | 182 | PASTEUR | 80 | 651.27 |
| 133 | AVAILABLE | 721 | 1040.40 | 183 | GUIDANCE | 216 | 646.01 |
| 134 | DOESN | 106 | 1036.74 | 184 | PUBLICATIONS | 182 | 641.42 |
| 135 | RESEARCH | 714 | 1032.69 | 185 | ENGLAND | 506 | 635.90 |
| 136 | CLINICAL | 284 | 1021.48 | 186 | SEVERE | 244 | 630.30 |
| 137 | SUPPLY | 437 | 1012.44 | 187 | ROUTINELY | 112 | 620.02 |
| 138 | REACTION | 347 | 1004.72 | 188 | COVERAGE | 184 | 618.83 |
| 139 | CSM | 121 | 1001.25 | 189 | MMRTHEFACTS | 60 | 616.50 |
| 140 | HTM | 95 | 976.13 | 190 | MAY | 1530 | 608.38 |
| 141 | INTRODUCED | 407 | 972.28 | 191 | FOLLOWING | 558 | 604.33 |
| 142 | NURSES | 258 | 969.44 | 192 | CAN | 2213 | 604.01 |
| 143 | ANTIBODIES | 184 | 961.20 | 193 | MEDICINE | 194 | 598.38 |
| 144 | PRACTICE | 539 | 923.70 | 194 | SEPTEMBER | 337 | 592.35 |
| 145 | PLEASE | 483 | 903.97 | 195 | ADULTS | 206 | 588.71 |
| 146 | PREGNANT | 232 | 892.78 | 196 | EFFECTIVE | 327 | 581.02 |
| 147 | OFFERED | 422 | 890.09 | 197 | EPIDEMIOLOGY | 88 | 578.39 |
| 148 | EFFECTS | 427 | 887.68 | 198 | PREGNANCY | 155 | 569.96 |
| 149 | PNEUMONIA | 149 | 871.65 | 199 | COMMUNICABLE | 69 | 566.15 |
| 150 | LEAFLET | 190 | 865.16 | 200 | SAFEST | 96 | 559.06 |

| 201 | BRAIN | 225 | 548.99 | 251 | ILLNESSES | 83 | 378.33 |
|---|---|---|---|---|---|---|---|
| 202 | TRUSTS | 146 | 540.85 | 252 | BLOOD | 261 | 377.15 |
| 203 | CARE | 505 | 539.43 | 253 | UNWELL | 66 | 369.71 |
| 204 | RARE | 220 | 529.07 | 254 | LETTER | 304 | 369.05 |
| 205 | REDNESS | 74 | 527.54 | 255 | DELIVERY | 160 | 365.44 |
| 206 | ORDERLINE | 52 | 524.39 | 256 | REPORTS | 273 | 361.86 |
| 207 | INFECT | 75 | 524.19 | 257 | RECEIVED | 298 | 359.20 |
| 208 | SURGERIES | 77 | 512.30 | 258 | CLINIC | 110 | 355.52 |
| 209 | CHRONIC | 148 | 511.35 | 259 | ONLINE | 83 | 355.29 |
| 210 | INCREASE | 401 | 510.91 | 260 | INDIVIDUALS | 228 | 353.45 |
| 211 | ENSURE | 308 | 502.51 | 261 | FURTHER | 539 | 350.48 |
| 212 | RECEIVE | 263 | 497.39 | 262 | PRACTICES | 173 | 349.44 |
| 213 | CONCERNS | 194 | 486.59 | 263 | LEVELS | 281 | 348.20 |
| 214 | DOWNLOAD | 59 | 486.52 | 264 | TYPE | 342 | 344.35 |
| 215 | THIS | 3905 | 485.29 | 265 | HEADACHE | 87 | 342.77 |
| 216 | BACTERIAL | 101 | 479.82 | 266 | DUE | 319 | 341.33 |
| 217 | ANY | 1387 | 478.68 | 267 | BORN | 227 | 338.65 |
| 218 | INJECTIONS | 90 | 478.20 | 268 | SWOLLEN | 83 | 338.60 |
| 219 | REVIEW | 290 | 475.35 | 269 | FUNDING | 166 | 337.41 |
| 220 | EFFICACY | 98 | 474.93 | 270 | HAS | 2246 | 327.31 |
| 221 | CAUSAL | 123 | 467.64 | 271 | LEAFLETS | 86 | 326.06 |
| 222 | ORDERING | 119 | 463.65 | 272 | MATERIALS | 201 | 324.12 |
| 223 | COMBINED | 200 | 461.34 | 273 | SORE | 87 | 323.81 |
| 224 | GROUPS | 411 | 457.77 | 274 | PROTECTING | 97 | 320.91 |
| 225 | SCIENTIFIC | 224 | 455.07 | 275 | AWARENESS | 148 | 320.82 |
| 226 | STRAINS | 105 | 449.78 | 276 | PAPER | 323 | 313.83 |
| 227 | YEAR | 904 | 448.31 | 277 | ASSOCIATED | 231 | 312.23 |
| 228 | EN | 128 | 445.98 | 278 | SALISBURY | 81 | 306.11 |
| 229 | DISORDER | 133 | 444.52 | 279 | SAFE | 193 | 302.26 |
| 230 | REPORTED | 312 | 440.70 | 280 | COMMON | 352 | 301.22 |
| 231 | DON | 137 | 440.37 | 281 | DR | 265 | 300.80 |
| 232 | B | 477 | 439.75 | 282 | INCLUDING | 395 | 300.34 |
| 233 | WALES | 272 | 439.17 | 283 | RISKS | 124 | 299.62 |
| 234 | DOH | 73 | 438.48 | 284 | MANUFACTURED | 88 | 298.40 |
| 235 | LABORATORY | 157 | 433.28 | 285 | CHIEF | 251 | 298.37 |
| 236 | BE | 5209 | 432.69 | 286 | CAUSING | 136 | 297.93 |
| 237 | SOURCE | 268 | 430.59 | 287 | CONDITION | 213 | 294.17 |
| 238 | GLANDS | 81 | 417.06 | 288 | PROVIDES | 213 | 292.55 |
| 239 | REVIEWED | 123 | 415.45 | 289 | DOCUMENT | 167 | 287.21 |
| 240 | IS | 7355 | 410.06 | 290 | CONFIRMED | 161 | 286.50 |
| 241 | EPIDEMIC | 89 | 409.17 | 291 | EXPOSURE | 115 | 284.74 |
| 242 | DIAGNOSED | 97 | 407.37 | 292 | GATEWAY | 71 | 283.65 |
| 243 | OCCUR | 206 | 405.94 | 293 | FINDINGS | 133 | 281.51 |
| 244 | NUMBER | 695 | 405.26 | 294 | TYPES | 215 | 279.94 |
| 245 | ASSOCIATION | 293 | 403.20 | 295 | CATCH | 159 | 277.34 |
| 246 | WILL | 2323 | 402.61 | 296 | UPDATES | 51 | 276.58 |
| 247 | VIRAL | 78 | 386.81 | 297 | BETWEEN | 962 | 272.71 |
| 248 | STUDIES | 314 | 385.80 | 298 | J | 232 | 272.37 |
| 249 | OLDER | 248 | 381.58 | 299 | RECOMMEND | 101 | 271.04 |
| 250 | C | 510 | 379.81 | 300 | CURRENTLY | 187 | 268.85 |

| 301 | CONCLUDED | 126 | 267.75 | 351 | ADDITIONAL | 166 | 198.45 |
|---|---|---|---|---|---|---|---|
| 302 | RATES | 244 | 267.54 | 352 | NEED | 607 | 194.56 |
| 303 | USED | 756 | 267.44 | 353 | NURSING | 103 | 193.92 |
| 304 | CATCHING | 89 | 266.14 | 354 | ORAL | 93 | 193.72 |
| 305 | SIGNS | 153 | 262.28 | 355 | SURGERY | 98 | 193.16 |
| 306 | CONTAINS | 149 | 261.87 | 356 | PREVIOUS | 218 | 192.07 |
| 307 | UNDER | 709 | 259.53 | 357 | AUTHORISED | 66 | 190.67 |
| 308 | LIKELY | 364 | 259.01 | 358 | PREVENTION | 77 | 190.10 |
| 309 | ORDERS | 159 | 258.93 | 359 | CONDUCTED | 102 | 187.65 |
| 310 | EFFECTIVENESS | 104 | 257.74 | 360 | OFFICER | 177 | 186.42 |
| 311 | DEVELOP | 202 | 254.61 | 361 | DAMAGE | 173 | 186.13 |
| 312 | FEBRUARY | 197 | 248.36 | 362 | ADVISE | 85 | 184.91 |
| 313 | RESOURCES | 223 | 248.03 | 363 | CONTAINING | 118 | 184.86 |
| 314 | SUMMARY | 116 | 246.89 | 364 | TESTED | 97 | 184.80 |
| 315 | AFFECTS | 87 | 246.07 | 365 | NERVOUS | 100 | 184.46 |
| 316 | PREVENT | 174 | 244.66 | 366 | RECEIVING | 105 | 183.56 |
| 317 | BREATHING | 105 | 243.90 | 367 | EXPERTS | 104 | 182.55 |
| 318 | MONITORING | 113 | 242.54 | 368 | LOCALLY | 80 | 182.07 |
| 319 | INCREASED | 250 | 242.49 | 369 | THESE | 1124 | 181.02 |
| 320 | COPYRIGHT | 75 | 238.51 | 370 | COUNTRIES | 259 | 180.20 |
| 321 | DIRECT | 234 | 237.72 | 371 | STRAIN | 91 | 177.04 |
| 322 | DETAILS | 232 | 235.10 | 372 | LINKED | 119 | 176.57 |
| 323 | WEEKS | 270 | 234.05 | 373 | COMPONENT | 92 | 172.44 |
| 324 | E-MAIL | 45 | 233.85 | 374 | RECOMMENDS | 47 | 168.61 |
| 325 | SPREAD | 161 | 228.01 | 375 | PRODUCED | 213 | 165.33 |
| 326 | POPULATION | 243 | 225.60 | 376 | PATIENTS | 258 | 164.66 |
| 327 | CONTAIN | 136 | 223.53 | 377 | HAVE | 3337 | 163.94 |
| 328 | USUALLY | 304 | 223.13 | 378 | FAX | 65 | 163.42 |
| 329 | YOUNG | 431 | 221.82 | 379 | EXPERT | 111 | 162.31 |
| 330 | ADVISED | 102 | 221.50 | 380 | RARELY | 111 | 161.94 |
| 331 | RECOMMENDATIONS | 104 | 220.16 | 381 | AFTER | 1029 | 160.10 |
| 332 | AUGUST | 181 | 216.32 | 382 | MEDIA | 157 | 159.83 |
| 333 | ILL | 138 | 215.74 | 383 | THROAT | 93 | 157.43 |
| 334 | TREATMENT | 229 | 215.62 | 384 | EXPOSED | 88 | 157.31 |
| 335 | IMPACT | 173 | 215.50 | 385 | PREVENTING | 62 | 154.78 |
| 336 | TESTING | 116 | 211.59 | 386 | MONTHLY | 76 | 151.67 |
| 337 | SUPPLIED | 115 | 211.07 | 387 | RESPONSE | 196 | 150.24 |
| 338 | VISITOR | 91 | 210.20 | 388 | MONITORED | 53 | 146.63 |
| 339 | SUPPLIES | 111 | 209.04 | 389 | TEST | 212 | 146.20 |
| 340 | GROUP | 490 | 207.84 | 390 | SURVEY | 153 | 145.21 |
| 341 | UPDATED | 67 | 204.55 | 391 | PRACTITIONERS | 74 | 145.03 |
| 342 | SKIN | 160 | 204.04 | 392 | DEVELOPING | 131 | 144.63 |
| 343 | HOSPITAL | 251 | 202.36 | 393 | PROVIDED | 241 | 141.65 |
| 344 | WHO | 1654 | 201.73 | 394 | WORLDWIDE | 76 | 140.72 |
| 345 | REPORTING | 100 | 201.51 | 395 | SCHOOL | 409 | 140.00 |
| 346 | MARCH | 263 | 201.23 | 396 | DOES | 522 | 138.55 |
| 347 | PAINFUL | 86 | 200.61 | 397 | DOCTORS | 107 | 137.90 |
| 348 | PREVIOUSLY | 160 | 200.50 | 398 | SITE | 166 | 134.49 |
| 349 | ISSUES | 223 | 200.40 | 399 | PUBLIC | 423 | 133.99 |
| 350 | REPORT | 372 | 199.40 | 400 | TEMPERATURE | 104 | 133.05 |

| 401 | VISIT | 192 | 132.78 | 451 | INCLUDE | 194 | 90.52 |
|---|---|---|---|---|---|---|---|
| 402 | APPROPRIATE | 183 | 132.63 | 452 | LEAD | 187 | 90.51 |
| 403 | ANNUAL | 148 | 131.89 | 453 | MULTIPLE | 64 | 89.73 |
| 404 | USING | 303 | 131.09 | 454 | FEEDBACK | 46 | 87.81 |
| 405 | TEL | 63 | 129.64 | 455 | PATIENT | 118 | 87.42 |
| 406 | MONTH | 205 | 129.28 | 456 | COMPARED | 132 | 86.16 |
| 407 | PERSONAL | 238 | 128.93 | 457 | TEN | 230 | 85.89 |
| 408 | SUPPORT | 353 | 126.37 | 458 | RESPONSIBLE | 137 | 85.28 |
| 409 | WOMEN | 371 | 123.95 | 459 | POSSIBLE | 343 | 85.05 |
| 410 | MANUFACTURERS | 79 | 122.76 | 460 | D | 189 | 84.35 |
| 411 | PARENT | 90 | 121.72 | 461 | AROUND | 416 | 82.95 |
| 412 | HUMAN | 251 | 121.60 | 462 | RESPOND | 74 | 82.21 |
| 413 | APPOINTMENT | 100 | 120.90 | 463 | HIGH | 376 | 82.12 |
| 414 | OCTOBER | 168 | 120.29 | 464 | SHOWN | 186 | 82.01 |
| 415 | SIDE | 360 | 119.39 | 465 | LOW | 202 | 81.79 |
| 416 | NATIONAL | 404 | 118.85 | 466 | ORDER | 344 | 81.33 |
| 417 | DIARRHOEA | 43 | 117.48 | 467 | BEEN | 1904 | 80.62 |
| 418 | SYSTEM | 454 | 116.80 | 468 | REDUCE | 112 | 79.61 |
| 419 | ACUTE | 70 | 116.45 | 469 | FULLY | 129 | 79.59 |
| 420 | TRUST | 157 | 114.84 | 470 | TARGET | 106 | 78.60 |
| 421 | FOR | 5992 | 114.25 | 471 | JANUARY | 140 | 77.35 |
| 422 | CURRENT | 200 | 114.21 | 472 | BIRTH | 90 | 75.79 |
| 423 | PAGE | 166 | 113.47 | 473 | IDENTIFIED | 101 | 75.07 |
| 424 | DATE | 227 | 113.40 | 474 | ISSUED | 95 | 73.85 |
| 425 | LIVE | 224 | 113.36 | 475 | JOINT | 107 | 72.06 |
| 426 | SPECIFIC | 170 | 111.28 | 476 | ARTICLE | 103 | 71.36 |
| 427 | STATISTICS | 81 | 110.47 | 477 | PEOPLE | 919 | 70.80 |
| 428 | OTHER | 1238 | 110.04 | 478 | DETAILED | 100 | 69.39 |
| 429 | E | 283 | 109.75 | 479 | QUESTIONS | 170 | 69.04 |
| 430 | CENTRE | 272 | 109.35 | 480 | NEWS | 171 | 68.92 |
| 431 | APRIL | 201 | 108.74 | 481 | RELEVANT | 114 | 68.26 |
| 432 | AFFECTED | 113 | 107.87 | 482 | THOSE | 714 | 68.11 |
| 433 | ADMINISTRATION | 118 | 107.05 | 483 | LONG-TERM | 75 | 67.61 |
| 434 | NOVEMBER | 148 | 105.91 | 484 | BEFORE | 698 | 67.16 |
| 435 | BODY | 283 | 105.54 | 485 | COLLEAGUE | 44 | 67.02 |
| 436 | AFFECTING | 58 | 104.26 | 486 | OF | 19105 | 66.39 |
| 437 | FACTS | 103 | 103.85 | 487 | DECEMBER | 126 | 66.30 |
| 438 | CO | 100 | 101.31 | 488 | IF | 1826 | 66.28 |
| 439 | CHECK | 123 | 101.00 | 489 | PERSON | 244 | 65.61 |
| 440 | ALSO | 1018 | 100.54 | 490 | PROVIDE | 233 | 63.89 |
| 441 | M | 225 | 100.49 | 491 | ORDERED | 83 | 63.50 |
| 442 | RESULTS | 202 | 100.35 | 492 | WORKERS | 170 | 62.59 |
| 443 | SHOWS | 167 | 100.25 | 493 | GENERAL | 354 | 62.59 |
| 444 | INDEPENDENT | 158 | 98.01 | 494 | PROGRAMMES | 96 | 61.53 |
| 445 | USE | 574 | 97.89 | 495 | KEY | 150 | 61.45 |
| 446 | GIVING | 173 | 96.77 | 496 | MOST | 779 | 61.35 |
| 447 | COPIES | 79 | 94.95 | 497 | CHEST | 66 | 60.30 |
| 448 | PRIOR | 94 | 94.89 | 498 | LONDON | 312 | 59.71 |
| 449 | RECENT | 201 | 93.83 | 499 | STOCK | 110 | 59.52 |
| 450 | PRODUCT | 157 | 91.57 | 500 | YELLOW | 74 | 58.27 |

| 501 | PRESS | 151 | 57.36 | 551 | NOT | 2867 | 38.69 |
|---|---|---|---|---|---|---|---|
| 502 | ORGANIZATION | 90 | 57.10 | 552 | IDENTIFY | 69 | 38.55 |
| 503 | CONDITIONS | 172 | 56.89 | 553 | BACKGROUND | 79 | 38.26 |
| 504 | RATE | 201 | 56.60 | 554 | STARTS | 60 | 38.04 |
| 505 | SEPARATELY | 43 | 56.51 | 555 | TREAT | 56 | 37.16 |
| 506 | REFERENCE | 110 | 56.43 | 556 | PROPERLY | 73 | 36.40 |
| 507 | DRUGS | 82 | 55.97 | 557 | BENEFITS | 92 | 36.17 |
| 508 | REQUIRED | 179 | 55.70 | 558 | ANSWERS | 54 | 34.84 |
| 509 | BEING | 696 | 54.53 | 559 | EVENTS | 112 | 34.09 |
| 510 | TELEPHONE | 105 | 54.18 | 560 | JULY | 125 | 34.07 |
| 511 | SERVICES | 245 | 54.08 | 561 | DIRECTOR | 125 | 34.01 |
| 512 | SECOND | 364 | 54.06 | 562 | CONTINUES | 57 | 33.93 |
| 513 | OCCURS | 60 | 53.85 | 563 | INCLUDES | 82 | 33.75 |
| 514 | REASONS | 130 | 53.07 | 564 | BELOW | 142 | 33.69 |
| 515 | SEPARATE | 115 | 52.38 | 565 | INCREASING | 93 | 33.64 |
| 516 | N | 121 | 52.33 | 566 | COLLEAGUES | 70 | 33.16 |
| 517 | VIA | 73 | 52.10 | 567 | OCCURRED | 69 | 32.55 |
| 518 | AMOUNTS | 56 | 51.88 | 568 | PRODUCTS | 113 | 32.54 |
| 519 | VULNERABLE | 49 | 51.28 | 569 | THREE | 592 | 32.28 |
| 520 | NEW | 936 | 50.34 | 570 | TOTAL | 166 | 32.22 |
| 521 | CHANGES | 196 | 50.19 | 571 | OVER | 937 | 31.94 |
| 522 | GUIDE | 84 | 49.62 | 572 | CONTENT | 74 | 31.48 |
| 523 | AFFECT | 75 | 49.41 | 573 | HIGHLY | 99 | 31.05 |
| 524 | AWARE | 125 | 49.14 | 574 | MEASURE | 77 | 30.76 |
| 525 | CROWN | 76 | 48.98 | 575 | PROPORTION | 78 | 30.60 |
| 526 | PROVIDING | 97 | 48.95 | 576 | COPY | 72 | 29.95 |
| 527 | LATEST | 89 | 48.63 | 577 | FREQUENTLY | 70 | 29.41 |
| 528 | CONTINUE | 134 | 47.95 | 578 | NEEDED | 151 | 28.63 |
| 529 | NUMBERS | 130 | 47.89 | 579 | FREE | 185 | 28.39 |
| 530 | ADDITION | 119 | 47.29 | 580 | PROBLEMS | 233 | 28.38 |
| 531 | RECOGNISED | 75 | 46.28 | 581 | JUNE | 140 | 28.38 |
| 532 | HOW | 750 | 45.49 | 582 | INDIVIDUAL | 167 | 27.89 |
| 533 | SINCE | 412 | 45.47 | 583 | SUFFICIENT | 70 | 27.26 |
| 534 | SECTION | 189 | 45.41 | 584 | SHORTLY | 51 | 27.17 |
| 535 | HEART | 145 | 45.34 | 585 | AVOID | 86 | 26.87 |
| 536 | LINKS | 66 | 44.11 | 586 | FRIDAY | 64 | 26.17 |
| 537 | LISTED | 55 | 43.70 | 587 | GREEN | 134 | 25.81 |
| 538 | RELATED | 105 | 43.65 | 588 | HEALTHY | 48 | 25.55 |
| 539 | IMPORTANT | 336 | 43.61 | 589 | RECORDS | 82 | 25.15 |
| 540 | CARD | 78 | 43.17 | 590 | EXTREMELY | 75 | 24.98 |
| 541 | WINTER | 90 | 42.82 | 591 | DISCUSS | 65 | 24.88 |
| 542 | MONDAY | 72 | 42.74 | | | | |
| 543 | STATEMENT | 114 | 42.22 | | | | |
| 544 | ACCESS | 123 | 42.12 | | | | |
| 545 | FOUND | 394 | 41.73 | | | | |
| 546 | ESTIMATED | 70 | 41.57 | | | | |
| 547 | RISE | 119 | 41.31 | | | | |
| 548 | CANNOT | 206 | 41.16 | | | | |
| 549 | MINOR | 70 | 41.15 | | | | |
| 550 | CONSIDERED | 141 | 39.77 | | | | |

**Appendix 3** Keywords in the *NHSvax* corpus with distribution across at least 10 % of corpus files, arranged according to semantic and functional categories and ordered according to LL.

| | | |
|---|---|---|
| **vaccination and pharmaceutical products** | vaccines | *vaccine*, *mmr*, *vaccines*, *ipv* (inactivated polio vaccine), *dtap* (diphtheria, tetanus and pertussis), *menc* (meningococcal conjugate C vaccine), *bcg* (anti-tuberculosis vaccine) |
| | vaccination | *immunisation*, *vaccination*, *immunisations*, *immunised*, *vaccinated*, *vaccinations* |
| | vaccination: pre-modifiers | *fully (immunised)* |
| | vaccination: post-modifiers | *component* |
| | vaccine types | *conjugate*, *inactivated (polio vaccine)*, *live*, *oral (polio vaccine)*, *single*, *combined*, *multiple*, *separate* |
| | mode of delivery | *separately* |
| | processes | *manufactured* |
| | other pharmaceutical products | *medicines*, *products*, *product*, *drugs* |
| | vaccine additives | *thiomersal* |
| **health, disease and medical conditions** | health and disease: general terms | *disease*, *health*, *infection*, *diseases*, *infections*, *illness*, *illnesses* |
| | health and disease: adjectives | *unwell*, *ill*, *healthy* |
| | diseases and infections | *measles*, *hib* (haemophilus influenzae type b), *rubella*, *mumps*, *hpv* (human papilloma virus), *pneumococcal* (*disease*/*infection*), *meningitis*, *flu*, *influenza*, *polio*, *virus*, *pertussis*, *tetanus*, *diphtheria*, *cervical (cancer)*, *viruses*, *whooping* (*cough*), *haemophilus (influenza(e) type b*), (*whooping*) *cough*, *influenzae*, *bacteria*, *cancer*, *pneumonia* |
| | diseases and infections: non-evaluative pre-modifiers | *infectious*, *congenital*, *communicable*, *chronic*, *bacterial*, *viral*, *seasonal*, *acute*, *long-term* |
| | diseases and infections: verbs | *infected*, *infect*, *catching* |
| | bacteria and virus types | *type*, *types*, *strains*, *b*, *c*, *strain* |
| | immune system and nervous system | *immune*, *immunity*, *antibodies*, *nervous (system)*, *(immune/nervous) system* |
| | immune system processes | *response*, *respond* |
| | medical conditions: general terms | *disorders*, *syndrome*, *condition*, *conditions*, *disorder* |
| | specific conditions | *autism*, *autistic(spectrum disorders)* |
| | non-pathological conditions | *pregnant*, *pregnancy* |
| | parts of the body | *bowel*, *brain*, *glands*, *blood*, *skin*, *throat*, *body*, *chest*, *heart* |
| | symptoms and signs | *fever*, *symptoms*, *rash*, *redness*, *headache*, *swelling*, *signs*, *temperature*, *diarrhoea* |
| | symptoms and signs: pre-modifiers | *sore*, *swollen* |

| | | |
|---|---|---|
| **people, places and institutions** | nhs users/vaccination target groups | *children, child, babies, parents, girls, infants, baby, adults, (age/risk/patient) groups, individuals, (general/target) population, young, group, patients, patient, women, people, person* |
| | pre-modifiers referring to groups of people | *older, young, human, target (groups/population), vulnerable (patients)* |
| | healthcare practitioners and medical experts | *gp, professionals, gps, nurses, nurse, doctor, (health) visitor, experts, expert, (general) practitioners, doctors, colleague, (healthcare) workers, colleagues* |
| | named individuals | *(dr andrew) wakefield,(dr david) salisbury* |
| | job titles and honorifics | *cmo (chief medical officer), professor, dr, chief (medical officer), director* |
| | public institutions | *pcts, primary (car/nhse trusts), (primary care/nhs) trusts, nhs, hpa, agency, department, committee, organization, joint (committee), csm (committee on the safety of medicines), (primary care/meningitis/nhs) trust, centre* |
| | specific institutions | *(royal) free (hospital)* |
| | industries | *manufacturers* |
| | private companies | *(aventis/sanofi) pasteur* |
| | places | *laboratory, clinic, surgeries, surgery, hospital, school* |
| | geographical references | *uk, countries, england, london, wales, national, locally, worldwide* |
| **science, medicine and healthcare** | general references to science, medicine and healthcare | *medical, healthcare, clinical, (general) practice, medicine, scientific, nursing, diagnosed, practices, treatment, treat* |
| | terms related to administration of vaccination programme | *programme, routine, update, ordering, campaign, supply, introduced, introduction (of), changes, schedule, delivery, funding, updates, orders, supplies, updated, supplied, public (health), order, programmes, resources, services, appointment, feedback, issued, routinely, catch(-up programme), ordered, responsible, stock, access, reference, (yellow) card, yellow (card scheme), measure* |
| | administration of vaccines | *dose, booster, doses, injection, jab, injections, (vaccination) site* |
| | references to age and time periods | *childhood, age, pre-school, months, aged, years, september, year, born, february, weeks, august, march, monthly, annual, month, october, date, april, november, january, birth, december, winter, monday, july, june, friday* |
| | references to epidemiological research and monitoring | *deaths, uptake, surveillance, coverage, outbreaks, epidemiology, epidemic, levels, monitoring, survey, spread, cases, incidence, rate, proportion* |
| | quantities and quantifiers | *number, numbers, amounts* |
| | reference to scientific research | *results, data, study, studies, findings, research, statistics* |

| | | |
|---|---|---|
| | research processes | *conducted* |
| | nhs information and communication | *information, advice, details, support, questions, answers, guidance* |
| | scientific and medical procedures and checks | *confirmed, testing, tested, test, check* |
| **sources** | sources of information: general | *media, news, press* |
| | references to publications: general | *journal, publications, paper, reports, summary, review, article* |
| | specific publications and websites | *lancet, greenbook, green (book), mmrthefacts* |
| | nhs documents | *copies, document, source, factsheet, leaflet, leaflets, letter, prolog, section, materials, page, guide, copy, content, records* |
| | written communication: processes | *published, reviewed, produced* |
| | miscellaneous | *et, al* |
| | means or channel of communication | *telephone, email, website, online, gateway, fax, tel, (dh publications) orderline* |
| | legal issues | *licensed, copyright, authorised, crown (copyright)* |
| **likelihood, causation, and temporal relations** | possibility | *risk, risks* |
| | cause and effect | *cause, link, caused , causes, associated, causing, due* (*to*), *impact, linked, association, associated , affects, affected, affecting, lead* (*to*), *affect, links, related* |
| | pre-modifiers | *causal* |
| | reason | *reasons* |
| | ergative verbs indicating change | *increase, develop, increased, developing, reduce, increasing, rise* |
| | temporal relations | *since, following, follow, followed, after, prior, before, previous, previously* |
| | danger | *complications, damage, catch, catching* |
| **expressions with modal meaning** | epistemic modality | *may, can, will, likely, possible, cannot* |
| | necessity/desirability | *recommended, should, recommend, advised, recommendations, need, advise, recommends, required, needed* |
| | directives | *contact, discuss* |
| **conditionality** | | *if* |
| **expressions of evaluation** | terms used to express positive evaluation of vaccines | *protection, protect, safety, protects, protected, effective, safest, efficacy, ensure, protecting, safe, prevent, prevention, preventing, benefits* |
| | potential ill-effects of vaccines | *adverse (events/effects/reactions), reactions, anaphylactic (shock/reaction), reaction, effects, side (effects), (adverse) events* |
| | sources of harm | *exposure (to mercury/thiomersal),exposed (to mercury)* |
| | diseases and infections: evaluative terms | *serious, common, mild, rare, severe, minor* |
| | diseases and infections: potential effects | *breathing (problems), painful (disease/swollen glands)* |
| | evaluation of information/research | *latest* |
| | other evaluative terms | *high, higher, important, key, low, properly* |
| | hedging and boosting | *usually, rarely, highly, frequently, extremely* |
| | status nouns: the ‘sign’ group | *evidence* |

| | | |
|---|---|---|
| | status nouns: factivity | *facts, issues, problems* |
| | status nouns: affect | *concerns* |
| | comparison/contrast | *or, compared, other* |
| **other nouns, verbs, and phrases referring to events, actions, and states of affairs** | verbal process/ attribution | *reported, concluded, reporting, statement* |
| | showing | *shows, shown* |
| | discovering | *finding, found, identified, identify* |
| | cognition | *awareness, (BE) aware, recognised, estimated, considered* |
| | occurring | *occur, occurred, occurs* |
| | durative | *continues* |
| | giving and receiving | *given, offered, receive, received, provides, receiving, provided, giving, provide, providing* |
| | starting and stopping | *avoid, starts* |
| | material processes: miscellaneous | *use, used, visit, using* |
| | relational processes | *contains, contain, containing, contains, include, includes* |
| | expression with meta-discoursal function | *listed* |
| **other grammatical items** | negative polarity | *doesn', none, don, 't* |
| | possessive marker | *'s* |
| | determiners and pronouns | *any, most, these, this, those, your* |
| | prepositions and adverbs | *against, also, (in) addition (to), between, including, for, under, below, over* |
| | *of* | *of* |
| | WH words | *how, who* |
| | delexicalised verbs/ auxiliaries / cupola | *are, be, been, being, does, has, have, is,* |
| | numerals | *#, second, ten, three* |
| **miscellaneous items** | politeness markers | *please* |
| | additive expressions | *additional, further* |
| | miscellaneous pre-and post-modifiers | *direct, usually, personal, specific, independent, new, other, individual, relevant, sufficient* |
| | miscellaneous nouns | *background, total* |
| | electronic communication | *www, http, org, download, com, htm, dh, en, gov, website, doh, gsi, http, pdf, co* |
| | abbreviations | *d, e, m, j* |

**Appendix 4** The one hundred most frequent three to six-word clusters from concordance of *mmr* in *NHSvax* corpus, with frequencies

| Rank | Cluster | Frequency |
|---|---|---|
| 1 | THE MMR VACCINE | 505 |
| 2 | MMR VACCINE AND | 181 |
| 3 | MMR AND AUTISM | 162 |
| 4 | BETWEEN MMR AND | 149 |
| 5 | LINK BETWEEN MMR | 146 |
| 6 | OF MMR VACCINE | 130 |
| 7 | OF THE MMR | 123 |
| 8 | BETWEEN MMR AND AUTISM | 114 |
| 9 | LINK BETWEEN MMR AND | 104 |
| 10 | LINK BETWEEN MMR AND AUTISM | 91 |
| 11 | DOSE OF MMR | 85 |
| 12 | MMR VACCINE WAS | 78 |
| 13 | MMR VACCINE AND AUTISM | 77 |
| 14 | VACCINE AND AUTISM | 77 |
| 15 | OF THE MMR VACCINE | 76 |
| 16 | MUMPS AND RUBELLA | 70 |
| 17 | BETWEEN MMR VACCINE | 70 |
| 18 | DOSES OF MMR | 68 |
| 19 | MEASLES MUMPS AND | 67 |
| 20 | A LINK BETWEEN | 67 |
| 21 | ASSOCIATION BETWEEN MMR | 67 |
| 22 | MEASLES MUMPS AND RUBELLA | 64 |
| 23 | BETWEEN MMR VACCINE AND | 62 |
| 24 | MMR VACCINE IS | 61 |
| 25 | MMR VACCINE THE | 60 |
| 26 | TWO DOSES OF | 59 |
| 27 | THE MMR VACCINE AND | 56 |
| 28 | THAT MMR IS | 52 |
| 29 | TWO DOSES OF MMR | 52 |
| 30 | A LINK BETWEEN MMR | 52 |
| 31 | VACCINE WAS INTRODUCED | 50 |
| 32 | MMR VACCINE WAS INTRODUCED | 50 |
| 33 | MMR VACCINATION AND | 49 |
| 34 | THE MMR VACCINE WAS | 48 |
| 35 | SAFETY OF MMR | 47 |
| 36 | MMR THE FACTS | 46 |
| 37 | THE SAFETY OF | 42 |
| 38 | SECOND DOSE OF | 42 |
| 39 | A LINK BETWEEN MMR AND | 41 |
| 40 | TO THE MMR | 41 |
| 41 | MMR VACCINE HAS | 41 |
| 42 | MUMPS AND RUBELLA MMR | 39 |
| 43 | AND RUBELLA MMR | 39 |
| 44 | MEASLES MUMPS AND RUBELLA MMR | 39 |
| 45 | THE MMR VACCINATION | 39 |
| 46 | THE MMR VACCINE THE | 38 |

| | | |
|---|---|---|
| 47 | THE MMR VACCINE WAS INTRODUCED | 38 |
| 48 | A LINK BETWEEN MMR AND AUTISM | 38 |
| 49 | AN ASSOCIATION BETWEEN MMR | 37 |
| 50 | SECOND DOSE OF MMR | 37 |
| 51 | NO LINK BETWEEN | 37 |
| 52 | THE USE OF | 37 |
| 53 | AN ASSOCIATION BETWEEN | 37 |
| 54 | THE INTRODUCTION OF | 36 |
| 55 | OF MMR AND | 36 |
| 56 | MMR VACCINE IN | 35 |
| 57 | USE OF MMR | 35 |
| 58 | EVIDENCE THAT MMR | 34 |
| 59 | BETWEEN MMR VACCINE AND AUTISM | 34 |
| 60 | IN THE MMR | 33 |
| 61 | AND THE MMR | 33 |
| 62 | A SECOND DOSE | 33 |
| 63 | THAT MMR VACCINE | 32 |
| 64 | A SECOND DOSE OF | 32 |
| 65 | BETWEEN THE MMR | 31 |
| 66 | THAT THE MMR | 30 |
| 67 | THE USE OF MMR | 30 |
| 68 | THE MMR VACCINE AND AUTISM | 30 |
| 69 | MMR AND THE | 30 |
| 70 | LINK BETWEEN THE | 30 |
| 71 | THE SAFETY OF MMR | 29 |
| 72 | AND THE MMR VACCINE | 29 |
| 73 | IN THE UK | 28 |
| 74 | IMMUNISED WITH MMR | 28 |
| 75 | WITH THE MMR | 28 |
| 76 | BETWEEN THE MMR VACCINE AND | 28 |
| 77 | BETWEEN THE MMR VACCINE | 28 |
| 78 | ABOUT THE MMR | 28 |
| 79 | LINK BETWEEN THE MMR | 28 |
| 80 | A SECOND DOSE OF MMR | 28 |
| 81 | DOSE OF MMR VACCINE | 28 |
| 82 | INTRODUCTION OF MMR | 28 |
| 83 | OF A LINK | 28 |
| 84 | THERE IS NO | 28 |
| 85 | NO LINK BETWEEN MMR | 27 |
| 86 | HAVE THE MMR | 27 |
| 87 | AND AUTISM AND | 27 |
| 88 | THE MMR IMMUNISATION | 27 |
| 89 | BETWEEN MMR VACCINATION AND | 27 |
| 90 | WAS INTRODUCED IN | 27 |
| 91 | BETWEEN MMR VACCINATION | 27 |
| 92 | THE MMR VACCINE IS | 26 |
| 93 | ASSOCIATION BETWEEN MMR AND | 26 |
| 94 | VACCINE HAS BEEN | 26 |
| 95 | MMR IS THE | 26 |
| 96 | AND AUTISM THE | 26 |

| | | |
|---|---|---|
| 97 | BETWEEN THE MMR VACCINE AND AUTISM | 26 |
| 98 | OF A LINK BETWEEN | 26 |
| 99 | MMR VACCINE HAS BEEN | 26 |
| 100 | LINK BETWEEN THE MMR VACCINE | 25 |

**Appendix 5** The one hundred most frequent three- to six-word clusters from concordance of *autism* in *NHSvax* corpus, with frequencies

| Rank | Cluster | Frequency |
|---|---|---|
| 1 | MMR AND AUTISM | 161 |
| 2 | LINK BETWEEN MMR | 116 |
| 3 | BETWEEN MMR AND | 114 |
| 4 | BETWEEN MMR AND AUTISM | 109 |
| 5 | LINK BETWEEN MMR AND | 91 |
| 6 | LINK BETWEEN MMR AND AUTISM | 90 |
| 7 | VACCINE AND AUTISM | 88 |
| 8 | MMR VACCINE AND | 80 |
| 9 | MMR VACCINE AND AUTISM | 76 |
| 10 | CHILDREN WITH AUTISM | 67 |
| 11 | OF AUTISM IN | 59 |
| 12 | THE MMR VACCINE | 58 |
| 13 | A LINK BETWEEN | 46 |
| 14 | THE INCIDENCE OF | 42 |
| 15 | INFLAMMATORY BOWEL DISEASE | 42 |
| 16 | PREVALENCE OF AUTISM | 41 |
| 17 | INCIDENCE OF AUTISM | 41 |
| 18 | THE INCIDENCE OF AUTISM | 40 |
| 19 | A LINK BETWEEN MMR | 39 |
| 20 | INCREASE IN AUTISM | 38 |
| 21 | A LINK BETWEEN MMR AND | 38 |
| 22 | A LINK BETWEEN MMR AND AUTISM | 38 |
| 23 | OF AUTISM AND | 37 |
| 24 | RISK OF AUTISM | 36 |
| 25 | AND AUTISM AND | 34 |
| 26 | CAUSES OF AUTISM | 34 |
| 27 | BETWEEN MMR VACCINE | 33 |
| 28 | BETWEEN MMR VACCINE AND AUTISM | 33 |
| 29 | MUMPS AND RUBELLA | 33 |
| 30 | BETWEEN MMR VACCINE AND | 33 |
| 31 | CASES OF AUTISM | 32 |
| 32 | THE MMR VACCINE AND | 30 |
| 33 | THE MMR VACCINE AND AUTISM | 30 |
| 34 | VACCINATION AND AUTISM | 30 |
| 35 | AUTISM SPECTRUM DISORDERS | 29 |
| 36 | THE PREVALENCE OF AUTISM | 28 |
| 37 | AND AUTISM THE | 28 |
| 38 | THE PREVALENCE OF | 28 |
| 39 | BETWEEN THE MMR | 27 |
| 40 | DIAGNOSIS OF AUTISM | 26 |
| 41 | BETWEEN THE MMR VACCINE | 26 |
| 42 | BETWEEN THE MMR VACCINE AND | 26 |
| 43 | BETWEEN THE MMR VACCINE AND AUTISM | 26 |
| 44 | BOWEL DISEASE AND | 25 |
| 45 | IN CHILDREN WITH | 25 |
| 46 | AUTISM AND THE | 25 |

| | | |
|---|---|---|
| 47 | NO LINK BETWEEN | 24 |
| 48 | VACCINES AND AUTISM | 24 |
| 49 | OF AUTISM THE | 23 |
| 50 | ASSOCIATION BETWEEN MMR | 23 |
| 51 | RATE OF AUTISM | 22 |
| 52 | AND BOWEL DISEASE | 22 |
| 53 | IN CHILDREN WITH AUTISM | 21 |
| 54 | INCREASED RISK OF | 21 |
| 55 | DEVELOPMENT OF AUTISM | 21 |
| 56 | NO LINK BETWEEN MMR AND | 20 |
| 57 | NO LINK BETWEEN MMR AND AUTISM | 20 |
| 58 | NO LINK BETWEEN MMR | 20 |
| 59 | FORM OF AUTISM | 20 |
| 60 | LINKED TO AUTISM | 20 |
| 61 | THE DEVELOPMENT OF AUTISM | 19 |
| 62 | THE DEVELOPMENT OF | 19 |
| 63 | DISEASE AND AUTISM | 19 |
| 64 | DIAGNOSED WITH AUTISM | 18 |
| 65 | NOT CAUSE AUTISM | 18 |
| 66 | THE RATE OF | 18 |
| 67 | AUTISM AND BOWEL | 18 |
| 68 | DOES NOT CAUSE AUTISM | 18 |
| 69 | AUTISM AND BOWEL DISEASE | 18 |
| 70 | DOES NOT CAUSE | 18 |
| 71 | INCREASE IN THE | 18 |
| 72 | OF AUTISM SPECTRUM | 18 |
| 73 | AUTISM IN CHILDREN | 18 |
| 74 | AND AUTISM OR | 17 |
| 75 | OF AUTISM SPECTRUM DISORDERS | 17 |
| 76 | BOWEL DISEASE OR | 17 |
| 77 | LINK BETWEEN MMR VACCINE | 17 |
| 78 | LINK BETWEEN MMR VACCINE AND | 17 |
| 79 | LINK BETWEEN MMR VACCINE AND AUTISM | 17 |
| 80 | RISE IN AUTISM | 17 |
| 81 | MMR VACCINATION AND AUTISM | 17 |
| 82 | MMR VACCINATION AND | 17 |
| 83 | INCREASED RISK OF AUTISM | 17 |
| 84 | MMR AND AUTISM AND | 17 |
| 85 | AN INCREASE IN | 16 |
| 86 | AND AUTISM IN | 16 |
| 87 | THE INCREASE IN | 16 |
| 88 | AND AUTISM THIS | 16 |
| 89 | AN INCREASE IN AUTISM | 16 |
| 90 | OF REGRESSIVE AUTISM | 16 |
| 91 | WITH AUTISM AND | 16 |
| 92 | DISEASE OR AUTISM | 16 |
| 93 | THE RATE OF AUTISM | 16 |
| 94 | SIGNS OF AUTISM | 15 |
| 95 | CHILDREN DIAGNOSED WITH AUTISM | 15 |
| 96 | CHILDREN DIAGNOSED WITH | 15 |

| | | |
|---|---|---|
| 97 | BOWEL DISEASE OR AUTISM | 15 |
| 98 | AND THE DEVELOPMENT | 15 |
| 99 | OF CHILDREN DIAGNOSED WITH AUTISM | 15 |
| 100 | AND THE DEVELOPMENT OF AUTISM | 15 |

**Appendix 6** The 200 most frequent words in the *JABS* corpus

| Rank | Word | Frequency | Percentage of corpus | No. of texts | Percentage of texts |
|---|---|---|---|---|---|
| 1 | THE | 201588 | 4.8542 | 1994 | 94.907 |
| 2 | # | 130781 | 3.1492 | 2004 | 95.383 |
| 3 | OF | 108188 | 2.6051 | 1951 | 92.861 |
| 4 | TO | 101985 | 2.4558 | 1977 | 94.098 |
| 5 | AND | 90424 | 2.1774 | 1923 | 91.528 |
| 6 | A | 73342 | 1.7661 | 1928 | 91.766 |
| 7 | IN | 70897 | 1.7072 | 1914 | 91.099 |
| 8 | THAT | 51663 | 1.244 | 1821 | 86.673 |
| 9 | IS | 48966 | 1.1791 | 1837 | 87.435 |
| 10 | I | 35537 | 0.8557 | 1483 | 70.585 |
| 11 | FOR | 34166 | 0.8227 | 1849 | 88.006 |
| 12 | IT | 30424 | 0.7326 | 1712 | 81.485 |
| 13 | ON | 28391 | 0.6837 | 1789 | 85.15 |
| 14 | HAVE | 28319 | 0.6819 | 1731 | 82.389 |
| 15 | WITH | 27097 | 0.6525 | 1721 | 81.913 |
| 16 | ARE | 25184 | 0.6064 | 1668 | 79.391 |
| 17 | WAS | 24851 | 0.5984 | 1528 | 72.727 |
| 18 | NOT | 24634 | 0.5932 | 1621 | 77.154 |
| 19 | BY | 24561 | 0.5914 | 1728 | 82.247 |
| 20 | THIS | 24377 | 0.587 | 1754 | 83.484 |
| 21 | AS | 23825 | 0.5737 | 1678 | 79.867 |
| 22 | BE | 23449 | 0.5646 | 1691 | 80.485 |
| 23 | VACCINE | 22780 | 0.5485 | 1382 | 65.778 |
| 24 | YOU | 22014 | 0.5301 | 1266 | 60.257 |
| 25 | FROM | 20038 | 0.4825 | 1642 | 78.153 |
| 26 | CHILDREN | 16357 | 0.3939 | 1302 | 61.97 |
| 27 | OR | 16004 | 0.3854 | 1499 | 71.347 |
| 28 | AT | 15990 | 0.385 | 1589 | 75.631 |
| 29 | THEY | 15856 | 0.3818 | 1471 | 70.014 |
| 30 | HAS | 15166 | 0.3652 | 1590 | 75.678 |
| 31 | BUT | 14878 | 0.3583 | 1525 | 72.584 |
| 32 | HAD | 13003 | 0.3131 | 1305 | 62.113 |
| 33 | AN | 12887 | 0.3103 | 1536 | 73.108 |
| 34 | VACCINES | 12582 | 0.303 | 1204 | 57.306 |
| 35 | BEEN | 12420 | 0.2991 | 1457 | 69.348 |
| 36 | WE | 12227 | 0.2944 | 1315 | 62.589 |
| 37 | WHO | 12004 | 0.2891 | 1445 | 68.777 |
| 38 | IF | 11831 | 0.2849 | 1395 | 66.397 |
| 39 | Â | 11590 | 0.2791 | 819 | 38.981 |
| 40 | MMR | 11503 | 0.277 | 836 | 39.791 |
| 41 | ALL | 11467 | 0.2761 | 1405 | 66.873 |
| 42 | AUTISM | 11406 | 0.2747 | 945 | 44.979 |
| 43 | HE | 11364 | 0.2736 | 1097 | 52.213 |
| 44 | MY | 11239 | 0.2706 | 1074 | 51.119 |
| 45 | THERE | 11114 | 0.2676 | 1386 | 65.969 |
| 46 | WERE | 10893 | 0.2623 | 1289 | 61.352 |
| 47 | ABOUT | 10768 | 0.2593 | 1432 | 68.158 |
| 48 | WHICH | 10509 | 0.2531 | 1412 | 67.206 |
| 49 | THEIR | 10508 | 0.253 | 1376 | 65.493 |
| 50 | ONE | 10207 | 0.2458 | 1398 | 66.54 |
| 51 | NO | 10086 | 0.2429 | 1286 | 61.209 |

| | | | | | |
|---|---|---|---|---|---|
| 52 | WOULD | 9841 | 0.237 | 1259 | 59.924 |
| 53 | CAN | 9626 | 0.2318 | 1369 | 65.159 |
| 54 | SO | 9552 | 0.23 | 1294 | 61.59 |
| 55 | WILL | 9133 | 0.2199 | 1315 | 62.589 |
| 56 | HTTP | 8987 | 0.2164 | 1758 | 83.674 |
| 57 | WHAT | 8964 | 0.2159 | 1222 | 58.163 |
| 58 | YOUR | 8907 | 0.2145 | 948 | 45.121 |
| 59 | HEALTH | 8834 | 0.2127 | 1299 | 61.828 |
| 60 | MORE | 8737 | 0.2104 | 1344 | 63.97 |
| 61 | HIS | 8211 | 0.1977 | 973 | 46.311 |
| 62 | DO | 8098 | 0.195 | 1183 | 56.307 |
| 63 | ANY | 7720 | 0.1859 | 1190 | 56.64 |
| 64 | WWW | 7585 | 0.1826 | 1626 | 77.392 |
| 65 | S | 7492 | 0.1804 | 949 | 45.169 |
| 66 | VACCINATION | 7434 | 0.179 | 934 | 44.455 |
| 67 | WHEN | 7337 | 0.1767 | 1205 | 57.354 |
| 68 | ALSO | 7042 | 0.1696 | 1238 | 58.924 |
| 69 | DR | 7025 | 0.1692 | 959 | 45.645 |
| 70 | OTHER | 6966 | 0.1677 | 1211 | 57.639 |
| 71 | SAID | 6875 | 0.1655 | 1031 | 49.072 |
| 72 | THESE | 6767 | 0.1629 | 1184 | 56.354 |
| 73 | SOME | 6698 | 0.1613 | 1245 | 59.257 |
| 74 | PARENTS | 6676 | 0.1608 | 989 | 47.073 |
| 75 | UP | 6587 | 0.1586 | 1197 | 56.973 |
| 76 | SHE | 6533 | 0.1573 | 795 | 37.839 |
| 77 | OUT | 6422 | 0.1546 | 1189 | 56.592 |
| 78 | MEASLES | 6402 | 0.1542 | 592 | 28.177 |
| 79 | HER | 6372 | 0.1534 | 740 | 35.221 |
| 80 | AFTER | 6294 | 0.1516 | 1141 | 54.307 |
| 81 | UK | 6135 | 0.1477 | 967 | 46.026 |
| 82 | THAN | 6103 | 0.147 | 1131 | 53.832 |
| 83 | YEARS | 6079 | 0.1464 | 1125 | 53.546 |
| 84 | PEOPLE | 5934 | 0.1429 | 1051 | 50.024 |
| 85 | JUST | 5918 | 0.1425 | 1115 | 53.07 |
| 86 | CHILD | 5909 | 0.1423 | 924 | 43.979 |
| 87 | COM | 5856 | 0.141 | 1292 | 61.495 |
| 88 | ONLY | 5787 | 0.1394 | 1114 | 53.022 |
| 89 | MANY | 5782 | 0.1392 | 1121 | 53.356 |
| 90 | MEDICAL | 5726 | 0.1379 | 1063 | 50.595 |
| 91 | NOW | 5617 | 0.1353 | 1129 | 53.736 |
| 92 | MAY | 5453 | 0.1313 | 1124 | 53.498 |
| 93 | STUDY | 5411 | 0.1303 | 750 | 35.697 |
| 94 | DISEASE | 5392 | 0.1298 | 879 | 41.837 |
| 95 | OUR | 5365 | 0.1292 | 1005 | 47.834 |
| 96 | VERY | 5357 | 0.129 | 1033 | 49.167 |
| 97 | COULD | 5311 | 0.1279 | 1092 | 51.975 |
| 98 | KNOW | 5304 | 0.1277 | 1012 | 48.168 |
| 99 | LIKE | 5150 | 0.124 | 1025 | 48.786 |
| 100 | TIME | 5134 | 0.1236 | 1098 | 52.261 |
| 101 | HOW | 5037 | 0.1213 | 1015 | 48.31 |
| 102 | ME | 4981 | 0.1199 | 858 | 40.838 |
| 103 | BECAUSE | 4916 | 0.1184 | 1054 | 50.167 |
| 104 | THEM | 4875 | 0.1174 | 1034 | 49.215 |
| 105 | BEING | 4788 | 0.1153 | 1014 | 48.263 |
| 106 | SHOULD | 4770 | 0.1149 | 1022 | 48.644 |

| | | | | | |
|---|---|---|---|---|---|
| 107 | US | 4660 | 0.1122 | 983 | 46.787 |
| 108 | RESEARCH | 4574 | 0.1101 | 909 | 43.265 |
| 109 | INTO | 4525 | 0.109 | 1025 | 48.786 |
| 110 | EVIDENCE | 4504 | 0.1085 | 750 | 35.697 |
| 111 | YEAR | 4464 | 0.1075 | 993 | 47.263 |
| 112 | GET | 4408 | 0.1061 | 967 | 46.026 |
| 113 | ITS | 4405 | 0.1061 | 950 | 45.217 |
| 114 | THEN | 4283 | 0.1031 | 926 | 44.074 |
| 115 | RECEIVED | 4268 | 0.1028 | 602 | 28.653 |
| 116 | CASES | 4215 | 0.1015 | 734 | 34.936 |
| 117 | OVER | 4133 | 0.0995 | 1042 | 49.595 |
| 118 | EVEN | 4083 | 0.0983 | 973 | 46.311 |
| 119 | TWO | 4009 | 0.0965 | 961 | 45.74 |
| 120 | NEW | 4004 | 0.0964 | 1016 | 48.358 |
| 121 | MOST | 3959 | 0.0953 | 986 | 46.93 |
| 122 | AGAINST | 3873 | 0.0933 | 926 | 44.074 |
| 123 | FIRST | 3868 | 0.0931 | 984 | 46.835 |
| 124 | VIRUS | 3864 | 0.093 | 522 | 24.845 |
| 125 | DOES | 3820 | 0.092 | 870 | 41.409 |
| 126 | SUCH | 3811 | 0.0918 | 929 | 44.217 |
| 127 | THINK | 3796 | 0.0914 | 823 | 39.172 |
| 128 | CASE | 3751 | 0.0903 | 756 | 35.983 |
| 129 | WHY | 3751 | 0.0903 | 827 | 39.362 |
| 130 | CAUSE | 3742 | 0.0901 | 873 | 41.552 |
| 131 | WELL | 3724 | 0.0897 | 954 | 45.407 |
| 132 | POSTED | 3565 | 0.0858 | 545 | 25.94 |
| 133 | GIVEN | 3563 | 0.0858 | 840 | 39.981 |
| 134 | DID | 3554 | 0.0856 | 822 | 39.124 |
| 135 | THOSE | 3546 | 0.0854 | 941 | 44.788 |
| 136 | SEE | 3495 | 0.0842 | 839 | 39.933 |
| 137 | FOUND | 3463 | 0.0834 | 905 | 43.075 |
| 138 | AM | 3452 | 0.0831 | 758 | 36.078 |
| 139 | BEFORE | 3444 | 0.0829 | 901 | 42.884 |
| 140 | MERCURY | 3416 | 0.0823 | 419 | 19.943 |
| 141 | CANCER | 3412 | 0.0822 | 383 | 18.229 |
| 142 | RISK | 3347 | 0.0806 | 701 | 33.365 |
| 143 | INFORMATION | 3343 | 0.0805 | 821 | 39.077 |
| 144 | BETWEEN | 3298 | 0.0794 | 842 | 40.076 |
| 145 | SAY | 3293 | 0.0793 | 855 | 40.695 |
| 146 | DON'T | 3270 | 0.0787 | 737 | 35.079 |
| 147 | GOVERNMENT | 3165 | 0.0762 | 708 | 33.698 |
| 148 | SINGLE | 3152 | 0.0759 | 500 | 23.798 |
| 149 | STUDIES | 3046 | 0.0733 | 643 | 30.604 |
| 150 | GOOD | 3028 | 0.0729 | 799 | 38.03 |
| 151 | MAKE | 3022 | 0.0728 | 875 | 41.647 |
| 152 | OLD | 3006 | 0.0724 | 747 | 35.554 |
| 153 | IMMUNE | 2978 | 0.0717 | 565 | 26.892 |
| 154 | DOCTORS | 2925 | 0.0704 | 755 | 35.935 |
| 155 | MUCH | 2914 | 0.0702 | 878 | 41.79 |
| 156 | WHERE | 2909 | 0.07 | 804 | 38.267 |
| 157 | JABS | 2900 | 0.0698 | 535 | 25.464 |
| 158 | HERE | 2886 | 0.0695 | 777 | 36.982 |
| 159 | MESSAGE | 2886 | 0.0695 | 204 | 9.7097 |
| 160 | SYSTEM | 2883 | 0.0694 | 694 | 33.032 |
| 161 | VACCINATED | 2869 | 0.0691 | 607 | 28.891 |

| | | | | | |
|---|---|---|---|---|---|
| 162 | USED | 2851 | 0.0687 | 810 | 38.553 |
| 163 | AUTISTIC | 2844 | 0.0685 | 549 | 26.13 |
| 164 | MONTHS | 2844 | 0.0685 | 670 | 31.89 |
| 165 | PUBLIC | 2834 | 0.0682 | 751 | 35.745 |
| 166 | VACCINATIONS | 2806 | 0.0676 | 726 | 34.555 |
| 167 | SAME | 2801 | 0.0674 | 809 | 38.505 |
| 168 | MADE | 2791 | 0.0672 | 797 | 37.934 |
| 169 | ORIGINALLY | 2786 | 0.0671 | 448 | 21.323 |
| 170 | IT'S | 2763 | 0.0665 | 722 | 34.365 |
| 171 | CO | 2711 | 0.0653 | 771 | 36.697 |
| 172 | WAKEFIELD | 2703 | 0.0651 | 311 | 14.802 |
| 173 | STILL | 2694 | 0.0649 | 793 | 37.744 |
| 174 | EFFECTS | 2689 | 0.0648 | 668 | 31.794 |
| 175 | DISEASES | 2670 | 0.0643 | 675 | 32.128 |
| 176 | USE | 2640 | 0.0636 | 799 | 38.03 |
| 177 | WAY | 2634 | 0.0634 | 822 | 39.124 |
| 178 | LAST | 2622 | 0.0631 | 866 | 41.218 |
| 179 | HIM | 2595 | 0.0625 | 566 | 26.94 |
| 180 | SON | 2591 | 0.0624 | 514 | 24.465 |
| 181 | BRAIN | 2575 | 0.062 | 515 | 24.512 |
| 182 | LINK | 2563 | 0.0617 | 703 | 33.46 |
| 183 | JAB | 2555 | 0.0615 | 424 | 20.181 |
| 184 | DEATH | 2498 | 0.0602 | 532 | 25.321 |
| 185 | NEVER | 2470 | 0.0595 | 717 | 34.127 |
| 186 | AGE | 2455 | 0.0591 | 650 | 30.938 |
| 187 | HTML | 2454 | 0.0591 | 881 | 41.932 |
| 188 | ANYONE | 2420 | 0.0583 | 664 | 31.604 |
| 189 | DAMAGE | 2412 | 0.0581 | 567 | 26.987 |
| 190 | SAFETY | 2412 | 0.0581 | 571 | 27.178 |
| 191 | TAKE | 2411 | 0.0581 | 725 | 34.507 |
| 192 | ORG | 2396 | 0.0577 | 725 | 34.507 |
| 193 | SINCE | 2392 | 0.0576 | 757 | 36.03 |
| 194 | NEED | 2387 | 0.0575 | 762 | 36.268 |
| 195 | PROBLEMS | 2385 | 0.0574 | 636 | 30.271 |
| 196 | THREE | 2383 | 0.0574 | 726 | 34.555 |
| 197 | DRUG | 2368 | 0.057 | 551 | 26.226 |
| 198 | NEWS | 2324 | 0.056 | 814 | 38.743 |
| 199 | GO | 2311 | 0.0556 | 682 | 32.461 |
| 200 | ADVERSE | 2300 | 0.0554 | 455 | 21.656 |

**Appendix 7** Keywords in *JABS* corpus occurring in at least 10% of corpus files, with *BNC World* as reference corpus and log-likelihood as measure of significance

| Rank | Keyword | Freq. | LL | Rank | Keyword | Freq. | LL |
|---|---|---|---|---|---|---|---|
| 1 | VACCINE | 22209 | 140465.92 | 51 | HTM | 1159 | 7524.31 |
| 2 | S | 23711 | 90568.63 | 52 | CAUSE | 3735 | 7492.72 |
| 3 | VACCINES | 12665 | 80428.20 | 53 | REACTIONS | 1982 | 7476.58 |
| 4 | Â | 11604 | 75362.92 | 54 | LINK | 2581 | 7252.92 |
| 5 | MMR | 11424 | 74138.18 | 55 | MERCK | 1179 | 7252.72 |
| 6 | AUTISM | 11253 | 72669.05 | 56 | BRAIN | 2460 | 7223.72 |
| 7 | HTTP | 9020 | 58575.40 | 57 | CASES | 4217 | 7165.53 |
| 8 | WWW | 7605 | 49347.08 | 58 | EVIDENCE | 4442 | 6890.98 |
| 9 | VACCINATION | 7330 | 45508.27 | 59 | FDA | 1124 | 6792.30 |
| 10 | # | 126712 | 44425.13 | 60 | MEDICINE | 1999 | 6754.48 |
| 11 | T | 11598 | 43839.26 | 61 | DRUG | 2334 | 6383.59 |
| 12 | CHILDREN | 16954 | 43115.14 | 62 | CO | 2331 | 6344.93 |
| 13 | MEASLES | 6226 | 38394.25 | 63 | IMMUNITY | 1336 | 6252.92 |
| 14 | COM | 5882 | 35930.85 | 64 | PHARMA | 960 | 6170.14 |
| 15 | DR | 7046 | 21423.67 | 65 | RISK | 3188 | 6138.10 |
| 16 | HEALTH | 8657 | 20114.30 | 66 | IMMUNISATION | 1056 | 6131.03 |
| 17 | POSTED | 3567 | 19370.64 | 67 | TETANUS | 1071 | 6088.39 |
| 18 | JABS | 2935 | 18543.27 | 68 | MENINGITIS | 1118 | 6037.76 |
| 19 | VIRUS | 3800 | 18495.59 | 69 | DISORDERS | 1365 | 6009.00 |
| 20 | VACCINATIONS | 2823 | 17860.38 | 70 | DOESN | 927 | 5936.20 |
| 21 | MEDICAL | 5672 | 17754.29 | 71 | DEATHS | 1746 | 5861.75 |
| 22 | WAKEFIELD | 3044 | 17547.07 | 72 | VACCINATE | 911 | 5770.21 |
| 23 | AUTISTIC | 2760 | 17261.23 | 73 | RESEARCH | 4556 | 5738.53 |
| 24 | VACCINATED | 2738 | 17126.79 | 74 | WEBSITE | 882 | 5725.95 |
| 25 | PARENTS | 6663 | 17095.83 | 75 | DIDN | 897 | 5639.96 |
| 26 | DISEASE | 5349 | 16605.21 | 76 | IMMUNIZATION | 918 | 5618.01 |
| 27 | HTML | 2459 | 15964.76 | 77 | LL | 977 | 5478.46 |
| 28 | MERCURY | 3156 | 15795.77 | 78 | RE | 1824 | 5473.62 |
| 29 | JAB | 2562 | 15656.75 | 79 | HI | 1263 | 5155.34 |
| 30 | ORG | 2398 | 15367.65 | 80 | INFECTION | 1638 | 5095.73 |
| 31 | DON | 3342 | 15202.61 | 81 | GOV | 800 | 5078.00 |
| 32 | UK | 6184 | 14865.57 | 82 | STUDIES | 3044 | 5015.14 |
| 33 | IMMUNE | 2790 | 13902.81 | 83 | SON | 2875 | 4997.85 |
| 34 | CHILD | 6372 | 13301.92 | 84 | M | 3469 | 4987.53 |
| 35 | MUMPS | 1870 | 11876.40 | 85 | BR | 1301 | 4968.68 |
| 36 | FLU | 2237 | 11664.86 | 86 | INFECTIONS | 1126 | 4925.41 |
| 37 | CANCER | 3309 | 11500.59 | 87 | DOSE | 1374 | 4895.60 |
| 38 | POLIO | 1926 | 11438.83 | 88 | DAMAGE | 2410 | 4878.89 |
| 39 | DISEASES | 2672 | 11412.06 | 89 | SAFETY | 2398 | 4749.90 |
| 40 | THIMEROSAL | 1731 | 11221.18 | 90 | SYMPTOMS | 1633 | 4734.27 |
| 41 | ADVERSE | 2295 | 10538.31 | 91 | ISN | 747 | 4727.12 |
| 42 | RUBELLA | 1680 | 10479.13 | 92 | HAVE | 28319 | 4634.30 |
| 43 | CDC | 1657 | 10139.69 | 93 | DONT | 817 | 4522.20 |
| 44 | STUDY | 5470 | 9894.13 | 94 | DOSES | 1044 | 4517.37 |
| 45 | VIRUSES | 1878 | 9828.97 | 95 | ARTICLE | 2052 | 4363.81 |
| 46 | EDITED | 2071 | 9555.11 | 96 | GP | 1079 | 4331.25 |
| 47 | DOCTORS | 2927 | 9456.05 | 97 | SYNDROME | 1133 | 4191.18 |
| 48 | RECEIVED | 4264 | 9373.32 | 98 | PHARMACEUTICAL | 921 | 4187.65 |
| 49 | VE | 1725 | 9066.69 | 99 | DOCTOR | 2322 | 4169.80 |
| 50 | ORIGINALLY | 2786 | 8802.32 | 100 | INFANTS | 1009 | 4052.97 |

| 101 | CHILDHOOD | 1401 | 3996.86 | 151 | NEWS | 2229 | 2554.44 |
|---|---|---|---|---|---|---|---|
| 102 | SINGLE | 3134 | 3993.62 | 152 | NOT | 24616 | 2494.57 |
| 103 | ID | 882 | 3968.54 | 153 | KIDS | 1222 | 2480.98 |
| 104 | DIAGNOSED | 943 | 3887.86 | 154 | D | 2266 | 2470.04 |
| 105 | PDF | 641 | 3882.89 | 155 | WASN | 389 | 2455.53 |
| 106 | FORUM | 1189 | 3880.00 | 156 | EXPOSURE | 951 | 2431.66 |
| 107 | EPIDEMIC | 876 | 3854.79 | 157 | MEDICINES | 624 | 2387.56 |
| 108 | CHRONIC | 1158 | 3804.77 | 158 | BLOOD | 1752 | 2328.03 |
| 109 | EFFECTS | 2332 | 3724.90 | 159 | ET | 1308 | 2325.88 |
| 110 | WHALE | 853 | 3718.56 | 160 | CAUSES | 1202 | 2260.93 |
| 111 | DISORDER | 1111 | 3653.66 | 161 | AL | 1324 | 2259.26 |
| 112 | CLINICAL | 1358 | 3651.24 | 162 | CENTER | 566 | 2236.14 |
| 113 | DRUGS | 1684 | 3636.24 | 163 | ETC | 1486 | 2191.76 |
| 114 | BOWEL | 1018 | 3585.58 | 164 | CELLS | 1503 | 2191.07 |
| 115 | MY | 11258 | 3573.44 | 165 | SERIOUS | 1901 | 2179.84 |
| 116 | DAMAGED | 1254 | 3515.56 | 166 | THAT | 52651 | 2123.83 |
| 117 | ADMINISTERED | 902 | 3338.91 | 167 | DEATH | 2488 | 2120.40 |
| 118 | PHP | 517 | 3319.65 | 168 | MONTHS | 2820 | 2118.66 |
| 119 | AASA | 504 | 3271.92 | 169 | PLEASE | 2016 | 2113.37 |
| 120 | ANYONE | 2457 | 3197.29 | 170 | PUBLISHED | 1749 | 2107.85 |
| 121 | AM | 3452 | 3185.48 | 171 | HEALTHY | 1033 | 2105.76 |
| 122 | VIRAL | 697 | 3150.74 | 172 | CENTERS | 367 | 2101.91 |
| 123 | REPORTED | 2235 | 3107.02 | 173 | TOXIC | 695 | 2099.83 |
| 124 | RESEARCHERS | 1144 | 3093.41 | 174 | PROFESSOR | 1185 | 2098.07 |
| 125 | INFO | 633 | 3091.70 | 175 | BABY | 1507 | 2096.42 |
| 126 | SAFE | 1686 | 3068.46 | 176 | DIAGNOSIS | 777 | 2085.72 |
| 127 | CGI | 518 | 3062.72 | 177 | IS | 48957 | 2050.27 |
| 128 | CAUSED | 1950 | 2994.89 | 178 | ARE | 25202 | 2047.69 |
| 129 | WOULDN | 466 | 2979.45 | 179 | FINDINGS | 981 | 2032.41 |
| 130 | VACCINATING | 468 | 2974.33 | 180 | KILLS | 536 | 2027.03 |
| 131 | U | 1378 | 2973.63 | 181 | RISKS | 884 | 2019.64 |
| 132 | SCIENTIFIC | 1550 | 2947.45 | 182 | SEVERE | 1120 | 1989.67 |
| 133 | ILLNESS | 1216 | 2937.15 | 183 | AGED | 1129 | 1982.40 |
| 134 | NEUROLOGICAL | 573 | 2892.10 | 184 | HAS | 15167 | 1978.45 |
| 135 | BABIES | 1065 | 2840.16 | 185 | WHO | 12161 | 1966.99 |
| 136 | JOURNAL | 1047 | 2799.24 | 186 | TRUTH | 1433 | 1916.07 |
| 137 | ASP | 494 | 2797.75 | 187 | ILLNESSES | 500 | 1885.11 |
| 138 | DAUGHTER | 1806 | 2797.15 | 188 | DATA | 2236 | 1866.76 |
| 139 | THANKS | 1568 | 2783.14 | 189 | COMPLICATIONS | 561 | 1824.30 |
| 140 | FEVER | 808 | 2777.36 | 190 | PERCENT | 866 | 1779.49 |
| 141 | REACTION | 1451 | 2761.13 | 191 | YOUR | 8889 | 1744.76 |
| 142 | INJECTION | 787 | 2728.11 | 192 | PARENT | 925 | 1715.40 |
| 143 | I | 39280 | 2720.62 | 193 | WARNINGS | 521 | 1703.01 |
| 144 | NHS | 1042 | 2703.76 | 194 | THIS | 24397 | 1698.53 |
| 145 | SCIENCE | 1958 | 2702.87 | 195 | SCIENTISTS | 907 | 1689.78 |
| 146 | INCIDENCE | 917 | 2684.10 | 196 | SPECTRUM | 659 | 1680.86 |
| 147 | TRIALS | 976 | 2665.28 | 197 | AGE | 2373 | 1657.17 |
| 148 | INFECTIOUS | 633 | 2643.45 | 198 | TREATMENT | 1650 | 1578.14 |
| 149 | GENETIC | 922 | 2623.35 | 199 | ARTICLES | 803 | 1566.71 |
| 150 | SHOTS | 907 | 2607.68 | 200 | GIRLS | 1394 | 1553.69 |

| 201 | CONCERNS | 894 | 1536.46 | 251 | CONDITION | 1061 | 927.96 |
|---|---|---|---|---|---|---|---|
| 202 | GIVEN | 3555 | 1525.12 | 252 | TESTING | 653 | 926.53 |
| 203 | BELIEVE | 2206 | 1480.95 | 253 | TESTS | 805 | 926.16 |
| 204 | SITE | 1413 | 1476.44 | 254 | REPORTS | 1258 | 922.81 |
| 205 | REGARDING | 719 | 1470.11 | 255 | ONLINE | 314 | 912.33 |
| 206 | PATIENTS | 1983 | 1470.04 | 256 | ASSOCIATED | 1114 | 902.27 |
| 207 | INFORMATION | 3327 | 1442.25 | 257 | MULTIPLE | 546 | 890.10 |
| 208 | EXPERTS | 797 | 1436.86 | 258 | PAPER | 1612 | 876.00 |
| 209 | LINKED | 899 | 1377.13 | 259 | PATIENT | 958 | 874.52 |
| 210 | CAUSING | 770 | 1345.09 | 260 | CELL | 799 | 848.14 |
| 211 | YEARS | 6059 | 1332.70 | 261 | MEDIA | 985 | 838.40 |
| 212 | ANDREW | 854 | 1323.94 | 262 | BBC | 661 | 832.28 |
| 213 | HAVEN | 415 | 1322.54 | 263 | REVIEW | 1106 | 830.20 |
| 214 | CASE | 3639 | 1313.64 | 264 | COMMENT | 783 | 829.30 |
| 215 | POST | 1310 | 1288.51 | 265 | EXPERT | 680 | 828.07 |
| 216 | HOSPITAL | 1702 | 1281.44 | 266 | PREVENTION | 422 | 818.57 |
| 217 | HEARING | 953 | 1274.21 | 267 | ALTERNATIVE | 1025 | 810.77 |
| 218 | ANY | 7728 | 1271.06 | 268 | CARE | 2038 | 807.34 |
| 219 | PROTECT | 925 | 1256.86 | 269 | ACCURATE | 558 | 799.99 |
| 220 | INFANT | 545 | 1248.02 | 270 | PUBLIC | 2815 | 781.86 |
| 221 | MANUFACTURERS | 681 | 1218.30 | 271 | EXPOSED | 526 | 774.27 |
| 222 | WHY | 3746 | 1218.13 | 272 | ADVICE | 1126 | 769.68 |
| 223 | DIED | 1582 | 1198.26 | 273 | INSTITUTE | 729 | 763.40 |
| 224 | DOES | 3823 | 1187.91 | 274 | RECOMMENDED | 620 | 760.84 |
| 225 | ADULTS | 728 | 1183.30 | 275 | RATES | 1202 | 755.40 |
| 226 | TESTED | 663 | 1173.92 | 276 | REPORT | 2158 | 745.45 |
| 227 | COMMENTS | 823 | 1164.46 | 277 | TRIAL | 820 | 724.11 |
| 228 | FAMILIES | 1156 | 1154.50 | 278 | OFFICIALS | 795 | 713.56 |
| 229 | TOPIC | 620 | 1138.00 | 279 | ALSO | 7042 | 702.99 |
| 230 | READ | 2102 | 1100.46 | 280 | LINKS | 648 | 702.70 |
| 231 | PROGRAM | 773 | 1100.17 | 281 | LIVE | 1506 | 697.39 |
| 232 | MANY | 5781 | 1073.39 | 282 | DANGEROUS | 739 | 671.79 |
| 233 | DUE | 1634 | 1064.92 | 283 | FOLLOWING | 2039 | 639.34 |
| 234 | PROBLEMS | 2385 | 1060.55 | 284 | SMOKING | 499 | 637.21 |
| 235 | HUMAN | 1879 | 1044.88 | 285 | AFFECTED | 741 | 629.05 |
| 236 | RECEIVING | 668 | 1019.49 | 286 | PREVENT | 793 | 622.29 |
| 237 | INFORMED | 714 | 997.51 | 287 | RESULTS | 1356 | 614.04 |
| 238 | HARM | 618 | 991.84 | 288 | BY | 24521 | 606.62 |
| 239 | ABOUT | 10772 | 984.60 | 289 | FULLY | 932 | 603.50 |
| 240 | CONSENT | 685 | 980.18 | 290 | CLAIMS | 936 | 599.41 |
| 241 | PROTECTION | 1059 | 979.86 | 291 | STORY | 1214 | 597.69 |
| 242 | SCHEDULE | 576 | 976.32 | 292 | STATED | 665 | 594.22 |
| 243 | FOUND | 3467 | 966.69 | 293 | PROOF | 460 | 591.10 |
| 244 | RARE | 774 | 966.43 | 294 | COMBINED | 604 | 584.01 |
| 245 | ILL | 789 | 956.59 | 295 | THOUSANDS | 676 | 579.87 |
| 246 | AGAINST | 3884 | 941.24 | 296 | OUR | 5391 | 574.90 |
| 247 | RESPONSE | 1323 | 939.93 | 297 | MILLIONS | 453 | 574.38 |
| 248 | REPORTING | 578 | 938.54 | 298 | SUFFERED | 664 | 567.17 |
| 249 | QUOTE | 446 | 936.46 | 299 | LEVELS | 1118 | 561.29 |
| 250 | DEPARTMENT | 1670 | 929.17 | 300 | PROVE | 680 | 553.11 |

| 301 | CONDUCTED | 474 | 552.78 | 351 | MENTION | 493 | 338.48 |
|---|---|---|---|---|---|---|---|
| 302 | B | 1906 | 552.54 | 352 | RECEIVE | 681 | 333.87 |
| 303 | EFFECTIVE | 974 | 550.18 | 353 | EVENTS | 855 | 333.42 |
| 304 | SYSTEM | 2890 | 548.24 | 354 | SICK | 478 | 331.77 |
| 305 | FACTS | 645 | 541.85 | 355 | INCREASE | 1220 | 328.39 |
| 306 | AFTER | 6323 | 538.14 | 356 | USA | 515 | 325.91 |
| 307 | DEVELOPED | 1112 | 537.01 | 357 | CONTAIN | 483 | 323.45 |
| 308 | STONE | 818 | 532.61 | 358 | HAVING | 2125 | 321.45 |
| 309 | THESE | 6761 | 529.53 | 359 | TEST | 1035 | 318.40 |
| 310 | ISSUE | 1347 | 516.24 | 360 | NORMAL | 953 | 316.64 |
| 311 | JOHN | 2161 | 512.32 | 361 | COMPARED | 749 | 314.71 |
| 312 | POPULATION | 1141 | 508.99 | 362 | WEEKS | 1114 | 314.29 |
| 313 | INJURY | 586 | 508.45 | 363 | SUGGEST | 748 | 310.69 |
| 314 | SUFFERING | 551 | 503.51 | 364 | ADMINISTRATION | 598 | 304.35 |
| 315 | US | 4642 | 503.01 | 365 | EDITOR | 415 | 304.33 |
| 316 | OPINION | 780 | 501.40 | 366 | BEEN | 12417 | 301.35 |
| 317 | DEVELOPING | 695 | 501.13 | 367 | LIVES | 817 | 290.11 |
| 318 | POTENTIALLY | 406 | 500.36 | 368 | POTENTIAL | 867 | 287.48 |
| 319 | INVESTIGATION | 614 | 499.26 | 369 | YEAR | 3728 | 277.05 |
| 320 | DIE | 634 | 496.76 | 370 | CONNECTION | 489 | 274.03 |
| 321 | AGO | 1503 | 494.34 | 371 | HELP | 2197 | 273.23 |
| 322 | INCLUDING | 1758 | 486.32 | 372 | HOW | 5065 | 263.16 |
| 323 | AUTHOR | 515 | 480.55 | 373 | STATEMENT | 765 | 256.05 |
| 324 | DEBATE | 753 | 467.81 | 374 | FEBRUARY | 681 | 255.72 |
| 325 | SHOT | 790 | 454.81 | 375 | CLAIM | 842 | 252.95 |
| 326 | CAN | 10633 | 451.26 | 376 | UNIVERSITY | 1093 | 252.65 |
| 327 | AGENCY | 631 | 449.34 | 377 | PRIOR | 457 | 249.00 |
| 328 | LEGAL | 1089 | 434.60 | 378 | KNOWN | 1568 | 240.56 |
| 329 | CONTENT | 644 | 426.82 | 379 | BAD | 1032 | 236.06 |
| 330 | FROM | 20078 | 421.85 | 380 | UNFORTUNATELY | 435 | 232.44 |
| 331 | WRONG | 1225 | 420.44 | 381 | CONCLUSION | 464 | 231.25 |
| 332 | INTERESTING | 864 | 420.06 | 382 | FREE | 1321 | 231.10 |
| 333 | J | 916 | 410.32 | 383 | OCCUR | 494 | 227.69 |
| 334 | QUESTIONS | 1133 | 409.66 | 384 | PAGE | 793 | 225.60 |
| 335 | BIRTH | 574 | 409.03 | 385 | INTRODUCED | 668 | 223.66 |
| 336 | GOVERNMENT | 3335 | 405.75 | 386 | NOTE | 775 | 223.62 |
| 337 | ACCORDING | 1220 | 401.83 | 387 | IF | 11829 | 213.47 |
| 338 | ASSOCIATION | 967 | 391.78 | 388 | TRUST | 728 | 212.69 |
| 339 | CONFIRMED | 543 | 384.92 | 389 | RECENTLY | 859 | 210.95 |
| 340 | TREAT | 455 | 379.91 | 390 | RECORDS | 595 | 204.98 |
| 341 | PEOPLE | 6116 | 369.04 | 391 | STORIES | 422 | 204.01 |
| 342 | NET | 653 | 367.19 | 392 | FEAR | 677 | 203.13 |
| 343 | BODY | 1656 | 366.56 | 393 | ACTUALLY | 1531 | 201.18 |
| 344 | INCREASED | 1034 | 364.50 | 394 | FIND | 2278 | 198.26 |
| 345 | RELATED | 770 | 357.94 | 395 | LACK | 718 | 193.49 |
| 346 | CAMPAIGN | 817 | 357.47 | 396 | SEEMS | 1285 | 191.44 |
| 347 | BEING | 4759 | 352.39 | 397 | TREATED | 549 | 189.60 |
| 348 | SUFFER | 421 | 348.78 | 398 | BENEFITS | 593 | 189.27 |
| 349 | STATES | 1291 | 342.27 | 399 | NONE | 664 | 187.86 |
| 350 | WONDER | 612 | 340.90 | 400 | SORRY | 759 | 185.44 |

| 401 | CHOICE | 817 | 185.01 | 451 | PREVIOUSLY | 478 | 116.03 |
|---|---|---|---|---|---|---|---|
| 402 | AMERICAN | 1038 | 184.31 | 452 | TRYING | 1036 | 115.85 |
| 403 | THANK | 808 | 184.12 | 453 | REVEALED | 391 | 113.72 |
| 404 | GETTING | 1217 | 182.25 | 454 | NATURAL | 850 | 113.57 |
| 405 | RECENT | 1014 | 181.77 | 455 | MONTH | 838 | 113.46 |
| 406 | AGREE | 605 | 180.17 | 456 | UNDERSTAND | 892 | 113.16 |
| 407 | INDIVIDUALS | 599 | 178.97 | 457 | FACT | 1919 | 111.01 |
| 408 | BORN | 613 | 177.80 | 458 | DO | 8098 | 110.46 |
| 409 | SURE | 1412 | 175.49 | 459 | JUST | 5908 | 110.08 |
| 410 | INTERESTED | 641 | 175.22 | 460 | OLDER | 572 | 107.93 |
| 411 | PRESS | 857 | 174.38 | 461 | SPREAD | 434 | 107.39 |
| 412 | CONTAINS | 397 | 172.68 | 462 | SUPPORT | 1632 | 106.40 |
| 413 | GIVING | 833 | 171.94 | 463 | SAY | 3289 | 105.57 |
| 414 | EXPLAIN | 576 | 170.01 | 464 | READING | 668 | 104.12 |
| 415 | C | 1771 | 167.17 | 465 | HUGE | 511 | 103.88 |
| 416 | CURRENTLY | 536 | 166.91 | 466 | IDENTIFIED | 435 | 103.80 |
| 417 | BASED | 1137 | 163.62 | 467 | LIST | 760 | 103.75 |
| 418 | RATE | 1158 | 163.03 | 468 | TOLD | 1847 | 103.16 |
| 419 | LETTER | 867 | 162.08 | 469 | CORRECT | 459 | 102.68 |
| 420 | WROTE | 680 | 161.93 | 470 | SIMILAR | 1040 | 102.15 |
| 421 | PROPERLY | 446 | 161.58 | 471 | POSSIBLE | 1762 | 101.80 |
| 422 | MOTHER | 1281 | 161.18 | 472 | POSITIVE | 544 | 100.73 |
| 423 | CONTACT | 728 | 159.82 | 473 | AVAILABLE | 1450 | 100.03 |
| 424 | QUESTION | 1485 | 159.82 | 474 | GROUP | 2041 | 99.79 |
| 425 | OCCURRED | 433 | 157.68 | 475 | RESPONSIBLE | 592 | 99.18 |
| 426 | CONDITIONS | 970 | 157.60 | 476 | DAILY | 498 | 98.10 |
| 427 | APPEARS | 557 | 155.07 | 477 | DAYS | 1678 | 96.11 |
| 428 | BECAUSE | 4915 | 154.70 | 478 | HAPPENED | 781 | 96.10 |
| 429 | ANSWER | 909 | 153.70 | 479 | COMMON | 1102 | 93.90 |
| 430 | NUMBERS | 748 | 153.32 | 480 | SIDE | 1674 | 92.97 |
| 431 | DEVELOP | 607 | 152.62 | 481 | COMMITTEE | 1023 | 92.37 |
| 432 | FINDING | 490 | 150.54 | 482 | CANNOT | 1171 | 89.81 |
| 433 | SERIOUSLY | 438 | 149.20 | 483 | NOTED | 413 | 89.05 |
| 434 | HUNDREDS | 341 | 149.14 | 484 | OTHER | 6969 | 88.37 |
| 435 | SIGNIFICANT | 785 | 147.97 | 485 | CONCERNED | 906 | 85.67 |
| 436 | CONCERN | 687 | 141.77 | 486 | FACTORS | 543 | 85.35 |
| 437 | HOPE | 1031 | 135.98 | 487 | HIGHLY | 560 | 85.18 |
| 438 | SAYING | 1072 | 135.77 | 488 | SCHOOL | 1835 | 83.68 |
| 439 | EXTREMELY | 490 | 135.73 | 489 | HERE | 3189 | 83.39 |
| 440 | PAIN | 507 | 133.14 | 490 | REMOVED | 389 | 82.40 |
| 441 | EVERYONE | 761 | 132.25 | 491 | THEY | 16394 | 81.63 |
| 442 | P | 1873 | 128.85 | 492 | POSSIBLY | 454 | 80.37 |
| 443 | DECISION | 998 | 128.02 | 493 | RELEASE | 428 | 79.71 |
| 444 | COMPLETELY | 571 | 126.85 | 494 | SEEM | 918 | 79.68 |
| 445 | SHOWED | 679 | 125.81 | 495 | SOMEONE | 939 | 79.60 |
| 446 | FAMILY | 1770 | 122.53 | 496 | APPARENTLY | 481 | 79.20 |
| 447 | WORSE | 491 | 122.37 | 497 | MAYBE | 602 | 78.57 |
| 448 | REASON | 1070 | 121.76 | 498 | PROPER | 416 | 78.52 |
| 449 | MENTIONED | 487 | 121.64 | 499 | SHOWING | 420 | 77.31 |
| 450 | LIKELY | 1288 | 116.31 | 500 | SOURCE | 551 | 77.09 |

| 501 | LATEST | 420 | 76.98 | 551 | SAME | 2800 | 38.44 |
|---|---|---|---|---|---|---|---|
| 502 | COMPANIES | 972 | 76.19 | 552 | SEVERAL | 1151 | 38.23 |
| 503 | AWARE | 619 | 74.55 | 553 | HEAR | 685 | 37.61 |
| 504 | PRODUCTS | 624 | 74.29 | 554 | CONSIDERED | 688 | 36.36 |
| 505 | SINCE | 2390 | 74.18 | 555 | HIGHER | 817 | 35.89 |
| 506 | PROGRAMME | 1020 | 73.46 | 556 | NEVER | 2447 | 35.84 |
| 507 | DESPITE | 805 | 73.07 | 557 | YET | 1606 | 35.48 |
| 508 | WHETHER | 1780 | 72.57 | 558 | LOW | 842 | 35.30 |
| 509 | PROBLEM | 1466 | 72.52 | 559 | DISCOVERED | 351 | 33.95 |
| 510 | EVEN | 4078 | 71.63 | 560 | WOMEN | 1562 | 33.83 |
| 511 | TYPE | 933 | 67.50 | 561 | NO | 9870 | 32.72 |
| 512 | INDEPENDENT | 622 | 66.91 | 562 | WE | 12831 | 32.22 |
| 513 | CONTINUE | 664 | 66.52 | 563 | EVERY | 1820 | 31.54 |
| 514 | CHECK | 439 | 64.26 | 564 | HEART | 673 | 31.27 |
| 515 | SHOWS | 657 | 64.20 | 565 | HISTORY | 933 | 31.27 |
| 516 | GET | 4412 | 64.08 | 566 | BELIEVED | 431 | 31.10 |
| 517 | OR | 16001 | 63.55 | 567 | NOVEMBER | 493 | 30.80 |
| 518 | SHOULD | 4793 | 63.41 | 568 | HIGH | 1778 | 30.55 |
| 519 | ADD | 489 | 63.07 | 569 | MILLION | 1172 | 30.53 |
| 520 | RESULT | 1140 | 62.42 | 570 | DIRECTOR | 612 | 30.34 |
| 521 | KNOWS | 482 | 61.73 | 571 | SPECIFIC | 581 | 30.25 |
| 522 | DONE | 1697 | 60.67 | 572 | ASK | 913 | 29.76 |
| 523 | ISSUES | 680 | 60.08 | 573 | WOULD | 9841 | 29.13 |
| 524 | TIMES | 1566 | 57.68 | 574 | COUNTRIES | 818 | 28.64 |
| 525 | HAPPEN | 498 | 57.08 | 575 | DOUBT | 607 | 28.50 |
| 526 | DAVID | 784 | 55.50 | 576 | REAL | 1085 | 27.47 |
| 527 | ASKING | 380 | 53.17 | 577 | ADDED | 741 | 27.24 |
| 528 | SEPTEMBER | 580 | 53.00 | 578 | BENEFIT | 551 | 26.10 |
| 529 | GIVE | 2101 | 51.84 | 579 | ANYTHING | 1290 | 26.04 |
| 530 | VERY | 5361 | 50.92 | 580 | CONSIDER | 586 | 25.65 |
| 531 | WITHOUT | 2133 | 50.45 | 581 | F | 489 | 25.29 |
| 532 | AMONG | 1134 | 49.30 | 582 | WATCH | 472 | 24.74 |
| 533 | CURRENT | 756 | 49.15 | | | | |
| 534 | POSSIBILITY | 412 | 48.91 | | | | |
| 535 | BOY | 644 | 48.84 | | | | |
| 536 | FIGURES | 618 | 48.29 | | | | |
| 537 | FOOD | 960 | 48.27 | | | | |
| 538 | EFFECT | 1160 | 46.69 | | | | |
| 539 | FAILED | 513 | 46.46 | | | | |
| 540 | USING | 1215 | 45.75 | | | | |
| 541 | KNOW | 5294 | 44.63 | | | | |
| 542 | NUMBER | 2289 | 43.93 | | | | |
| 543 | MAJORITY | 538 | 42.53 | | | | |
| 544 | AUTHORITIES | 679 | 41.35 | | | | |
| 545 | SUGGESTS | 385 | 41.12 | | | | |
| 546 | TODAY | 1138 | 40.80 | | | | |
| 547 | PRODUCT | 592 | 40.70 | | | | |
| 548 | EVER | 1297 | 40.43 | | | | |
| 549 | CLEARLY | 775 | 40.25 | | | | |
| 550 | LEAD | 743 | 39.21 | | | | |

**Appendix 8** Keywords in the *JABS* corpus with distribution across at least 10 % of corpus files, arranged according to semantic and functional categories and ordered according to LL

| | | |
|---|---|---|
| **vaccination and pharmaceutical products** | vaccines | *vaccine, vaccines, mmr* |
| | vaccination | *vaccination, vaccinations, vaccinated, immunisation, vaccinate, immunization, vaccinating* |
| | vaccination: pre-modifiers | *fully (vaccinated)* |
| | vaccine types | *single, combined, multiple* |
| | administering vaccines | *jabs,*[76] *jab, dose, doses, injection, shots, shot* |
| | other pharmaceutical products | *drugs, medicines, products, product* |
| | vaccine additives | *mercury, thimerosal* |
| | toxicity | *toxic* |
| **health, disease and medical conditions** | health and disease: general terms | *health, disease, diseases, infection, infections, illness, illnesses* |
| | health and disease: adjectives | *healthy, ill, sick* |
| | diseases and infections | *measles, virus, mumps, flu, cancer, polio, rubella, viruses, tetanus, meningitis* |
| | diseases and infections: non-evaluative pre-modifiers | *chronic, viral, infectious, genetic* |
| | disease types | *type* |
| | immune system and nervous system | *immune, immunity, (immune/nervous) system* |
| | immune system processes | *response* |
| | medical conditions: general terms | *disorders, syndrome, disorder, condition, conditions* |
| | specific conditions | *autism, autistic (spectrum), spectrum* |
| | conditions: misc. pre-modifiers | *neurological* |
| | diseases and conditions: verbs | *suffering (from), suffer* |
| | parts of the body | *brain, bowel, blood, cells, cell, body, heart* |
| | symptoms and signs | *symptoms, fever, pain* |
| | diet | *food* |
| **people, places and institutions** | nhs users/laypeople | *children, parents, child, son, infants, babies, daughter, kids, baby, parent, girls, patients, adults, families, patient, (the) public, population, people, individuals, mother, family, group, boy, women* |
| | pre-modifiers referring to people | *infant, human, older* |
| | healthcare practitioners and scientific and medical experts | *doctors, gp, doctor, researchers, scientists, experts, expert, officials* |
| | miscellaneous | *author* |
| | named publicly known individuals | *(dr andrew) wakefield, andrew (wakefield), david (kirby/ southall/ salisbury,/ thrower/ elliman)* |
| | JABS forum posters | *aasa,(john) stone, john (stone)* |
| | job titles and honorifics | *dr, professor, director* |
| | public institutions: | *center, centers, department, institute, agency,* |

[76] Roughly 50% of citations of *jabs* refer to injections; the other 50% refer to the JABS group.

| | | |
|---|---|---|
| | general | *government, authorities* |
| | public institutions: specific | *nhs* |
| | industries | *drug (companies), pharma, pharmaceutical (companies) manufacturers, companies* |
| | private companies | *merck* |
| | places | *hospital, university, school* |
| | geographical references | *usa, countries* |
| | nationalities | *american* |
| **science, medicine and healthcare** | general medical, healthcare and scientific terms | *medical, medicine, diagnosed, clinical, scientific, science, diagnosis, treatment, alternative, care, treat, treated, (case) history* |
| | terms related to administration of vaccination programme | *program, schedule, introduced, programme* |
| | references to age and time periods | *childhood, months, aged, age, years, birth, weeks, year, february, born, month, daily, days, september, november, today, times* |
| | references to epidemiological research and monitoring | *cases, deaths, epidemic, incidence, case, rates, levels, rate, spread* |
| | other technical terms | *factors* |
| | measurement | *percent* |
| | research processes | *conducted* |
| | miscellaneous | *et, al* |
| | tests and investigations | *trials, tested, testing, tests, trial, investigation confirmed, test, check* |
| **sources** | sources of information: general terms | *news, journal, media, press, source* |
| | scientific sources | *journal, reports, paper, review, report* |
| | reference to scientific research | *study, research, studies, findings, data, paper, results, figures* |
| | news sources and websites | *jabs, cdc, fda, whale, bbc* |
| | sources/channels of communication | *website, site, online* |
| | written communication | *letter, list* |
| **the concerns and business of JABS** | information, advice and requests for advice | *info, information, content, advice, help, contact, question, answer, support* |
| | political action | *campaign* |
| | legal issues | *hearing, legal* |
| **likelihood, causation, and temporal relations** | possibility | *risk, risks, possibility* |
| | cause and effect | *cause, link, caused , causes, linked, causing, due (to), links, affected, association, connection, because, lead (to), result (in)* |
| | adverbs and prepositions: temporal | *following, after, prior, recently, recent, currently, previously, since, current* |
| | reason | *reason* |
| | ergative verbs indicating change | *develop, developed, developing, , increased, increase, develop* |
| | effects, danger and harm | *adverse (events/effects/reactions), reactions, damage, effects, (vaccine) damaged, death, complications, harm, dangerous, injury, (adverse) events, side (effects), effect* |
| | sources of harm | *exposure (to mercury/thimerosal), exposed (to)* |

| | | |
|---|---|---|
| **expressions with modal meaning** | epistemic modality | *'ll, potentially, can, potential, likely, possible, cannot, possibly, maybe, likely* |
| | necessity/desirability | *recommended, should* |
| **conditionality** | | *wouldn', if, whether, would* |
| **expressions of evaluation** | terms used to express evaluation of vaccines | *safety, safe, protect, protection, prevention, prevent, effective, benefits* |
| | terms used to express evaluation of science/authority | *lack (of evidence/ information/ understanding/ knowledge), failed* |
| | evaluative terms: disease and reactions | *serious, severe, rare* |
| | evaluation of quality of information | *accurate, latest* |
| | evaluation: clarity | *apparently, clearly* |
| | other evaluative terms | *wrong, interesting, bad, unfortunately, sure, interested, normal, properly, seriously, significant, worse, correct , concerned, proper, higher, low, high, real, positive, responsible, benefit* |
| | hedging and boosting | *actually, extremely, completely, just, huge, highly, very, even, ever* |
| | status nouns: the 'sign' group | *evidence, proof* |
| | status nouns: factivity | *truth, problems, facts, issue, opinion, fact, problem, issues* |
| | status nouns: affect | *concerns, fear, concern, hope* |
| | comparison/contrast | *compared, similar, other, or, same* |
| | neutrally connotative terms used evaluatively in some contexts | *natural* |
| **other nouns, verbs, and phrases referring to events, actions, and states of affairs** | verbal processes | *reported, comments, reporting, quote, comment, claims, stated, according (to), states, mention, debate, suggest, statement, conclusion, claim, concluded, agree, explain, saying, mentioned, say, told, asking, suggests, ask* |
| | recounting | *story, stories* |
| | written communication | *published, read, wrote, reading* |
| | making decisions | *informed (consent/choice/decision),(informed) consent, choice, decision* |
| | perception and cognition | *believe, wonder, known, note, trust, understand, know, noted, knows, (BE) aware, considered, hear, consider, doubt, watch, believed* |
| | seeming | *seems, appears, seem* |
| | showing | *prove, showed, revealed, showing, shows* |
| | discovering | *found, find, finding, identified, discovered* |
| | occurring | *occur, occurred, happened, happen* |
| | death | *died, kills, die* |
| | durative | *continue* |
| | adding | *add, added* |
| | other material processes | *try, removed, release, using* |
| | existential/relational processes or states | *live, lives, contain, contains* |
| **other grammatical items** | negative polarity | *'t, don', doesn' didn', isn', don't, not, wasn', haven', none, never, no* |
| | *that* | *that* |
| | abbreviations | *re* |
| | determiners and | *my, anyone, i, this, your, any, many, our, these, us,* |

| | | |
|---|---|---|
| | pronouns | *everyone, they, someone, we, several, anything, several, every* |
| | spatial deixis | *here* |
| | prepositions and adverbs | *regarding, about, against, also, by, ago, including, from, despite, without, among, yet* |
| | WH words | *who, why, how,* |
| | delexicalised verbs/ auxiliaries / cupola | *'ve, have, am, is, are, has, does, being, having, been, do, done* |
| | numerals | *million* |
| | quantities and quantifiers | *thousands, numbers, millions, hundreds, number, majority* |
| **miscellaneous items** | greetings and politeness markers | *hi, thanks, please, sorry, thank* |
| | miscellaneous nouns, verbs, and abbreviations | *re, etc, smoking* |
| | miscellaneous pre-and post-modifiers | *related, free, based, available, common, independent, specific* |
| | electronic communication | *http, www, com, posted, html, org, uk, edited, originally (posted by), gov, br, htm, id, pdf, forum, php, cgi, post, net, page,* |
| | abbreviations | *m, u, d, b, c, p* |

**Appendix 9** Fifty most frequent three to six-word clusters from concordances of *mmr* in *JABS* and *NHSvax* corpora, with frequencies

| | ***JABS*** | | ***NHSvax*** | |
|---|---|---|---|---|
| **Rank** | **Cluster** | **Freq.** | **Cluster** | **Freq.** |
| 1 | THE MMR VACCINE | 1005 | THE MMR VACCINE | 505 |
| 2 | OF THE MMR | 484 | MMR VACCINE AND | 181 |
| 3 | THE MMR JAB | 404 | MMR AND AUTISM | 162 |
| 4 | THAT THE MMR | 238 | BETWEEN MMR AND | 149 |
| 5 | MMR AND AUTISM | 200 | LINK BETWEEN MMR | 146 |
| 6 | MMR VACCINE AND | 192 | OF MMR VACCINE | 130 |
| 7 | TO THE MMR | 180 | OF THE MMR | 123 |
| 8 | LINK BETWEEN MMR | 178 | BETWEEN MMR AND AUTISM | 114 |
| 9 | HAD THE MMR | 177 | LINK BETWEEN MMR AND | 104 |
| 10 | OF THE MMR VACCINE | 174 | LINK BETWEEN MMR AND AUTISM | 91 |
| 11 | THE MMR AND | 168 | DOSE OF MMR | 85 |
| 12 | BETWEEN MMR AND | 162 | MMR VACCINE WAS | 78 |
| 13 | WITH THE MMR | 156 | MMR VACCINE AND AUTISM | 77 |
| 14 | THE MMR VACCINATION | 151 | VACCINE AND AUTISM | 77 |
| 15 | THE COMBINED MMR | 144 | OF THE MMR VACCINE | 76 |
| 16 | IN THE MMR | 139 | MUMPS AND RUBELLA | 70 |
| 17 | BY THE MMR | 135 | BETWEEN MMR VACCINE | 70 |
| 18 | BETWEEN MMR AND AUTISM | 125 | DOSES OF MMR | 68 |
| 19 | LINK BETWEEN MMR AND | 122 | MEASLES MUMPS AND | 67 |
| 20 | MMR VACCINE IS | 122 | A LINK BETWEEN | 67 |
| 21 | MMR RIP EDITED | 122 | ASSOCIATION BETWEEN MMR | 67 |
| 22 | EDITED BY GUS | 121 | MEASLES MUMPS AND RUBELLA | 64 |
| 23 | BY GUS THE | 120 | BETWEEN MMR VACCINE AND | 62 |
| 24 | MMR RIP EDITED BY GUS | 120 | MMR VACCINE IS | 61 |
| 25 | RIP EDITED BY | 120 | MMR VACCINE THE | 60 |
| 26 | MMR RIP EDITED BY GUS THE | 120 | TWO DOSES OF | 59 |
| 27 | RIP EDITED BY GUS THE | 120 | THE MMR VACCINE AND | 56 |
| 28 | RIP EDITED BY GUS | 120 | THAT MMR IS | 52 |
| 29 | MMR RIP EDITED BY | 120 | TWO DOSES OF MMR | 52 |
| 30 | EDITED BY GUS THE | 120 | A LINK BETWEEN MMR | 52 |
| 31 | THE SAFETY OF | 111 | VACCINE WAS INTRODUCED | 50 |
| 32 | MEASLES MUMPS AND | 109 | MMR VACCINE WAS INTRODUCED | 50 |
| 33 | MUMPS AND RUBELLA | 109 | MMR VACCINATION AND | 49 |
| 34 | OF MMR VACCINE | 102 | THE MMR VACCINE WAS | 48 |
| 35 | FOR THE MMR | 102 | SAFETY OF MMR | 47 |
| 36 | MEASLES MUMPS AND RUBELLA | 102 | MMR THE FACTS | 46 |
| 37 | THAT THE MMR VACCINE | 101 | THE SAFETY OF | 42 |
| 38 | LINK BETWEEN MMR AND AUTISM | 100 | SECOND DOSE OF | 42 |
| 39 | HAVE THE MMR | 100 | A LINK BETWEEN MMR AND | 41 |
| 40 | THAT MMR IS | 99 | TO THE MMR | 41 |
| 41 | ABOUT THE MMR | 99 | MMR VACCINE HAS | 41 |
| 42 | AFTER THE MMR | 98 | MUMPS AND RUBELLA MMR | 39 |
| 43 | THE MMR VACCINE AND | 95 | AND RUBELLA MMR | 39 |

| | | | | |
|---|---|---|---|---|
| 44 | PAGE 1 OF | 94 | MEASLES MUMPS AND RUBELLA MMR | 39 |
| 45 | GIVEN THE MMR | 92 | THE MMR VACCINATION | 39 |
| 46 | HAVING THE MMR | 92 | THE MMR VACCINE THE | 38 |
| 47 | THE MMR IS | 87 | THE MMR VACCINE WAS INTRODUCED | 38 |
| 48 | ON THE MMR | 87 | A LINK BETWEEN MMR AND AUTISM | 38 |
| 49 | AND THE MMR | 84 | AN ASSOCIATION BETWEEN MMR | 37 |
| 50 | A LINK BETWEEN | 83 | SECOND DOSE OF MMR | 37 |

**Appendix 10** Fifty most frequent three to six-word clusters from concordances of *autism* in *JABS* and *NHSvax* corpora, with frequencies

| | ***JABS*** | | ***NHSvax*** | |
|---|---|---|---|---|
| **Rank** | **Cluster** | **Freq.** | **Cluster** | **Freq.** |
| 1 | CHILDREN WITH AUTISM | 556 | MMR AND AUTISM | 161 |
| 2 | MMR AND AUTISM | 202 | LINK BETWEEN MMR | 116 |
| 3 | VACCINES AND AUTISM | 183 | BETWEEN MMR AND | 114 |
| 4 | WITH AUTISM AND | 173 | BETWEEN MMR AND AUTISM | 109 |
| 5 | OF CHILDREN WITH | 162 | LINK BETWEEN MMR AND | 91 |
| 6 | AUTISM SPECTRUM DISORDERS | 157 | LINK BETWEEN MMR AND AUTISM | 90 |
| 7 | OF AUTISM IN | 148 | VACCINE AND AUTISM | 88 |
| 8 | PAGE 1 OF | 146 | MMR VACCINE AND | 80 |
| 9 | LINK BETWEEN MMR | 145 | MMR VACCINE AND AUTISM | 76 |
| 10 | OF AUTISM AND | 140 | CHILDREN WITH AUTISM | 67 |
| 11 | OF CHILDREN WITH AUTISM | 137 | OF AUTISM IN | 59 |
| 12 | DIAGNOSED WITH AUTISM | 135 | THE MMR VACCINE | 58 |
| 13 | BETWEEN MMR AND | 131 | A LINK BETWEEN | 46 |
| 14 | AUTISM AND BOWEL | 130 | THE INCIDENCE OF | 42 |
| 15 | AGE OF AUTISM | 128 | INFLAMMATORY BOWEL DISEASE | 42 |
| 16 | BETWEEN AUTISM AND | 127 | PREVALENCE OF AUTISM | 41 |
| 17 | AUTISM IN THE | 127 | INCIDENCE OF AUTISM | 41 |
| 18 | CAUSE OF AUTISM | 126 | THE INCIDENCE OF AUTISM | 40 |
| 19 | BETWEEN MMR AND AUTISM | 126 | A LINK BETWEEN MMR | 39 |
| 20 | IN CHILDREN WITH | 121 | INCREASE IN AUTISM | 38 |
| 21 | AUTISM SPECTRUM DISORDER | 116 | A LINK BETWEEN MMR AND | 38 |
| 22 | VACCINE AND AUTISM | 116 | A LINK BETWEEN MMR AND AUTISM | 38 |
| 23 | PAGE 1 OF 1 | 111 | OF AUTISM AND | 37 |
| 24 | 1 OF 1 | 111 | RISK OF AUTISM | 36 |
| 25 | OF THE AUTISM | 109 | AND AUTISM AND | 34 |
| 26 | CASES OF AUTISM | 105 | CAUSES OF AUTISM | 34 |
| 27 | LINK BETWEEN MMR AND | 105 | BETWEEN MMR VACCINE | 33 |
| 28 | LINK BETWEEN MMR AND AUTISM | 101 | BETWEEN MMR VACCINE AND AUTISM | 33 |
| 29 | IN CHILDREN WITH AUTISM | 101 | MUMPS AND RUBELLA | 33 |
| 30 | BETWEEN VACCINES AND | 98 | BETWEEN MMR VACCINE AND | 33 |
| 31 | BETWEEN VACCINES AND AUTISM | 95 | CASES OF AUTISM | 32 |
| 32 | WITH REGRESSIVE AUTISM | 92 | THE MMR VACCINE AND | 30 |
| 33 | THE MMR VACCINE | 92 | THE MMR VACCINE AND AUTISM | 30 |
| 34 | AUTISM AND THE | 91 | VACCINATION AND AUTISM | 30 |
| 35 | THE AUTISM EPIDEMIC | 91 | AUTISM SPECTRUM DISORDERS | 29 |
| 36 | A LINK BETWEEN | 90 | THE PREVALENCE OF AUTISM | 28 |
| 37 | INCIDENCE OF AUTISM | 84 | AND AUTISM THE | 28 |
| 38 | AND BOWEL DISEASE | 83 | THE PREVALENCE OF | 28 |
| 39 | NOT CAUSE AUTISM | 79 | BETWEEN THE MMR | 27 |
| 40 | DIAGNOSIS OF AUTISM | 78 | DIAGNOSIS OF AUTISM | 26 |
| 41 | INCREASE IN AUTISM | 77 | BETWEEN THE MMR VACCINE | 26 |

| | | | | |
|---|---|---|---|---|
| 42 | LINK BETWEEN AUTISM | 77 | BETWEEN THE MMR VACCINE AND | 26 |
| 43 | AUTISM IS A | 77 | BETWEEN THE MMR VACCINE AND AUTISM | 26 |
| 44 | AUTISM AND BOWEL DISEASE | 76 | BOWEL DISEASE AND | 25 |
| 45 | THAT AUTISM IS | 74 | IN CHILDREN WITH | 25 |
| 46 | BOWEL DISEASE AND | 73 | AUTISM AND THE | 25 |
| 47 | AND AUTISM AND | 72 | NO LINK BETWEEN | 24 |
| 48 | CHILD WITH AUTISM | 71 | VACCINES AND AUTISM | 24 |
| 49 | AUTISM PAGE 1 OF | 71 | OF AUTISM THE | 23 |
| 50 | AUTISM PAGE 1 | 71 | ASSOCIATION BETWEEN MMR | 23 |

**Appendix 11** The top 100 collocates of *risk* and *risks* in the *JABS* corpus, according to MI3

| Rank | *risk* | MI3 | *risks* | MI3 |
|---|---|---|---|---|
| 1 | RISK | 33.69 | RISKS | 31.85 |
| 2 | OF | 25.87 | OUTWEIGH | 24.43 |
| 3 | THE | 25.21 | BENEFITS | 24.26 |
| 4 | AT | 24.89 | THE | 23.54 |
| 5 | A | 23.24 | OF | 21.64 |
| 6 | INCREASED | 23.10 | AND | 19.91 |
| 7 | TO | 22.39 | ASSOCIATED | 19.08 |
| 8 | BENEFIT | 22.22 | ARE | 18.93 |
| 9 | IS | 21.79 | TO | 18.92 |
| 10 | AND | 21.71 | BALANCE | 17.90 |
| 11 | FOR | 21.34 | OUTWEIGHED | 17.85 |
| 12 | FACTORS | 21.11 | WITH | 17.81 |
| 13 | DEVELOPING | 21.04 | VACCINE | 17.42 |
| 14 | INCREASE | 20.75 | ABOUT | 17.30 |
| 15 | ARE | 20.26 | FAR | 17.12 |
| 16 | CHILDREN | 20.18 | REAL | 17.10 |
| 17 | CANCER | 20.14 | VACCINATION | 16.77 |
| 18 | HIGH | 20.09 | HEALTH | 16.70 |
| 19 | IN | 20.01 | ANY | 16.62 |
| 20 | PUT | 19.98 | ACKNOWLEGE | 16.59 |
| 21 | PUTTING | 19.97 | THAT | 16.57 |
| 22 | THAT | 19.90 | OVERWHELMINGLY | 16.36 |
| 23 | FROM | 19.88 | VACCINATING | 16.28 |
| 24 | GREATER | 19.70 | OUTWEIGHT | 16.21 |
| 25 | HIGHER | 19.47 | POSES | 16.16 |
| 26 | NOT | 19.38 | POSE | 16.09 |
| 27 | VACCINE | 19.37 | VACCINES | 15.99 |
| 28 | CHILD | 19.34 | INVOLVED | 15.83 |
| 29 | DISEASE | 19.29 | NOT | 15.79 |
| 30 | WITH | 19.29 | GREATER | 15.74 |
| 31 | BE | 19.23 | WEIGH | 15.72 |
| 32 | SERIOUS | 19.16 | SERIOUS | 15.72 |
| 33 | RELATIVE | 19.11 | PROPERLY | 15.66 |
| 34 | FACTOR | 19.10 | WEIGHING | 15.62 |
| 35 | OUTWEIGHS | 18.88 | POTENTIAL | 15.40 |
| 36 | OR | 18.71 | OUTWAY | 15.38 |
| 37 | CATCHING | 18.59 | FAVOURABLE | 15.32 |
| 38 | THERE | 18.48 | BUT | 15.26 |
| 39 | REDUCE | 18.35 | DISEASES | 15.23 |
| 40 | AN | 17.92 | MMR | 15.19 |
| 41 | CERVICAL | 17.83 | FROM | 15.17 |
| 42 | TAKE | 17.79 | POSSIBLE | 14.99 |
| 43 | COMPLICATIONS | 17.70 | WRAPPING | 14.96 |
| 44 | PRECANCEROUS | 17.67 | INFORMATION | 14.60 |
| 45 | MEASLES | 17.67 | FOR | 14.55 |
| 46 | AUTISM | 17.65 | BE | 14.52 |
| 47 | THAN | 17.57 | CARRY | 14.48 |
| 48 | MAY | 17.55 | CHILDREN | 14.43 |
| 49 | IT | 17.48 | BENFITS | 14.38 |
| 50 | NO | 17.42 | ITS | 14.35 |
| 51 | WORTH | 17.41 | OUTWEIGHS | 14.32 |
| 52 | ASSOCIATED | 17.40 | INCUR | 14.30 |

| | | | | |
|---|---|---|---|---|
| 53 | INFECTION | 17.30 | SINVESTIGATE | 14.21 |
| 54 | THEIR | 17.29 | OVERSHADOW | 14.21 |
| 55 | HEART | 17.24 | UNDERPLAYING | 14.21 |
| 56 | DEATH | 17.19 | SSMALL | 14.21 |
| 57 | AS | 17.16 | REALÂ | 14.21 |
| 58 | RATIO | 17.16 | THERE | 14.18 |
| 59 | BY | 17.15 | APPRECIABLE | 14.16 |
| 60 | ANY | 17.14 | EVALUATE | 14.09 |
| 61 | WAS | 17.08 | PROCEDURE | 14.06 |
| 62 | ASTHMA | 17.05 | IS | 14.03 |
| 63 | POSE | 17.02 | MINIMISING | 13.96 |
| 64 | INCREASES | 16.92 | PLAYED | 13.95 |
| 65 | CARRIES | 16.84 | EXCEED | 13.85 |
| 66 | RUN | 16.83 | INFORMED | 13.81 |
| 67 | ANALYST | 16.78 | IN | 13.81 |
| 68 | VACCINES | 16.78 | WERE | 13.78 |
| 69 | MORE | 16.78 | KNOW | 13.75 |
| 70 | CONTRACTING | 16.74 | CARRIES | 13.70 |
| 71 | ASSESSMENT | 16.72 | AWARE | 13.68 |
| 72 | BUT | 16.71 | SIGNIFICANT | 13.68 |
| 73 | THOSE | 16.69 | CLIFFORD | 13.64 |
| 74 | SMALL | 16.66 | UNVERIFIED | 13.62 |
| 75 | REDUCES | 16.55 | A | 13.53 |
| 76 | S | 16.50 | TAKE | 13.52 |
| 77 | VACCINATION | 16.46 | OR | 13.49 |
| 78 | THEY | 16.45 | SMALL | 13.33 |
| 79 | RON | 16.44 | KNOWN | 13.30 |
| 80 | COULD | 16.42 | IMPAIR | 13.30 |
| 81 | IF | 16.38 | AS | 13.28 |
| 82 | ADVERSE | 16.37 | HAVE | 13.26 |
| 83 | BECAUSE | 16.35 | BY | 13.24 |
| 84 | LESIONS | 16.27 | QUANTIFIABLE | 13.21 |
| 85 | GETTING | 16.27 | DISEASE | 13.19 |
| 86 | YOU | 16.24 | FETUSES | 13.16 |
| 87 | MIGHT | 16.23 | MAY | 13.16 |
| 88 | VACCINATING | 16.19 | REFUSES | 13.15 |
| 89 | REACTION | 16.03 | CHILD | 13.13 |
| 90 | LOW | 16.03 | ACCOMPANY | 13.06 |
| 91 | CARRY | 16.01 | ON | 13.03 |
| 92 | WOULD | 15.99 | NO | 13.03 |
| 93 | HEALTH | 15.96 | EFFECTS | 13.01 |
| 94 | HAVE | 15.90 | CONTRACTING | 12.96 |
| 95 | DYING | 15.86 | NTLWORLD | 12.92 |
| 96 | I | 15.83 | TBH | 12.89 |
| 97 | LIVES | 15.81 | STARKLY | 12.89 |
| 98 | THIS | 15.79 | EXAGGERATE | 12.89 |
| 99 | WHICH | 15.73 | ILLUSTRATED | 12.88 |
| 100 | FAR | 15.72 | WITHOUT | 12.86 |

**Appendix 12** All meaningful two- to six-word clusters with *risk* and *risks* , with a frequency of 7 occurrences or more in the *JABS* corpus

| Rank | *risk* | freq | *risks* | freq |
|---|---|---|---|---|
| 1 | THE RISK | 892 | THE RISKS | 391 |
| 2 | AT RISK | 497 | OUTWEIGH THE RISKS | 46 |
| 3 | A RISK | 211 | HEALTH RISKS | 42 |
| 4 | HIGH RISK | 177 | RISKS AND BENEFITS | 37 |
| 5 | INCREASED RISK | 161 | REAL RISKS | 27 |
| 6 | RISK FACTORS | 98 | ABOUT THE RISKS | 23 |
| 7 | RISK BENEFIT | 91 | THE REAL RISKS | 21 |
| 8 | RISK OF DEVELOPING | 75 | VACCINE RISKS | 18 |
| 9 | INCREASE THE RISK | 69 | POTENTIAL RISKS | 16 |
| 10 | HIGHER RISK | 64 | THE RISKS AND BENEFITS | 16 |
| 11 | RISK FACTOR | 57 | POSSIBLE RISKS | 14 |
| 12 | AN INCREASED RISK | 55 | THE BALANCE OF RISKS | 13 |
| 13 | RISK OF AUTISM | 55 | SERIOUS RISKS | 12 |
| 14 | BE AT RISK | 51 | BENEFITS OUTWEIGH THE RISKS | 12 |
| 15 | GREATER RISK | 43 | RISKS OF VACCINATION | 12 |
| 16 | RELATIVE RISK | 43 | RISKS OF MMR | 11 |
| 17 | CHILDREN AT RISK | 41 | WHAT ARE THE RISKS | 11 |
| 18 | AT HIGH RISK | 40 | THE BENEFITS OUTWEIGH THE RISKS | 10 |
| 19 | TAKE THE RISK | 36 | RISKS OF VACCINES | 10 |
| 20 | ARE AT RISK | 34 | THERE ARE RISKS | 10 |
| 21 | WORTH THE RISK | 34 | THE POTENTIAL RISKS | 10 |
| 22 | REDUCE THE RISK | 30 | BENEFITS AND RISKS | 10 |
| 23 | A RISK FACTOR | 29 | THESE RISKS | 9 |
| 24 | PUT AT RISK | 28 | KNOWN RISKS | 9 |
| 25 | RISK OF DEATH | 28 | THE HEALTH RISKS | 9 |
| 26 | CANCER RISK | 27 | THE RISKS OF MMR | 9 |
| 27 | NO RISK | 27 | FAR OUTWEIGH THE RISKS | 9 |
| 28 | AT INCREASED RISK | 26 | THE RISKS INVOLVED | 8 |
| 29 | RISK OF INFECTION | 26 | THE RISKS OF VACCINES | 8 |
| 30 | RISK OF CANCER | 26 | KNOW THE RISKS | 8 |
| 31 | RISK GROUPS | 26 | INCREASED RISKS | 8 |
| 32 | THE RISK OF AUTISM | 25 | THE TERRIBLE RISKS INVOLVED | 7 |
| 33 | THE RISK OF DEVELOPING | 25 | RELATIVE RISKS | 7 |
| 34 | RUN THE RISK | 24 | - | - |
| 35 | A HIGH RISK | 24 | - | - |
| 36 | VACCINE RISK | 24 | - | - |
| 37 | CHILD AT RISK | 22 | - | - |
| 38 | MORE AT RISK | 22 | - | - |
| 39 | SERIOUS RISK | 21 | - | - |
| 40 | DISEASE RISK | 21 | - | - |
| 41 | REAL RISK | 21 | - | - |
| 42 | MORE RISK | 21 | - | - |
| 43 | THEIR RISK | 21 | - | - |
| 44 | HEALTH RISK | 21 | - | - |
| 45 | RISK OF CATCHING | 20 | - | - |
| 46 | RISK BENEFIT RATIO | 20 | - | - |

| | | | | |
|---|---|---|---|---|
| 47 | SMALL RISK | 20 | - | - |
| 48 | A GREATER RISK | 19 | - | - |
| 49 | THERE IS A RISK | 18 | - | - |
| 50 | TO TAKE THE RISK | 18 | - | - |
| 51 | THE RELATIVE RISK | 18 | - | - |
| 52 | A HIGHER RISK | 18 | - | - |
| 53 | INCREASES THE RISK | 18 | - | - |
| 54 | RISK OF COMPLICATIONS | 18 | - | - |
| 55 | AT HIGHER RISK | 18 | - | - |
| 56 | RISK OF CONTRACTING | 18 | - | - |
| 57 | LOW RISK | 18 | - | - |
| 58 | NOT WORTH THE RISK | 17 | - | - |
| 59 | THE RISK BENEFIT RATIO | 17 | - | - |
| 60 | THE RISK OF DEATH | 16 | - | - |
| 61 | RISK ASSESSMENT | 16 | - | - |
| 62 | RISK OF MEASLES | 15 | - | - |
| 63 | PEOPLE AT RISK | 15 | - | - |
| 64 | HIGHEST RISK | 15 | - | - |
| 65 | SIGNIFICANT RISK | 15 | - | - |
| 66 | AT RISK OF CATCHING | 14 | - | - |
| 67 | THE HIGHEST RISK | 14 | - | - |
| 68 | RISK BENEFIT ANALYSIS | 14 | - | - |
| 69 | AT RISK REGISTER | 14 | - | - |
| 70 | ZERO RISK | 14 | - | - |
| 71 | POTENTIAL RISK | 14 | - | - |
| 72 | TO REDUCE THE RISK | 13 | - | - |
| 73 | HIGH RISK GROUPS | 13 | - | - |
| 74 | INCREASED RISK OF CANCER | 12 | - | - |
| 75 | HIGH RISK HPV | 12 | - | - |
| 76 | RISK OF ASTHMA | 12 | - | - |
| 77 | THE HIGH RISK | 12 | - | - |
| 78 | A SMALL RISK | 12 | - | - |
| 79 | REDUCES THE RISK | 12 | - | - |
| 80 | ABSOLUTE RISK | 12 | - | - |
| 81 | THIS RISK | 12 | - | - |
| 82 | CHILDREN AT HIGH RISK | 11 | - | - |
| 83 | BECAUSE OF THE RISK | 11 | - | - |
| 84 | WHO ARE AT RISK | 11 | - | - |
| 85 | MOST AT RISK | 11 | - | - |
| 86 | AT GREATER RISK | 11 | - | - |
| 87 | INCREASED THE RISK | 11 | - | - |
| 88 | WITH AN INCREASED RISK | 10 | - | - |
| 89 | THE RISK OF INFECTION | 10 | - | - |
| 90 | LEVEL OF RISK | 10 | - | - |
| 91 | NO INCREASED RISK | 10 | - | - |
| 92 | RISK OF DYING | 10 | - | - |
| 93 | RISK POLICY ANALYST | 10 | - | - |
| 94 | NOT AT RISK | 10 | - | - |
| 95 | OUTWEIGH THE RISK | 10 | - | - |
| 96 | MUCH HIGHER RISK | 10 | - | - |
| 97 | RISK POLICY | 10 | - | - |
| 98 | RISK MANAGEMENT | 10 | - | - |
| 99 | THE RISK FACTORS | 9 | - | - |
| 100 | LOWER RISK | 9 | - | - |
| 101 | RISK TO CHILDREN | 9 | - | - |

| | | | | |
|---|---|---|---|---|
| 102 | AUTISM RISK | 9 | - | - |
| 103 | TO BE AT RISK | 9 | - | - |
| 104 | INCREASE IN THE RISK | 9 | - | - |
| 105 | RISK CONDITIONS | 9 | - | - |
| 106 | BRAVE NEW WORLD OF ZERO RISK | 9 | - | - |
| 107 | ASSOCIATED WITH AN INCREASED RISK | 9 | - | - |
| 108 | THE RISK OF MEASLES | 9 | - | - |
| 109 | A RELATIVE RISK | 9 | - | - |
| 110 | REDUCED RISK | 9 | - | - |
| 111 | COULD INCREASE THE RISK | 9 | - | - |
| 112 | RISK FOR AUTISM | 9 | - | - |
| 113 | FAR OUTWEIGHS THE RISK | 9 | - | - |
| 114 | COULD BE AT RISK | 9 | - | - |
| 115 | LESS RISK | 8 | - | - |
| 116 | RISK BASED | 8 | - | - |
| 117 | THE RISK OF CONTRACTING | 8 | - | - |
| 118 | BEING PUT AT RISK | 8 | - | - |
| 119 | THE VACCINE RISK | 8 | - | - |
| 120 | PUT THEM AT RISK | 8 | - | - |
| 121 | TAKE A RISK | 8 | - | - |
| 122 | AT RISK GROUPS | 8 | - | - |
| 123 | AT MORE RISK | 8 | - | - |
| 124 | RISK AREAS | 8 | - | - |
| 125 | BE A RISK | 8 | - | - |
| 126 | THE RISK OF CANCER | 8 | - | - |
| 127 | MIGHT BE AT RISK | 8 | - | - |
| 128 | RISK OF BIAS | 8 | - | - |
| 129 | A SIGNIFICANT RISK | 8 | - | - |
| 130 | INCREASED RISK OF AUTISM | 8 | - | - |
| 131 | POSSIBLE RISK | 8 | - | - |
| 132 | HIGH RISK CHILDREN | 8 | - | - |
| 133 | ON THE AT RISK REGISTER | 8 | - | - |
| 134 | INCREASING THE RISK | 8 | - | - |
| 135 | RISK OF DISEASE | 8 | - | - |
| 136 | OTHER CHILDREN AT RISK | 8 | - | - |
| 137 | A MUCH HIGHER RISK | 8 | - | - |
| 138 | RISK FOR INFLUENZA | 8 | - | - |
| 139 | OTHER RISK FACTORS | 8 | - | - |
| 140 | WITH A REDUCED RISK | 7 | - | - |
| 141 | AT A GREATER RISK | 7 | - | - |
| 142 | AS RISK FACTORS | 7 | - | - |
| 143 | OBVIOUSLY AT RISK | 7 | - | - |
| 144 | FOR RISK ASSESSMENT | 7 | - | - |
| 145 | RISK IS OVERESTIMATED | 7 | - | - |
| 146 | PUT AT INCREASED RISK | 7 | - | - |
| 147 | WITHOUT RISK | 7 | - | - |
| 148 | DISEASE RISK IS OVERESTIMATED | 7 | - | - |
| 149 | ABOUT RISK | 7 | - | - |
| 150 | THE ABSOLUTE RISK | 7 | - | - |

**Appendix 13** Status nouns and reporting verbs and nouns in the *JABS* and *NHSvax* corpora, which collocate with *that* with a z-score of at least 4.0

| rank | ***JABS*** | | ***NHSvax*** | |
|---|---|---|---|---|
| | **collocate** | **z-score** | **collocate** | **z-score** |
| 1 | FACT | 44.83 | EVIDENCE | 28.92 |
| 2 | SUGGEST | 35.26 | CONCLUDED | 25.60 |
| 3 | EVIDENCE | 34.88 | SHOWS | 23.03 |
| 4 | CONCLUDED | 34.83 | SHOWED | 19.68 |
| 5 | SUGGESTS | 28.20 | ADVISED | 16.48 |
| 6 | BELIEVE | 27.61 | SUGGESTS | 16.22 |
| 7 | STATED | 25.94 | SUGGEST | 16.14 |
| 8 | SUGGESTED | 25.85 | NOTE | 16.05 |
| 9 | FOUND | 23.98 | ESTIMATED | 14.97 |
| 10 | KNOW | 22.49 | SHOWN | 13.98 |
| 11 | INDICATES | 21.99 | STATED | 13.46 |
| 12 | SHOWED | 21.94 | NOTED | 12.23 |
| 13 | NOTED | 21.25 | SUGGESTED | 12.02 |
| 14 | CLAIM | 21.24 | BELIEVE | 11.87 |
| 15 | REALISE | 19.83 | FOUND | 11.81 |
| 16 | ASSUME | 19.55 | CONCLUDES | 11.63 |
| 17 | SHOWN | 19.33 | RECOMMEND | 11.51 |
| 18 | REVEALED | 18.91 | RECOMMENDED | 11.09 |
| 19 | SHOW | 18.63 | SUGGESTION | 11.06 |
| 20 | THINK | 18.33 | SHOW | 10.78 |
| 21 | ARGUE | 18.31 | THEORY | 10.18 |
| 22 | BELIEF | 18.17 | FACT | 10.13 |
| 23 | ADMIT | 18.13 | SUGGESTING | 10.08 |
| 24 | SUGGESTING | 17.88 | SAID | 10.06 |
| 25 | CONCLUDE | 16.97 | THOUGHT | 9.98 |
| 26 | INDICATE | 16.84 | REMEMBER | 9.90 |
| 27 | POSSIBILITY | 16.83 | AGREED | 9.82 |
| 28 | CONVINCED | 16.13 | SAY | 9.82 |
| 29 | ACCEPT | 15.80 | RECOMMENDS | 9.78 |
| 30 | STATING | 15.78 | WISH | 8.96 |
| 31 | ARGUED | 15.34 | HOPE | 8.92 |
| 32 | INDICATED | 15.04 | CONCLUDING | 8.35 |
| 33 | SAY | 14.92 | STATING | 7.81 |
| 34 | DISCOVERED | 14.66 | CLAIMED | 7.62 |
| 35 | CONCLUSION | 14.64 | DEMONSTRATE | 7.23 |
| 36 | REALIZE | 14.63 | HYPOTHESIS | 7.17 |
| 37 | IMPRESSION | 14.61 | ACKNOWLEDGED | 7.00 |
| 38 | ASSURED | 14.33 | INDICATE | 6.82 |
| 39 | ASSUMPTION | 14.27 | KNOWS | 6.69 |
| 40 | CONCEDED | 13.97 | ARGUED | 6.60 |
| 41 | PROOF | 13.58 | ANNOUNCE | 6.58 |
| 42 | SUGGESTION | 13.53 | NOTING | 6.58 |
| 43 | DEMONSTRATES | 13.53 | SHOWING | 6.56 |
| 44 | HYPOTHESIS | 13.34 | KNOW | 6.54 |
| 45 | ANNOUNCED | 13.17 | SAYS | 6.38 |
| 46 | AGREE | 12.52 | CONFIRMS | 6.25 |
| 47 | CLAIMING | 12.41 | WARNED | 6.25 |
| 48 | BELIEVES | 12.20 | CONCLUDE | 6.25 |

| | | | | |
|---|---|---|---|---|
| 49 | IMPLICATION | 11.98 | CLAIM | 6.02 |
| 50 | HOPE | 11.89 | ASSUME | 5.86 |
| 51 | BELIEVING | 11.89 | STATES | 5.77 |
| 52 | CLAIMED | 11.82 | THINK | 5.45 |
| 53 | DOUBT | 11.72 | NOTION | 5.35 |
| 54 | SEEMS | 11.62 | PREDICTED | 5.35 |
| 55 | THEORY | 11.55 | ASSUMING | 5.35 |
| 56 | IMPLY | 11.53 | REMINDED | 5.33 |
| 57 | UNDERSTAND | 11.47 | PALE | 5.32 |
| 58 | IMPLYING | 11.33 | RECOGNISE | 5.31 |
| 59 | SURPRISED | 11.25 | STRONG | 5.31 |
| 60 | AGREED | 11.08 | DENIED | 5.29 |
| 61 | DEMONSTRATED | 11.04 | BELIEVES | 5.21 |
| 62 | REALIZED | 11.04 | STATE | 5.14 |
| 63 | NOTION | 10.75 | REALIZE | 5.11 |
| 64 | IDEA | 10.73 | RECOMMENDING | 4.82 |
| 65 | KNEW | 10.64 | CONCLUSION | 4.80 |
| 66 | FEEL | 10.59 | RECOGNISED | 4.74 |
| 67 | REPORTED | 10.55 | AGREE | 4.68 |
| 68 | CONVINCE | 10.50 | DEMONSTRATED | 4.64 |
| 69 | ARGUES | 10.26 | ARGUMENT | 4.57 |
| 70 | ARGUMENT | 10.25 | IMPRESSION | 4.51 |
| 71 | CLAIMS | 10.12 | ASSUMPTION | 4.28 |
| 72 | CONFIRM | 10.11 | WORRYING | 4.28 |
| 73 | CERTAINTY | 10.02 | MAINTAINS | 4.18 |
| 74 | ARGUING | 9.94 | SPECULATE | 4.18 |
| 75 | SUSPECT | 9.78 | WARN | 4.18 |
| 76 | ACKNOWLEDGE | 9.74 | | |
| 77 | NOTICED | 9.73 | | |
| 78 | STRESSED | 9.72 | | |
| 79 | FELT | 9.69 | | |
| 80 | ACKNOWLEDGED | 9.56 | | |
| 81 | ASSUMES | 9.56 | | |
| 82 | FIND | 9.54 | | |
| 83 | REASON | 9.53 | | |
| 84 | ADMITTING | 9.50 | | |
| 85 | ASSERTION | 9.49 | | |
| 86 | SAID | 9.49 | | |
| 87 | REALISED | 9.14 | | |
| 88 | CONFIRMS | 9.05 | | |
| 89 | ADMITTED | 8.99 | | |
| 90 | CONTENTION | 8.82 | | |
| 91 | ASSERTIONS | 8.76 | | |
| 92 | CONCEDE | 8.73 | | |
| 93 | ASSUMING | 8.66 | | |
| 94 | REVEALS | 8.65 | | |
| 95 | PROBLEM | 8.58 | | |
| 96 | WARNED | 8.51 | | |
| 97 | CONCLUDES | 8.44 | | |
| 98 | ADMITS | 8.31 | | |
| 99 | ASSURE | 8.30 | | |
| 100 | DEMONSTRATE | 8.30 | | |
| 101 | REASSURE | 8.14 | | |
| 102 | STATEMENT | 8.14 | | |
| 103 | BELIEVED | 8.04 | | |

| | | | | |
|---|---|---|---|---|
| 104 | CONCERNED | 8.03 | | |
| 105 | HOPED | 7.99 | | |
| 106 | ASSUMED | 7.91 | | |
| 107 | INSIST | 7.89 | | |
| 108 | THOUGHT | 7.82 | | |
| 109 | INSISTED | 7.78 | | |
| 110 | CONCLUDING | 7.67 | | |
| 111 | CONFIRMED | 7.57 | | |
| 112 | REASSURED | 7.56 | | |
| 113 | THEORIZED | 7.25 | | |
| 114 | ASSERT | 7.15 | | |
| 115 | UNDERSTOOD | 7.06 | | |
| 116 | REMEMBER | 7.05 | | |
| 117 | MAINTAINED | 7.03 | | |
| 118 | SAYS | 7.01 | | |
| 119 | MENTION | 7.01 | | |
| 120 | REMARK | 7.00 | | |
| 121 | IMAGINE | 6.98 | | |
| 122 | REVEAL | 6.87 | | |
| 123 | KNOWING | 6.85 | | |
| 124 | MAINTAINS | 6.80 | | |
| 125 | CONCERN | 6.79 | | |
| 126 | STATES | 6.55 | | |
| 127 | SPECULATED | 6.53 | | |
| 128 | ALLEGING | 6.47 | | |
| 129 | HYPOTHESIZED | 6.46 | | |
| 130 | ASSERTED | 6.46 | | |
| 131 | DOUBTED | 6.40 | | |
| 132 | ACKNOWLEDGMENT | 6.40 | | |
| 133 | SUSPICION | 6.37 | | |
| 134 | WARNS | 6.31 | | |
| 135 | ACKNOWLEDGES | 6.31 | | |
| 136 | SHOWING | 6.30 | | |
| 137 | SUPPOSITION | 6.20 | | |
| 138 | NOTING | 6.17 | | |
| 139 | CONTEND | 6.07 | | |
| 140 | ACCEPTED | 6.06 | | |
| 141 | FEELS | 6.06 | | |
| 142 | DENY | 6.00 | | |
| 143 | ASSERTING | 5.92 | | |
| 144 | ACKNOWLEGE | 5.92 | | |
| 145 | REALIZATION | 5.92 | | |
| 146 | CONTENDED | 5.89 | | |
| 147 | ASSERTS | 5.85 | | |
| 148 | ACKNOWLEDGING | 5.79 | | |
| 149 | OPINE | 5.79 | | |
| 150 | IMPLIED | 5.71 | | |
| 151 | ALLEGED | 5.70 | | |
| 152 | OPINION | 5.68 | | |
| 153 | FAITH | 5.65 | | |
| 154 | DECLARED | 5.65 | | |
| 155 | OBSERVATION | 5.48 | | |
| 156 | INSISTING | 5.48 | | |
| 157 | ADDING | 5.45 | | |
| 158 | EMPHASIZED | 5.44 | | |

| | | | | |
|---|---|---|---|---|
| 159 | PROPOSITION | 5.42 | | |
| 160 | CONTENDING | 5.40 | | |
| 161 | COMPLAINED | 5.40 | | |
| 162 | EMPHASIZES | 5.40 | | |
| 163 | GENERALISATION | 5.38 | | |
| 164 | REALISES | 5.34 | | |
| 165 | RECOGNISES | 5.34 | | |
| 166 | PROCLAIMED | 5.33 | | |
| 167 | RECOMMEND | 5.27 | | |
| 168 | CAUTIONED | 5.25 | | |
| 169 | PREDICTS | 5.25 | | |
| 170 | ADDED | 5.24 | | |
| 171 | PRAY | 5.24 | | |
| 172 | REALITY | 5.23 | | |
| 173 | PROPOSED | 5.23 | | |
| 174 | EXPLAINED | 5.21 | | |
| 175 | ANNOUNCE | 5.19 | | |
| 176 | COMMENTED | 5.16 | | |
| 177 | PERCEPTION | 5.15 | | |
| 178 | PREDICTED | 5.01 | | |
| 179 | INSISTS | 4.92 | | |
| 180 | HYPOTHESISE | 4.79 | | |
| 181 | ALLEGATION | 4.74 | | |
| 182 | DISCOVERY | 4.71 | | |
| 183 | ILLUSION | 4.54 | | |
| 184 | SPECULATE | 4.54 | | |
| 185 | RECOMMENDS | 4.53 | | |
| 186 | MENTIONS | 4.52 | | |
| 187 | MENTIONED | 4.49 | | |
| 188 | STATE | 4.44 | | |
| 189 | MYTH | 4.42 | | |
| 190 | OBSERVED | 4.41 | | |
| 191 | ACCPET | 4.19 | | |
| 192 | ACCEPTING | 4.17 | | |
| 193 | THEORIZE | 4.16 | | |
| 194 | EXPECTATION | 4.13 | | |
| 195 | RECOGNISING | 4.13 | | |
| 196 | RECKONS | 4.13 | | |
| 197 | KNOWN | 4.13 | | |
| 198 | ALLEGES | 4.09 | | |
| 199 | FEARS | 4.07 | | |
| 200 | WISHED | 4.06 | | |