

ANALYSIS OF HEALTH COMMUNICATION TEXTS

UK PRESS COVERAGE OF DEBATES ABOUT ASSISTED DYING

CLIVE SEALE

Conventionally, social researchers analyse media messages by reading text and coding it. This is time consuming and restricts many studies to small samples. Nowadays very large amounts of text are available in electronic form, offering potential insights into the health messages they contain, but which appear daunting to the analyst with limited resources using conventional methods. I explain and illustrate methods for the computer-assisted analysis of media texts, using specialised software (Wordsmith Tools and Wordstat). These allow the analyst to provide a statistical overview of the key features of texts, to compare them (looking at change over time, for example) and to select illustrative quotations that epitomise the trends identified.

Keywords: media texts, data analysis, computer assisted analysis, keyword analysis, text mining.

Analysing the health messages contained in mass media is an important aspect of health promotion. Such messages are often contained in text, large amounts of which are available to communications specialists, often by simply downloading it to a personal computer. There are databases of newspaper and magazine articles which make it easy to obtain very large samples of material relevant to health topics. For example, Nexis provides a searchable database of newspaper and magazine articles in a variety of languages from around the world. It is a simple matter to search for articles related to a particular topic and then download them. The challenge then lies in how to analyse such large amounts of material.

Conventionally, the approach taken has been to make a collection of materials and then to read them, designing a coding scheme to identify recurrent themes. Computer software, such as *NVivo*, exists to assist this kind of analysis. This is shown in a study I did with Hannah Parke concerning the UK media portrayal of policies that provide incentives (for example, cash) to people to encourage them to take up healthy behaviour (Parke, Ashcroft, Brown, Marteau, & Seale, 2013). These are examples of news coverage. We wanted to see if coverage was for or against these policies, what kind of arguments were put forward by critics or supporters, and whether this varied according to the health problem targeted. So

we designed a coding scheme. Some examples of arguments that we coded are shown below.

«Against» arguments:

- BETTER SPENT

The money being spent on incentive schemes could be better spent elsewhere, includes mentions of the cost of the schemes and references to taxpayers' money.

«Is NHS cash going to be channelled into dance lessons and vouchers for fatties when people need cancer drugs and better end-of-life care?» (Street Porter, 2008).

- BRIBE

Incentives coerce patients into making decisions they may not have otherwise made.

«The option of being paid to take a drug treatment could unduly influence people's decision making over whether the treatment is right for them» (Eaton, 2009).

«For» arguments:

- BENEFITS SOCIETY

Benefits to the rest of society, not just the recipient of the incentive.

«If it works to keep people in treatment there would be considerable benefits to the public» (*The Daily Telegraph*, 2007).

- BEST INTERESTS

Incentives help people to do what is in their best interests, including mentions of the health benefits to the individual.

«The one-off payment is intended to help pregnant mums stay fit and healthy in the run-up to the birth» (*South Wales Echo*, 2009).

Amongst other things, we found that a policy that gave expectant mothers cash to spend on healthy food in pregnancy was more popular than a policy that rewarded people who successfully lose weight. We were able to look at the arguments put forward for and against in each case, because we had coded them and could retrieve relevant text.

All of this took a great deal of time. We both had to read all of the articles, design some codes and agree on how they would be applied, modifying them when new examples came up that we could not agree on, or which did not fit existing codes. There were only about 100 articles in this study, so it was manageable.

Most analysis of this sort is done on small samples because it takes so long to do the analysis. This restricts the scope of the study and it is desirable to do much larger studies now that it is easy to obtain very large amounts of mass media text. But what do you do when a study involves thousands of articles and millions of words?

■ ANALYZING BIG SAMPLES

I faced this problem when I wanted to analyse postings to internet discussion forums by people with breast and prostate cancer, for which software called Wordsmith Tools proved useful (Seale, 2005, 2006). There were over 1.5 million words to analyse: I did not want to read them all! But I also wanted to analyse *all* of them, not just a subsample.

Wordsmith Tools allows you to compare word frequencies in two texts. For example, most common in the prostate cancer text were words like *PSA* (Prostate Specific Antigen test), *Brachytherapy* (a treatment available for prostate cancer). In the breast cancer text words like *Breast* and *Tamoxifen* were more common. All of this is quite predictable and unsurprising.

But some other words pointed to differences between men and women in what aspects of their experience they considered to be important when

«CONVENTIONALLY,
THE APPROACH TAKEN
HAS BEEN TO MAKE A
COLLECTION OF MATERIALS
AND THEN TO READ THEM,
DESIGNING A CODING
SCHEME TO IDENTIFY
RECURRENT THEMES»

facing a serious illness. Men were more emotionally restrained when they were greeting people typing *regards* rather than putting *X* (for a kiss) at the end of their messages for example. Women referred to much wider range of other people, as well as to themselves, in their messages, using words like *I, me, her, I'm, women, mum, they* and *my* much more often than men.



Thomas Lefebvre

Currently, there are newspaper and journal databases that make it easier to access and collect large samples of relevant material for health issues.

When doing this kind of analysis it is important, of course, to investigate particular words to ensure that they mean what you think (this is known as disambiguation). This is not a problem with words like *mastectomy* or *brachytherapy* but what about *X*? Does this mean a kiss or does it mean *X* as in *X-ray*?

The software makes it possible to look at individual words in their context (via «KWIC» or «Keyword in Context» displays) to establish this. It is usually best to exclude words with multiple meanings. In this collection of texts (or corpus as it is sometimes known) the word *X* was almost exclusively used for kisses rather than *X-ray*.

I have explained elsewhere (Seale & Charteris-Black, 2010) how you can move from such quantitative displays to selecting qualitative examples that illustrate key differences between texts, and how you



Wellcome Library, London

The analysis of assisted death in the British press showed, for instance, that in the 1930s people who supported voluntary euthanasia also supported eugenics as a path to societal improvement. Above, a pro-eugenics poster from the 1920s.

can do more interesting things with comparisons than simply showing women are more emotional and people-oriented than men. I have also demonstrated the usefulness of this simple Wordsmith Tools procedure in generating findings across a range of health topics, several of which involve the analysis of media text (Seale, 2008; Seale, Boden, Lowe, Steinberg, & Williams, 2007; Seale & Charteris-Black, 2008a, 2008b; Seale, Charteris-Black, Dumelow, Locock, & Ziebland, 2008; Seale, Charteris-Black, & Ziebland, 2006; Shepherd & Seale, 4).

Rather than repeating all of that, I now want to focus on a more sophisticated approach to the computer-assisted text analysis, by telling you about an unpublished study I am working on. This describes UK press coverage of debates about assisted dying (for example, euthanasia, assisted suicide) over

«WORDSTAT ANALYSIS INVOLVES THE CREATION OF A DICTIONARY WHICH IDENTIFIES THE THEMES THAT THE ANALYST IS INTERESTED IN»

time, starting in the early part of the 20th century. For this I used software called Wordstat. I used Nexis to download electronic text from five national newspapers (*The Guardian*, *The Times*, *The Daily Mail*, *The Mail on Sunday* and *The Observer*) as well as the *British Medical Journal*. Search terms were as follows (where ! denotes a wildcard suffix): *euthanasia OR assist! dying OR mercy killing OR right to die OR assist! suicide*

Existing studies of mass media coverage of assisted dying are based on small samples of text, because the approach their authors take to analysis (usually a coding approach) is so time-consuming. Wordstat enabled me to produce an analysis of all of the coverage of this topic in the newspapers and journals concerned since the modern debate about this topic began, a total of 4,340 articles.

Coverage increased over time (with a decline during World War Two). Modern debates started after 1935, with the formation of the Voluntary Euthanasia Legalisation Society and the first attempt to get Parliament to pass a law allowing euthanasia in certain cases.

Wordstat analysis involves the creation of a dictionary which identifies the themes that the analyst is interested in. Dictionary categories contain words and phrases that can be unambiguously assigned to that category after investigation of each word or phrase in its context, where necessary. The steps can be outlined as follows:

- 1) Words and phrases used in the documents are listed in descending order of TF*IDF score [Term Frequency*Inverse Document Frequency, measurement used to evaluate how important a word is in a document corpus], a measurement indicating terms that strongly distinguish documents.
- 2) The top 300 words and phrases are examined in context and those with consistently singular meanings are selected for inclusion in a user-defined dictionary whose categories group together words with semantic similarity.
- 3) Other words and phrases occurring in the texts are added to the dictionary on the basis of thesaurus similarity and further inspection of context.
- 4) The distribution of dictionary categories across comparison groups is examined.

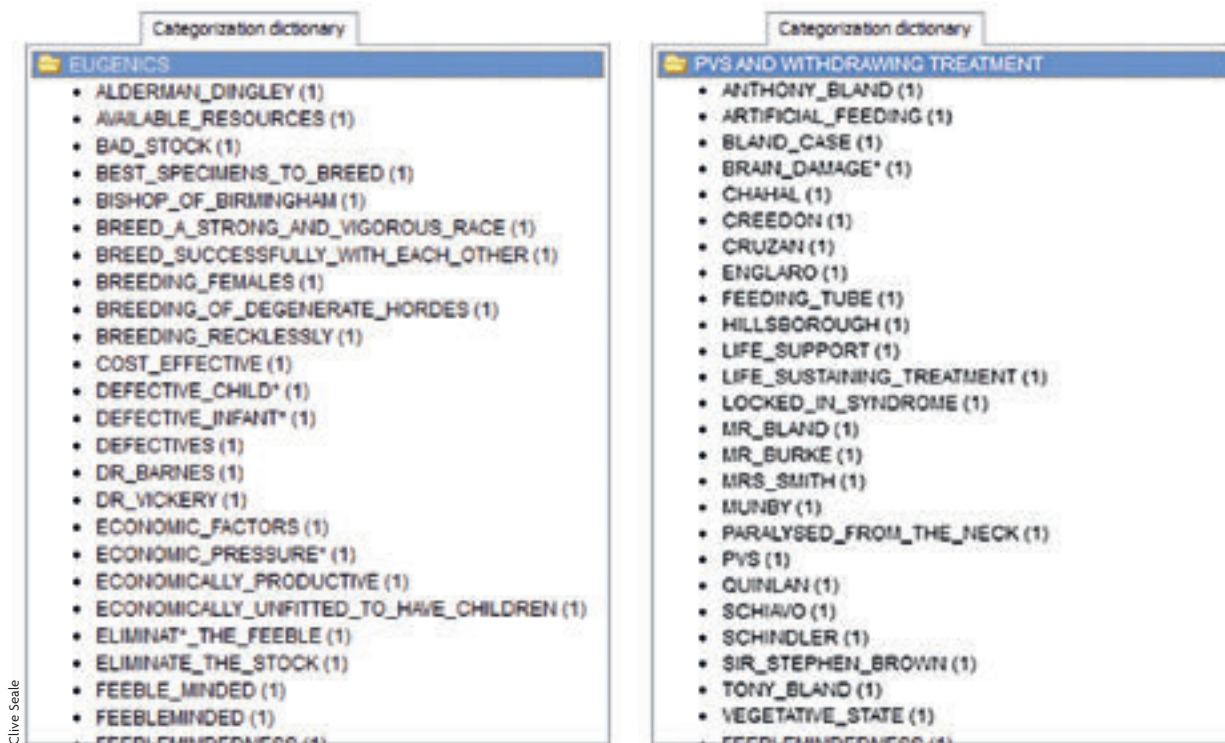


Figure 1. Examples of some categories from the dictionary created to analyse British press coverage of the debate on assisted dying (euthanasia, assisted suicide) from the early twentieth century.

In this study, the phrase showing the highest TF*IDF score was *assisted suicide*. This is the new terminology in debates about assisted dying, *voluntary euthanasia* being the old terminology, much more likely to occur in earlier texts and therefore also having a high TF*IDF score. This score identifies the words and phrases the analyst needs to focus on, because they unlock the analytic key to understanding differences between newspapers, or trends over time, or whatever comparison the analyst may be interested in.

The software also uses a built-in thesaurus to identify synonyms (words with a similar meaning). Synonyms for the word *assisted*, for example, include *aided*, *helped*, *supported* and *abetted*. Each suggested synonym needs to be inspected in its context (disambiguated) before deciding whether to include it in the dictionary. It could be, for example, that most of the uses of the word *helped* are *not* about helping someone to die, but are about helping with something else, so it would make no sense to include this word in a dictionary

category that was about assisting a death. Two or three word phrases are better to work with than individual words, as they are less likely to require disambiguation.

Some examples of dictionary categories used in the project are shown in Figure 1 above.

The full dictionary contained many more categories (you can see some more of them in Table 1) which then becomes a tool for comparing groups of texts. In this case, we can compare different newspapers, or track changes in the popularity of these themes over time. We can also use the categories to select illustrative quotations from newspapers that epitomise the category.

Some topics have declined over time, for example *Christian* perspectives less often enter the newspaper debate, and eugenicist arguments are out of fashion. Some rose in popularity and then declined, for example *Nazi*. Memories of euthanasia under the Nazis appear to be fading.

Other topics such as palliative care, human rights and assisted suicide frame the debate about

«IT IS DESIRABLE TO DO MUCH LARGER STUDIES NOW THAT IT IS EASY TO OBTAIN VERY LARGE AMOUNTS OF MASS MEDIA TEXT. BUT WHAT DO YOU DO WHEN A STUDY INVOLVES THOUSANDS OF ARTICLES AND MILLIONS OF WORDS?»

	1930s	1940s	1950s	1960s	1970s	1980s	1990s	2000s	2010-13
CATEGORIES DECLINING OVER TIME									
Christian	2.6	3.3	5.2	2.5	2.7	1.1	1.5	1.5	0.7
Voluntary euthanasia	3.2	1.8	1.6	1.1	0.9	0.9	0.8	0.4	0.1
Eugenics	0.8	4.2	2.4	0.8	0.1	0.1	0.1	0.1	0.1
CATEGORIES RISING THEN DECLINING									
Abortion	0.1	0.1	0.3	1.0	1.0	0.3	0.7	0.3	0.1
Advance declaration	0	0	0	0	0.1	0.04	0.5	0.4	0.1
Mercy killing	0.7	0.7	3.5	1.4	1.5	0.6	0.8	0.4	0.5
Nazi	0	1.3	1.2	3.1	0.3	0.4	0.4	0.3	0.2
PVS and treatment withdrawal	0.05	0	0	0.1	1.1	0.5	2.6	1.2	0.6
CATEGORIES RISING OVER TIME									
Assisted suicide	0	0	0	0.03	0.1	0.2	0.8	3.0	4.9
Human rights	0.1	0	0	0.02	0.02	0.02	0.03	0.22	0.12
Law and legal	0.02	0	0.23	0.1	0.4	1.5	1.1	1.9	2.4
Palliative Care	0	0	0	0.1	0.5	0.5	0.7	0.8	0.6
Pro legalisation cases	0	0	0.1	0	0.03	0.2	0.3	1.5	2.2
Slippery and vulnerable	0.2	0	0.3	0.4	0.2	0.2	0.5	0.6	0.7
Right to die	1.8	0.2	0.4	0.2	0.7	0.7	0.8	1.0	0.9
Wishing and wanting	0.1	0	0.2	0.2	0.3	0.3	0.4	0.4	0.5
TOTAL WORDS	40,392	8,307	30,722	86,990	125,767	140,844	336,965	687,168	259,087
Average words per year during each decade	4,039	831	3,072	8,699	12,577	14,084	33,697	68,717	64,772

Table 1. Selected dictionary categories by decade: words per 1,000. They show the importance of different topics, such as eugenics, decreasing over time, while others, such as human rights or end-of-life care, gain importance over time.

SOURCE: Clive Seale.

assisted dying more commonly nowadays than in the past. The arguments have moved on, old ones have been forgotten, or are deemed irrelevant, and new ones have been put forward by both opponents and proponents of legalisation. The software at this stage provides a statistical overview of these changes.

But because at every stage these statistical displays are linked to the underlying text, it can be retrieved to provide illustrative quotations and to allow for more detailed, qualitative scrutiny of key arguments. I will show you just one of these arguments over time: eugenicist arguments.

The Guardian on 5th September 1935 reported the views of Dr T. W. Hill, Deputy Medical Officer for Health in West Ham, who had addressed the New Health Society's summer school at Margate (UK), asking what they were to do about the thousands of low-grade imbeciles in their institutions:

When we see the hideous monsters that are kept alive and nursed by trained people... we are inclined to ask: Can civilisation devise no method of dealing with them? Are we not stretching humanity too far? Such creatures would be better out of the way. We are doing our best to promote the survival of the unfittest. If we are to breed

a strong and vigorous race we must reverse our false notions of the sacredness of human life. We need not return to the ruthless code of natural selection, but we can achieve the same results painlessly, quickly, and without cruelty, by scientific euthanasia.

(*The Guardian*, 1935)

Here you can see that Dr Hill is happy to describe people with cognitive and other disabilities as «imbeciles» and «monsters» and to kill them off in the interests of society as a whole. In the 1930s, people who supported voluntary euthanasia were often also supporters of eugenics as a way of improving society. The experience of the Nazis changed all of this, so that after the war, such arguments had to be put in much softer ways, and opposition to them was recorded by journalists. For example, on 30th April 1969 *The Guardian* reported Dr Kenneth Vickery, medical officer for health for Eastbourne (UK) noting the need «to start planning now for the growing number of old people in society, who were already swamping health and welfare services in areas like his own». He was not advocating euthanasia, but he did suggest that «people beyond a certain age should not be kept alive by



Wellcome Library, London

Modern debates on assisted death started circa 1935, with the formation of the Voluntary Euthanasia Legalisation Society and the first attempt of the UK Parliament to pass a law that would permit euthanasia in certain cases. The photograph above shows the founding members of the Voluntary Euthanasia Legalisation Society.

medical ingenuity but simply offered nursing care until they died» (Shearer, 1969). A spokesman from Help the Aged was said to have reacted to this view with horror and revulsion.

And nowadays, the only people who dare to put forward such unpopular ideas publicly are people who want to be notorious, self-publicists like the novelist Martin Amis who in 2010 was reported as having suggested «euthanasia booths» should be set up in the streets so that old people who had become a burden on society could end their lives with «a martini and a medal», thus conserving scarce resources and avoiding civil war between the old and the young. Dignity in Dying (the new name for the Voluntary Euthanasia Legalisation Society) were widely reported as having distanced themselves from his stance.

■ COMPUTER ASSISTED METHODS

These methods, using the power of computers and specialised software, allow us to analyse very large amounts of text in a way that is less influenced by the preferences of the analyst than the time-consuming coding approach. They provide a statistical overview of differences between texts, allowing for all kinds of interesting comparisons to be made. They are

**«COMPUTER ASSISTED
METHODS ALLOW US TO
ANALYSE VERY LARGE
AMOUNTS OF TEXT. THEY
PROVIDE A STATISTICAL
OVERVIEW OF DIFFERENCES
BETWEEN TEXTS, ALLOWING
FOR ALL KINDS OF
INTERESTING COMPARISONS
TO BE MADE»**

suiited to the conditions of today, where it is easy to obtain very large collections of electronic text from the internet, but difficult to analyse by conventional methods that depend on reading the text. ☺

REFERENCES

- Eaton, L. (2009, 12 October). Trial that pays patients to take antipsychotic drugs comes under attack. *British Medical Journal*, 339, b4202. doi: 10.1136/bmj.b4202
- Hall, C. (2007, January 27). NICE wants incentives for addicts days after refusing cancer drug. *The Daily Telegraph*, p. 8.
- Parke, H., Ashcroft, R., Brown, R., Marteau, T. M., & Seale, C. (2013). Financial incentives to encourage healthy behaviour: An analysis of UK media coverage. *Health Expectations*, 16(3), 292–304. doi: 10.1111/j.1369-7625.2011.00719.x
- Seale, C. (2005). Portrayals of treatment decision-making on popular breast and prostate cancer web sites. *European Journal of Cancer Care*, 14(2), 171–174. doi: 10.1111/j.1365-2354.2005.00544.x
- Seale, C. (2006). Gender accommodation in online cancer support groups. *Health*, 10(3), 345–360. doi: 10.1177/1363459306064495
- Seale, C. (2008). Mapping the field of medical sociology: A comparative analysis of journals. *Sociology of Health and Illness*, 30(5), 677–695. doi: 10.1111/j.1467-9566.2008.01090.x
- Seale, C., Boden, S., Lowe, P., Steinberg, D., & Williams, S. (2007). Media constructions of sleep and sleep disorders: A study of UK national newspapers. *Social Science and Medicine*, 65(3), 418–430.
- Seale, C., & Charteris-Black, J. (2008a). The interaction of class and gender in illness narratives. *Sociology*, 42(3), 453–469. doi: 10.1177/0038038508088835
- Seale, C., & Charteris-Black, J. (2008b). The interaction of age and gender in illness narratives. *Ageing and Society*, 28(7), 1025–1043. doi: 10.1017/S0144686X0800737X
- Seale, C., & Charteris-Black, J. (2010). Keyword analysis: A new tool for qualitative research. In I. L. Bourgeault, R. DeVries, & R. Dingwall (Eds.), *The SAGE handbook of qualitative methods in health research* (pp. 536–556). London: SAGE.
- Seale, C., Charteris-Black, J., Dumelow, C., Locock, J., & Ziebland, S. (2008). The effect of joint interviewing on the performance of gender. *Field Methods*, 20, 107–128. doi: 10.1177/1525822X07313837
- Seale, C., Charteris-Black, J., & Ziebland, S. (2006). Gender, cancer experience and internet use: A comparative keyword analysis of interviews and online cancer support groups. *Social Science and Medicine*, 62(10), 2577–2590. doi: 10.1016/j.socscimed.2005.11.016
- Shearer, A. (1969, 30 April). Elderly should be «allowed to die». *The Guardian*, p. 5.
- Shepherd, E., & Seale, C. (2014). Eating disorders in the media: The changing nature of newspaper reports. *European Eating Disorders Review*, 18(6), 486–495. In K. Eli, & S. Ulijaszek (Eds.), *Obesity, eating disorders and the media* (pp. 91–106). Farnham: Ashgate.
- South Wales Echo. (2009, 30 April). New grant for mums-to-be. *South Wales Echo*, p. 9.
- Street-Porter, J. (2008, 13 November). What we really need to solve our obesity epidemic. *Belfast Telegraph*, p. 36.
- The Guardian (1935, 5 September). Problem of human suffering: Scientific euthanasia. Doctor urges its adoption. *The Guardian*, p. 6.
- Clive Seale.** Professor of Sociology at Brunel University in London (UK). He was editor for the journal *Sociology of Health & Illness* 2006–2012 and is the author of many papers and books such as *Media & health* and *Qualitative research practice*. In recent years, his research has focused on end-of-life care and health care communication.