# Corpus-Assisted Discourse Studies (CADS)

## Video 3 transcript

Full resource, see: <https://www.ncrm.ac.uk/resources/online/all/?id=20855>

In this video we’re looking at a further example of CADS. This time we’re looking at a project I carried out on how social workers are portrayed in the UK press. So, my aim here was to explore the commonly held belief by social workers that UK newspapers portrayed them quite negatively, that they dwell on negativity rather than neutral occurrences or saying anything positive. So, my research questions here were to look at the proportions of positive, negative and neutral mentions of social worker in a few UK national newspapers, and then to look at the dominant categories within each of these groupings.

I built my own corpus using Nexis UK, now called LexisNexis, and which may be available if you’re in a UK university through your library. And here I used the search term social worker or social workers plural to look at all print and online UK national newspapers over a three-month period. So, I filtered it to include only UK national press. I found 987 hits within 736 articles covering over 5 million words of newspaper. And then I examined this corpus using corpus linguistic software.

So, I actually began this project from the discourse analytic side by reading a few newspapers to see what was going on, to see if they were really interesting to explore. I looked at LexisNexis and searched for social worker and downloaded a few examples to see what sort of negativity surrounded the job title social worker. Is there potential for a study on this? Will a three-month sample of newspapers give me enough to go on? And then I created my corpus and went from there.

And here’s what the data looks like in AntConc, sorted first, second, third words to the right, and you can see the different colour coding there.

I then exported lines to Excel because I feel there’s greater functionality there to add different categories. So, you can see here in column B I’ve got concordance lines, and you can see they’re quite long and they were actually a whole paragraph each time. You can decide how much data you want to export. And then I categorised them in different ways and I had a second in column D, subcategory, other areas, comments on classifying. Column G says whether it’s a children’s social worker or adult’s social worker. And you can see this is a sort of work-in-progress part of my categorisation within Excel. But it gave me the ability to sort in different ways and to really sift the data to delve further into it, to go back to individual texts and work out what was going on. So, this is really the discourse analytic phase where I’m perhaps going back to the corpus, searching for more data, looking for more examples of the same type, then back again to individual texts to read in a little bit more depth.

And here at the bottom I’ve given a kind of collapsed the data together, so I’ve got neutral, positive, and different types of mention, and I’m trying to begin the process of categorising thematically. So, some of them I categorise by job title, the social worker as a job title was mentioned. Some of them the social worker was mentioned but it was nothing to do with being a social worker, it was just somebody’s job that happened to be mentioned, and so on. So really this was a lengthy process of gradually sifting the data until I could account for all the data within my categories.

The broadest level of classification of each of the 974 instances of social worker in the entire corpus was negative, positive, or neutral mentions. And here I found that 25% of the instances were negative, so a raw count of 242. And only 6% were positive. So, the social workers were quite right in that the UK press tend to be quite negative towards them. And I then categorised within that grouping of negativity into the ones you see here on the screen. So, the largest by far here is the red line, failure to notice or act. And this denotes occurrences ranging from social workers being described as dismissive of concerns raised by others, or where it was deemed that they failed to take sufficient action. And this is far greater than the yellow over-zealous, where they were seen to take children away too quickly or act too swiftly in some way.

Here are the original negative concordance lines again, together with my classifications. So, you can see here that social workers are treated quite badly really by the UK press, and newspaper writers tend to dwell on cases where things go wrong and then perhaps hold social workers to account when it’s maybe rather a larger societal issue. I could have maybe come to this conclusion from reading newspapers, but by building a corpus I was able to be a little bit more rigorous about this and gather 1000 instances over a short period of time, a three-month period of the press reporting, and look at how many of those occurrences were negative, how many were positive, how many were neutral, and then systematically sort them and make some rigorous claims about how social workers are viewed in the press.

[0:06:55]

Ideally, you’d have a second rater for this. I rated myself over a two-month period, so I had two months between first rating and second rating, and largely I had the same findings the second time around. I think it’s better to have another person to have another objective view. I did try using ChatGPT but it wasn’t great, however it is improving, and it tended to come up with quite superficial versions of negativity. So here I think it’s really important that you’re transparent over how you’re categorising your data.

And the corpus is used as an entry point, followed by this extensive qualitative research where you’re going back and forth to the original texts. So, this is an example of CADS research and I would imagine you can think of other ways in which it can be carried out.

National Centre for Research Methods (NCRM)  
Social Sciences  
Murray Building (Bldg 58)  
University of Southampton  
Southampton SO17 1BJ  
United Kingdom

**Web** www.ncrm.ac.uk   
**Email** info@ncrm.ac.uk  
**Tel** +44 23 8059 4539  
**Twitter** @NCRMUK