Coding of Danish radio news 1984-2003

A data report

Christoffer Green-Pedersen

Rune Stubager

Department of Political Science
University of Aarhus
Bartholins Allé
8000 Århus C
Denmark
Email cgp@ps.au.dk/stubager@ps.au.dk
Web www.ps.au.dk/greenp

August 2007

Introduction

The following data rapport documents coding of Danish radio news in the period 1984-2003. It contains information on the different decisions made when coding the data and the different procedures set up to secure validity, reliability and quality of the data. The report thus serves two purposes. As background material for the publications from the project, it provides an introduction to the considerations behind the creation of the dataset. Hopefully, the report will provide the reader with the kind of insight into the details of the database which can never be found in academic publications. Second, the report is meant as a source of inspiration for other researchers aiming to build similar datasets. It covers the following issues: why code Danish radionews, the documentation of the radionews, the content code variables, other variables, coding procedures and coding problems.

Why code radio news?

The aim of creating this database was to have a long term time series measuring the content of Danish news coverage. Or to phrase it differently to have a measure of the mass media agenda over a long time period.

Mass media are many different things, i.e. newspapers, TV, internet, radio etc, so the question of what media to code is central. A dataset drawing on a number of different media would in a way be preferable, but building such a database for a long time period would be a giant task. Further, a study of Danish mass media coverage (Lund 2004) clearly shows how different mass media borrow stories form each other and this points to the sameness of media content across mass media. This study thus described a "food-chain" of mass media content where most stories are of two sorts. First, stories more or less copied from press agencies etc. for instance, describing events abroad etc. Second, stories started by the major Danish morning newspapers, which then in many but not all cases travel through radio and television news during the data. Further this study points to the central role of the 12 o'clock radio news (approx. 30 minutes) as a gatekeeper for mass media coverage during the day. The radio news contain few new stories, but pick among the stories of the morning news papers. Making it into the 12 o'clock radio news is thus necessary for a news story to travel during the day.

This picture of the radio news was confirmed by its former leader, Søren Elmqvist, and was reflected in the editorial guidelines from the early 1990s. The focus of the news was on presenting existing news in a qualified way rather than creating new news stories through its

journalists (Interview with Søren Elmquist, January 2007). This made the Danish radio news, especially at 12 o'clock an obvious choice as a media indicator. Finally, surveys made by "Danmarks Radio" revealed that the 12 o'clock news had more than a million listeners out of a population of some five million (op. cit.).

The main news at 12 o'clock was obvious to code. However, a potential problem related to only using this news broadcast was that stories starting due to events in the afternoon would not be covered. To some extent these events might come up in next day's radio news, but coding the 12 o'clock news only would contain a bias towards stories starting in the morning, not least stories coming from newspapers. Therefore, the radio news at 1830 (approx 20 minutes) were included as well. The content of this news can be expected to be close to the content of television news in the evening.

In format, the Danish radio news has remained very stable in the period. The exact length of the news has varied slightly – they have always been shorter in the weekends, but the format has stayed the same.[1] According to Søren Elmqvist who was leader of the radio new from the early 1980s to the mid 1990s, there were no significant changes to the idea behind the news, except a tendency to include shorter and thus more stories from the late 1980s.

Documentation of the radio news

A further reason for choosing to code radio news was that we have access to a summary of each news broadcast. These summaries contain a list of the features in the news containing a few lines on the content, the names of the actors appearing in them and the length of interviews, reports etc. included in them. They are used by the journalists reading the news. These summaries thus contained most of the information we were interested in and were an easy material to code.

The available material was almost complete, but approximately 20 news broadcasts are missing for instance due to strikes or the material has simply been lost with The Danish Broadcasting Corporation.

A question of course was whether these summaries were actually similar to the radio news. Would for instance last minutes changes constitute a problem when using these summaries for coding? To investigate these issues, we acquired tapes with the actual radio news for June 1 for

---

[1] Only in 2004, some changes were made. Before, the Danish national broadcasting corporation (DR) produced only one series of news, which was then sent out on all it channels every hour – in most case only lasting 5 minutes. This was changed in 2004, so there is special news broadcast for the different channels. The existing radio news was continued on channels 1 and 4 (P1 and P4) and we continued to code these. Further the long evening news was moved from 1830 to 1800 on weekdays and to 1700 in the weekends

the period 1984 to 2004. We listened to this news and compared with the summaries. The summaries turned out to be very close to the actual news. Besides a few case where the order of minor items was changed marginally, only technical breakdowns during broadcasting caused deviations between the summaries, which are written in advance, and the actual news. This insured us that when coding from the summaries, we were actually coding the radio news as broadcasted. Further, the summaries were set up as a list of the features in the news, which was perfect as features can be seen as a "news story", which is the unit of analysis.

What to code

In terms of the variables that have been coded for each feature, they can be divided into three types, descriptive variables, a content code and an actor code.

Descriptive codes

For each feature, day, month, and year of the feature was coded. This allows us to uniquely identify each feature. These variables also allowed us to automatically generate a variable identifying which news broadcast the feature belonged to. Further, the number of the feature as it appeared on the list in the summaries was coded. This variable thus allows us to possibly distinguish between top stories and marginal stories as the order of the features is likely to represent a journalistic judgment of importance of the news content. The length of the features was also of interest to us. Here, the summaries provided information on the length (in minutes and seconds) of for instance interviews, i.e. everything not read by the speaker. However, this gave us no indication of the length of cables and whatever introduction and further remarks were given by the speaker when features contained interviews etc. In order to remedy this, a student recorded the length of each such reading of a cable, introduction etc for the radio news on June every year 1 in the period. This provided us with a sample of the length of 241 cables and 271 introductions etc. Distributions of these lengths were somewhat skewed due to a few high values so we used the median of each type, 36 second and 20 seconds. 95% confidence intervals for the two figures are (33.58; 38.42) and (18.25; 21.75), respectively. Given the difference between the time it takes to read a cable and an introduction to an interview, a variable was coded indicating whether the feature was a cable or one containing an interview etc. In the first case, the value of 36 seconds was then automatically inserted. In the second case, the coders inserted the length of interviews etc. found in the summaries and 20 seconds were then automatically added to cover the introduction etc.

Finally the coders wrote a short summary of the news stories. The aim of this was to allow for keyword search in case the categories of the content codes were not appropriate for a specific purpose. These summaries are short keywords or sentences.

Content codes

The aim of the content codes was to have data on the content of media stories and to compare these with the content of the political agenda and voters' agenda. The design of the content codes was thus set up to match that of the coding of parliamentary activities (see Green-Pedersen 2005). These were coded in 236 subcategories and 19 main categories. However, for the media data, we decided that we could do with a less detailed coding scheme. The 236 subcategories in the parliamentary coding scheme had been compiled into 24 main categories like, traffic, justice, health, energy, the environment etc. We used these 24 main categories as the starting point. We then added two categories, namely covering the royal family and for special features in the radio news such as live transmission from sports events, weather forecasts etc. Further, we decided to subdivide some of the main categories for instance economics into unemployment, taxation. Especially, foreign policy had to be split into several categories, some of them referring to Danish foreign policy and others referring to events in foreign countries which take up a considerable part of the news. The EU was handled in the same way as in the parliamentary coding. A dummy variable was created to measure each time the EU was mentioned, and the content coding then followed policy substance for instance environment or agriculture. However, in some cases, EU was also policy substance, i.e. enlargement or a new commission, so therefore we also have a category (2000) for EU as a substantive area

Actor codes

In order to understand media processes better, the actors in a news feature are also coded. Two considerations have been crucial in this regard. First, who are the actors in a feature and, second, to develop a coding scheme for the actors. Concerning who are actors, this proved to be one of the most difficult coding elements. Actors who are interviewed or whose statements are cited etc. are clearly actors, but what about actors who are criticized, convicted etc.? Here, the rule for the coders is to focus on the news element in the story. If the stories start by describing an act or a statement from an actor, which is then commented or criticized by another actor, both actors are coded. However, if the story starts by referring to past statements or deeds and the news story is then the

comment etc. by someone else, only the latter is coded as an actor. With regard to the actors coding scheme, we divided actors into, government actors, political parties, interest organizations, and bureaucrats. With regard to the three former types, the coders also have to specify the exact actors (specific party, minister etc.).

Securing data quality

Securing quality and reliability of the data is an obvious issue with regard to coding of this sort and several steps have been taken to ensure both. Given the extent of coding-work, it had to involve a number of student coders and 11 students were hired. Many variables such as the number of a feature, length etc are easy to code – the values in the summaries just had to be typed in, but the content codes and actors are much more demanding partly because of the many categories and partly because of the judgement which inevitably is involved in deciding which category is more appropriate for a question. Therefore, several procedures were set up to secure a common understanding of the coding scheme.

Most of the coding was done in Microsoft Access. A programmer created forms which allowed several coders to code at the same time and the program could be given default values and a range of values to accept. This reduces the number of typing errors. The 11 students hired went through a week of intensive training in the content and actor coding schemes consisting of discussions of the coding scheme and test coding.

When the coding was initiated, we set up two procedures to secure the reliability of the content and actor codings (the other codes involved no judgment from the coder). First, the students could put a "?" in the comments field for each items where there was doubt about the coding. When each session had been coded, Christoffer Green-Pedersen went through all the items where the coders had been in doubt and sometimes recoded them. Via e-mail or orally, we also discussed the items that had caused the problem.[2] Second, systematic reliability test were set up (See Togeby 1974). These tests were done in the following way: For each coder, Christoffer Green-Pedersen recoded a number of randomly selected new summaries and calculated mistakes. For the content code this was simply when then coder had coded differently from Christoffer Green-Pedersen. For the content code, 90% was set as target, i.e. only 10% different codes. All coders

---

[2] Even though the codebook had been developed by coding approximately 2500 items, we of course constantly ran into issues that required some general decision about how they should be coded. Still, we found no revisions to the codebook necessary.

made fewer mistakes, typically 5-8%. Recoding continued until N was so large that we, based on confidence intervals, could be 95% sure that less than 10% mistakes were made. A minimum of 100 sentences was coded, and the maximum was 269. For the actor codes, this was done slightly differently. The actor coding of each news feature involved both a decision whether to code an actor and sometimes the coding of multiple actors.[3] Every mistake was counted as mistake, whereas each news feature only counted once when calculating the total number of codings. The number of coding mistakes could thus exceed the number of news features. In reality the N in connection with the actor coding was thus in reality considerably larger than the number of news features. Therefore only 85% was set as a target. Recoding thus continued until we could 95% sure that the actual number of mistakes would not exceed 15%. The actual number of mistakes varied from 5 to 9 and the 85% target could thus be reached with an N of 100 or a little more.

The reason for Christoffer Green-Pedersen doing the recoding during the coding period was that it provided a possibility for detecting coding mistakes and making the coders aware of them. In this regard it is important to be aware that there are several different reasons for coding mistakes, i.e. differences between the two codings of the same material, some of systematic and others of unsystematic nature. The systematic ones are clearly the most problematic since the unsystematic ones will even out due to the large N. One source of systematic errors is misunderstandings of the codebook. This is an important problem as coders would systematically code features in the wrong categories. Detecting such errors and correcting them is an important reason for performing reliability tests during the coding. Another source of systematic errors stems from the coders misunderstanding the content of the news and coding them in the wrong categories. This is probably the biggest source of systematic errors, but one that can only be remedied through training of the coders. With regard to unsystematic errors there are also two courses. One is that the features are occasionally very short and thus involves an element of coder judgement with regard to category. Another and related is that a feature can be about several issues and there could thus be several relevant categories, but only one can be coded. Here the coders had to make a judgement about the most appropriate category, and this leads to difference in coding. However, this difference is likely to be unsystematic.

---

[33] For a number of news features, mainly those referring to news from abroad, no actors were coded, and these features were disregarded with respect to the reliability tests of actor codes:.

References

Green-Pedersen, Christoffer (2005). *Coding of Parliamentary Activities in Denmark, 1953-2003. Data rapport*. Aarhus: Department of Political Science, University of Aarhus.

Lund, Anker B. (2002). *Den redigerende magt*, Århus: Aarhus Universitets Forlag.

Togeby, Lise (1974). *Kodningens reliabilitet* , Aarhus: Department of Political Science.