Qualitative and Quantitative Methods in Libraries (QQML) 1:263 –270, 2014

Understanding News Researchers through a Content Analysis of Dissertations and Theses

Mary Feeney

Associate Librarian, Research Services Team, The University of Arizona Libraries

Abstract. Understanding the information needs of researchers is an important role of academic libraries. This paper describes the methodology and results of a research study that explored the use of newspapers by scholars in different disciplines through a content analysis of dissertations and theses. The purpose of the study was to examine which disciplines utilize newspapers, whether U.S. or non-U.S. newspapers are used more frequently, and whether researchers typically consult historical or current newspapers for their research. Results indicate that scholars in a wide range of disciplines utilize newspapers in their research. Analysis of the time periods used by disciplines and the origin of newspapers used in research is included.
Keywords: Newspapers, Content Analysis, News Research, Information Needs, Dissertations, Graduate Student Research

1. Introduction

One role of academic libraries is to provide information resources needed by faculty, students, and researchers. Newspapers are one type of information resource that may be needed across disciplines. Much has been written about newspapers as sources of information for historians (such as Allen and Sieczkiewicz, 2010; Allen, Zhu, Sieczkiewicz, 2010; Bingham, 2010; Jones, n.d.; Salmon, 1923), as well as for scholars in journalism and communication. Jones (n.d.) notes that, "while historians are perhaps the largest community that uses historic newspapers, they also serve as important source materials for other groups of users as well." Deacon (2007) notes that "news archives are a key research resource for academics across the humanities and social sciences." Jones (n.d.) summarizes newspaper research done by scholars in other disciplines, such as linguists, geologists, and archaeologists, and concludes that

Received: 19.4.2013 / Accepted: 20.1.2014 ISSN 2241-1925

"historic newspapers are used for a variety of purposes by a large number of different communities." The aim of this research study was to learn more about which disciplines use newspapers in their research, the origin of the newspapers used, and whether newspapers from certain time periods are used more frequently. To better understand the needs of scholars, a content analysis was conducted to examine the use of newspapers in graduate student research.

2. Methodology

Content analysis is a research method for "defining, measuring, and analyzing both the substance and meaning of texts" (Beck and Manuel, 2008). It involves the classification of documents based on defined categories, usually delineated in a codebook. The author created a codebook for this study detailing eight variables to be coded in the content analysis:

- Relevancy to the study. Documents were considered relevant to the study if the content of newspapers was directly used in the dissertation or thesis research. Documents that were coded as "not relevant" included ones in which newspapers were referenced from secondary sources; newspapers were mentioned in the document, but the content of the newspaper was not used for research; and newspapers were used for other reasons, such as the recruitment of subjects for a study.
- Subject headings. The subject headings as listed in the Proquest Dissertations and Theses (PQDT) database record.
- University/college of the author. This information also came from the PQDT database record.
- Level of degree. Ph.D. or Master's degree
- Year of degree. Information from the PQDT database record.
- Discipline of the author. The Taxonomy of Disciplines from the Council of Graduate Schools/Educational Testing Service's Survey of Graduate Enrollment was used for this study.
- Origin of newspapers. There were three coding categories for newspapers used in the documents: United States, non-United States, or both.
- Time periods of newspapers. Time periods of newspapers used in the documents were tallied in these categories: pre-1700, 1700-1749, 1750-1799, 1800-1849, 1850-1899, 1900-1949, 1950-1999, 2000-2008, or unknown.

The ProQuest Dissertations and Theses (PQDT) Full Text database was used as the source of documents for this study. ProQuest describes the database as "the world's most comprehensive collection of dissertations and theses" and the "database of record for graduate research" (ProQuest 2013). It covers a wide range of research across disciplines. Initially, a search was conducted on the keyword "newspaper" in the full document text of dissertations and theses published during a five-year period, but this resulted in over 20,000 results. At the time of this study, only the first 1,000 results were presented in PQDT. In

addition, a scan of the results indicated that a full text search included many irrelevant results. For example, the term "newspaper" may have been used in passing somewhere in the full text of the dissertation, but newspapers were not used as part of the research.

Since this was an exploratory study, documents for content analysis were identified by searching for the term "newspaper" in the citation and abstract field for full text English language dissertations and theses published between 2004 and 2008. Other studies have used the abstracts of dissertations and theses as the basis for their research (Alidousti et al, 2009). The search resulted in 1,882 dissertations and theses, which were exported into RefWorks and then imported into a Microsoft Excel spreadsheet that served as the master coding sheet.

The author and a research assistant coded a sample of the results for pilot intercoder reliability testing. Intercoder reliability is "the extent to which independent coders evaluate a characteristic of a message or artifact and reach the same conclusion" (Lombard et al, 2002). One of the reasons that a satisfactory level of intercoder reliability is important is "to provide basic validation of a coding scheme…it must be established that more than one individual can use the coding scheme as a measurement tool, with similar results" (Neuendorf, 2002). An assessment of reliability should be done "at two points in a content analysis: pilot and final" (Neuendorf, 2002), and the pilot should be done before the full study begins. Forty-seven documents were coded by both coders, and Scott's pi was used to calculate intercoder reliability, which ranged from .75 to .84 for the different variables.

Revisions were made to the codebook to improve reliability. For example, additional codes were added to the discipline category to account for issues discovered during pilot testing. In some cases, the discipline of the author could not be determined from the document. To account for those cases, an additional code not listed in the Taxonomy of Disciplines was added to the codebook.

Each dissertation and thesis from the search results was assigned a document identification (ID) number. Next, a sample of 321 documents was identified for coding for the full study. The sample was selected randomly using a random number generator, and the document with the corresponding document ID was included in the sample. The sample size was determined using a table in Basic Research Methods for Librarians (Powell, 2004). For a population of 1,800, a sample of 317 is needed. The author and the research assistant coded documents in the sample using the schema outlined in the codebook.

A final intercoder reliability test was conducted on 50 randomly selected documents coded by both coders. One measure of reliability is simple percent agreement between coders, but some simple agreement may be due to chance. Two formulas that take chance agreement into account are Scott's pi, which Riffe, Lacy, and Fico (2005) describe as "one of the most frequently used" reliability tests, and Cohen's kappa that Neuendorf (2002) notes is one of "the most widely used reliability coefficient." Lombard et al. (2002) explain that Scott's pi "takes into account the number of categories as well as the

distribution of values across them," with Cohen's kappa using the "same conceptual formula."

There is not agreement in the literature about what is considered a satisfactory level of reliability. Neuendorf (2002) summarizes research on this topic, and concludes that "reliability coefficients of .90 or greater would be acceptable to all, .80 or greater would be acceptable in most situations." Lombard et al (2002) states that ".70 may be appropriate in some exploratory studies." Scott's pi and Cohen's kappa were calculated using ReCal ("Reliability Calculator"), an online program for calculating reliability (Freelon, 2010, and Freelon, 2013). Table 1 provides the intercoder reliability measures, calculated for each variable, using percent agreement, Scott's pi, and Cohen's kappa.

Table 1: Intercoder reliability results

Variable	Percent Agreement	Scott's pi	Cohen's kappa
Relevance	94	0.84	0.84
Discipline of author	86	0.83	0.83
Origin of newspapers	94	0.90	0.90
Time period of newspapers – pre-1700	94	0.84	0.84
Time period of newspapers – 1700-1749	94	0.85	0.85
Time period of newspapers – 1750-1799	94	0.85	0.85
Time period of newspapers – 1800-1849	90	0.79	0.79
Time period of newspapers – 1850-1899	90	0.84	0.84
Time period of newspapers – 1900-1949	92	0.87	0.87
Time period of newspapers – 1950-1999	90	0.85	0.85
Time period of newspapers – 2000-2008	88	0.82	0.82
Time period of newspapers – unknown	100	1.00	1.00

3. Results/Discussion

As described in the methodology section, dissertations and theses were considered relevant to the study if newspapers were directly used in the research; 82% of the sample documents were relevant to the study. For the documents that were not relevant, the term "newspaper" appeared in the abstract; but upon analysis of the full text, it was determined that newspapers were either referenced from secondary sources or mentioned in passing, or newspapers were used for other reasons, such as the recruitment of subjects for a study.

About two-thirds of the relevant documents were dissertations; the remaining one-third were theses. They were distributed fairly evenly by year of degree: 45 (2004), 58 (2005), 50 (2006), 61 (2007), 48 (2008).

Examining the disciplines of the authors who used newspapers in their research yielded some of the most interesting results. For about 41% of the dissertations and theses, the authors were in history (23.6%) or communication/journalism (17.5%). This is not surprising, given the research cited in the introduction regarding the use of newspapers by historians and communication/journalism scholars. However, about 60% of the documents were written by authors in a wide range of other disciplines, as table 2 demonstrates. The "undetermined" category includes dissertations and theses in which the author's department, program, or major was not stated explicitly in the document. For the purposes of this content analysis, the author decided not to look for this information beyond the document itself. While the number of dissertations and theses in some areas was very low, it is interesting to see how many different disciplines were represented in this sample.

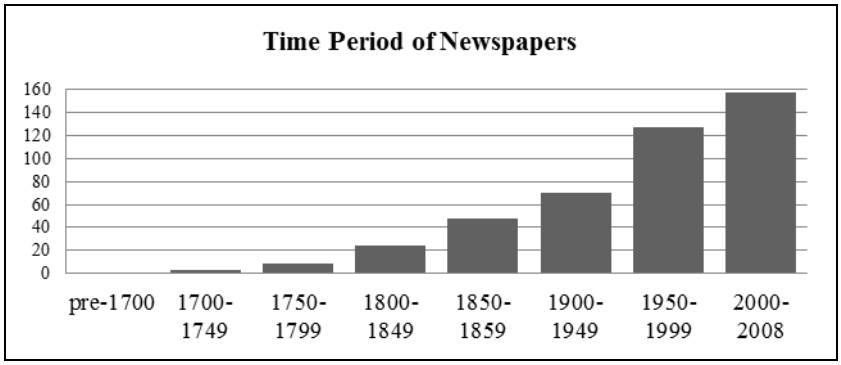
Table 2: Disciplines using newspapers

Discipline	**Percentage**
History	23.6%
Communication and Journalism	17.5%
Undetermined	10.6%
Multiple disciplines/more than one category	6.1%
Political Science	5.7%
Sociology	4.9%
Education, Other	4.6%
Social Sciences, Other	4.6%
English Language and Literature	3.8%
Arts - Performance and Studio	2.3%
Health and Medical Sciences	1.9%
Other Fields	1.9%
Arts - History, Theory, and Criticism	1.5%
Arts and Humanities, Other	1.5%
Architecture and Environmental Design	1.5%
Economics	1.1%
Foreign Languages and Literatures	1.1%
Agriculture, Natural Resources, and Conservation	0.8%
Education Administration	0.8%
Business Administration and Management	0.8%
Public Administration	0.8%
Religion and Theology	0.8%
Anthropology and Archaeology	0.4%
Psychology	0.4%
Curriculum and Instruction	0.4%
Social Work	0.4%

About 67% of the dissertations and theses used only newspapers from the United States, 23% used only non-U.S. newspapers, and about 10% utilized both U.S. and non-U.S. newspapers.

An analysis of the time periods of newspapers used also provided interesting results (see Figure 1). The newspapers used were largely from the mid- to late-twentieth century (1950-1999) and the beginning of the twenty-first century (2000-2008). Papers from the 1800s were used a fair amount, while eighteenth-century newspapers were very minimally represented, which could possibly indicate lack of access to these resources. Jones (n.d.) discusses the issue of access to newspapers and its impact on research, from the lack of availability of newspapers from certain regions to the overrepresentation of particular newspapers in research because they are available electronically.

Figure 1. Time period of newspapers used.



Almost half (47%) of the dissertations and theses used newspapers from only one time period. Another 42% used newspapers from two time periods. Only about 11% used newspapers from three or more time periods.

In looking more closely at time periods and the disciplines of the researchers, history was the only discipline in the sample that used newspapers from 1700-1749. For the time period 1750-1799, most of the dissertations and theses were in history, with just a couple of documents in performing arts and education. Newspapers from 1800-1849 were mostly used by the history discipline, with a few dissertations and theses in social sciences, performing arts, and English language and literature. Newspapers published in 1850-1899 were also largely used by history scholars, followed by English language and literature researchers, and a few dissertations and theses in political science, social sciences, performing arts, and foreign languages and literature. The largest time period of newspapers used by the history discipline was 1900-1949, followed by 1950-1999 and 1850-1859. Other disciplines that used newspapers from 1900-1949, but to a much lesser extent than history, included performing arts and religion and theology. The heaviest users of newspapers from 1950-1999 were in the communication/journalism and history disciplines, but others included

political science, sociology, and education. In the communication/journalism discipline, the newspapers used were exclusively twentieth century and forward, with the largest number from 2000-2008, and almost none from 1900-1949. Other disciplines, such as political science and sociology, used twentieth century newspapers.

Examination of the subject headings associated with the database records for the documents revealed that most were in American history, journalism, and mass media. Additional subject headings included European history, higher education, American literature, criminology, marketing, and more.

4. Conclusion

It is clear from this study that while newspapers are largely used by scholars in history and communication/journalism, there are researchers in many other disciplines who utilize newspapers. This is important for libraries to recognize as we seek to provide access to information resources needed by our users. Scholars across many academic disciplines utilize newspapers in their research, demonstrating the importance of these resources to a wide range of researchers, and emphasizing the need to ensure better access to newspapers. Understanding more about news researchers and how newspapers are being used across disciplines may create opportunities for new areas of research and interdisciplinary collaboration. Jones (n.d.) comments that "digitization of historic newspapers may lead to new and unexpected user groups." Bingham (2010) notes that "digital archives have suddenly made using newspapers in research a far more attractive proposition, and it is very likely that newspaper evidence will become far more prominent in historical writing." A wider interest in using newspapers for research may result in greater support for digitizing them and for libraries providing access to them.

Additional analysis will include an examination of the origin of newspapers used by discipline and a closer look at the subject headings associated with the dissertations and theses. It would be useful to also understand whether the newspapers used by these researchers were in print or microfilm format or accessible in digital format. Further research could include expanding the time period covered by this research study and analyzing documents beyond dissertations and theses.

Acknowledgement

Special thanks to Niamh Wallace, Special Assistant to the Dean, University of Arizona Libraries, who was the research assistant for this study. The research project was funded by a Support for Outstanding Academic Research (SOAR) Grant, University of Arizona Libraries.

References

Alidousti, S., Khosrowjerdi, M., Shahriari, P., Shirani, F., and Tarnoni, H.B., (2009). Designing a Model for Description of Theses and Dissertations Information on a Large Scale, Libri, Vol. 59, No. 3, 180-197.

Allen, R.B., and Sieczkiewicz, R. (2010) How Historians use Historical Newspapers. Proceedings of the American Society for Information Science and Technology, Vol. 47, No. 1, 1–4.

Allen, R. B., Zhu, W., and Sieczkiewicz, R. (2010). What to Do With a Million Pages of Digitized Historical Newspapers. IConference, Urbana-Champaign IL.

Beck, S. and Manuel, K., (2008). Practical Research Methods for Librarians and Information Professionals, Neal-Schuman, New York.

Bingham, A., (2010). The Digitization of Newspaper Archives: Opportunities and Challenges for Historians, Twentieth Century British History, Vol. 21, No. 2, 225-231.

Deacon, D., (2007). Yesterday's Papers and Today's Technology: Digital Newspaper Archives and 'Push Button' Content Analysis, European Journal of Communication, Vol. 22, No. 1, 5-25.

Freelon, D., (2010). ReCal: Intercoder Reliability Calculation as a Web Service, International Journal of Internet Science, Vol. 5, No. 1, 20-33.

Freelon, D., (2013). ReCal OIR: Ordinal, Interval, and Ratio Intercoder Reliability as a Web Service, International Journal of Internet Science, Vol. 8, No. 1, in press.

Jones, A. (n.d.) The Many Uses of Newspapers. Access date 03/25/2013. Available at: http://dlxs.richmond.edu/d/ddr/docs/papers/usesofnewspapers.pdf.

Krippendorff, K., (2004). Content Analysis: An Introduction to its Methodology, 2nd ed., Sage Publications, Thousand Oaks, CA.

Lombard, M., Snyder-Duch, J., and Bracken, C.C., (2002). Content Analysis in Mass Communication: Assessment and Reporting of Intercoder Reliability, Human Communication Research, Vol. 28, No. 4, 587-604.

Neuendorf, K., (2002). The Content Analysis Guidebook, Sage Publications, Thousand Oaks, CA.

Powell, R., and Connaway, L.S., (2004). Basic Research Methods for Librarians, 4th ed., Libraries Unlimited, Westport, CT.

ProQuest Dissertations and Theses Database. Access date 03/25/2013. Available at: http://www.proquest.com/en-US/catalogs/databases/detail/pqdt.shtml.

ReCal: reliability calculation for the masses. Access date 03/25/2013. Available at: http://dfreelon.org/utils/recalfront/.

Riffe, D., Lacy, S., and Fico, F., (2005). Analyzing Media Messages: Using Quantitative Content Analysis in Research, 2nd ed., Lawrence Erlbaum, Mahwah, NJ.

Salmon, L.M., (1923). The Newspaper and the Historian, Oxford University Press, New York.