Intercoder Reliability Report

Korean, Punjabi and Russian Newspapers 2011

Three newspapers were coded for this project. The Korean Times Daily publishes six days per week in Korean, the Russian Express is a weekly that publishes every Friday in Russian, and the Canadian Punjabi Post publishes six days per week in Punjabi. There were three coders: one coder fluent in Korean, a second coder fluent in Punjabi and a third who was fluent in Russian.

The coders coded 21 issues of the newspapers collected between January 4, 2011, and August 8, 2011, except for Russian Express, which was coded from January 14 to September 9 because we were unable to obtain 6 issues in the main study period. We excluded the weeks between March 17 and May 4 because a federal election took place during this period and the news was unrepresentative of typical content.

The issue of intercoder reliability was addressed in two ways. In the first instance, all coders were trained to code the Toronto Star. Once they reached acceptable levels of intercoder agreement we felt confident that their coding decisions for the respective ethnocultural newspapers would be consistent and accurate. This approach would also allow us, in the future, to analyze and compare data from the ethnic papers with 2011 Toronto Star data, which was subject to exactly the same intercoder reliability testing procedure using the same test issues and two of the same coders.

To further ensure the integrity of the coding, coders of the ethnic papers worked in partners. As they finished coding each issue of the paper, they manually/verbally reviewed their coding decisions. This process involved describing the contents of each local news item to the other coder, discussing the coding decision and making any necessary changes. This quality control measure was possible because most issues contained fewer than 15 local news items. We did this paired testing for all issues.

Coders received 35 hours of coder training for Toronto Star content. After random testing indicated they were coding in a similar/consistent manner with each other they then coded three Toronto Star test issues. The test issues were 35, 41, and 43.

Issue number	Date of publication
35	2/5/2011
41	5/28/2011
43	6/13/2011

Once it was clear that the coders' decision-making vis a vis the Toronto Star was consistent and met acceptable standards of intercoder reliability, the coders moved on to coding their respective ethnic newspapers.

The Toronto Star Test Results

Prior to running intercoder reliability tests, each coder verified the accuracy of the data entered by matching the entries with the coder's notes for each news item. We randomly selected 10 items from each issue, and verified the data for all the elements within those items. If any errors were found within the set, the errors were corrected and another 10 items were selected and checked. Each coder repeated the process until a set of 10 randomly selected items was found to be without data entry error.

We measured intercoder reliability by coding four of the randomly selected sample issues (of the total 21 issues in our data set). There were 365 local items in our Toronto Star test sample out of a total of 1382 local items in the three papers we examined for our study. Thus, given that our test sample represents over 10 per cent of our data, there is no reason to believe that the test sample is not representative of the data set.

Each of the three test issues was coded by three research assistants. After 35 hours of training, each coder coded the issues independently, meeting only once to reach a consensus on the number of local items before proceeding to independently code the item details for local items. All other discrepancies were resolved through discussion after the intercoder reliability tests were completed.

We measured intercoder reliability for interval variables using Lin's concordance, with the aid of PRAM (Program for Reliability Assessment with Multiple Coders) software. Lin's concordance measures the correlation between coder responses, and takes into account systematic coding errors (coder bias), the possible range of responses, the magnitude of difference between coders' responses, and the agreement expected by chance. We considered reliability to be acceptable at or above .700 using Lin's concordance.

We measured intercoder reliability for nominal variables using Cohen's Kappa, with the aid of PRAM (Program for Reliability Assessment with Multiple Coders) software. Cohen's Kappa is a relatively conservative index that measures the extent to which coders make identical coding decisions, and takes into account the agreement expected by chance. Research suggests that for three or more coders, values of between .400 and .750 may be taken to represent fair to good agreement beyond chance (Banerjee, Capozzoli, McSweeney & Sinha, 1999; Fleiss, 1981; Landis & Koch, 1977)

Intercoder reliability was at or above .476 for all variables.

Variable	**Lin's Concordance Test Result**
NoSptlRef	.936

Variable	**Cohen's Kappa Test Result**
ItemOrigin	0.947
ItemForm	0.952
FotoStatus	0.994

GenFotoCont	0.925
Religion	0.960
MinorityCmnty	0.689
GenSubj	0.784
SubjDetail	0.628
R/E1	0.665
R/E2	0.476
R/E3	1.000
R/E4	1.000

REFERENCES

Banerjee, Wayne, Laura McSweeney and Sinha Debajyoti. 1999. Beyond kappa: A review of interrater agreement measures. *The Canadian Journal of Statistics, 27*(1), 3-23.

Fleiss, J.L. 1981. *Statistical methods for rates and proportions* (2nd edition). New York: John Wiley.

Landis, J.R. and Koch, G.G. 1977. The measurement of observer agreement for categorical data. *Biometrics, 33*(1), 159-174.