Text Analysis: Further Work on Computer Assisted Techniques

Bernard L. Dugoni, Ph.D.

NORC at the University of Chicago

Introduction

The purpose of this study was to

1. Examine the usefulness of text analysis techniques developed for the assessment of open-ended survey items when used with more detailed, in depth verbal protocols; and to

2. Assess the ability of these techniques to reproduce or enhance findings with archival data.

To address the first objective, we considered several approaches used in qualitative research projects for the coding of responses to single open-ended survey questions (verbatims) in order to assess their usefulness in the coding of longer open-ended (unstructured) responses (verbal protocols).

The second objective seeks to assess the accuracy or deficiency of the use of these approaches in analyzing longer text responses using newer, automated techniques. To address this objective we used archival data collected, analyzed, and published previously prior to the existence of more current software assisted approaches, to see whether the newer techniques support previous analytic conclusions or point to flaws in the earlier analytics.

Method and Materials

The data for this study were collected by Dugoni and Ilgen (1981). These researchers collected verbal protocols for the construction of "realistic job previews." Respondents to that study were asked to describe situations they had encountered when they had started working their new jobs. In the 1981 study, these were analyzed using traditional approaches such as "Subject" reviewers, the Q-Sort approach, and expert reviewers, and the results summarized in the Journal of Applied Psychology.

In a study reported at JSM last year, Dugoni and Brown (2014) developed a 3-stage approach for analyzing verbatim responses which combines the use of text-analysis software with researcher review techniques for a tailored solution that is computer *assisted* but retains the human element, refined by iterative assessment and psychometric evaluation.

Procedure

A set of 149 verbal protocols from the Dugoni and Ilgen (1981) study were re-analyzed using the WordStat and SimStat text analytic software developed by Provalis Research (Peladeau, 1996, 2014). These software routines allow the exploration and computer assisted classification of text responses into categories developed through the evaluation of word frequency counts and the use of co-occurrence matrices to develop keyword phrases and facilitate keyword-in-context analysis.

Using iterative runs and review of these descriptive computations, categories were developed and compared to those developed in the Dugoni and Ilgen (1981) study. This was combined with the approach developed by Dugoni and Brown (2014) to assess results in three stages. At the first stage, word frequencies were computed by the Provalis software and the keywords with the highest frequencies were compared with those observed in the Dugoni and Ilgen Study.

At the second stage, expert review of the keywords as well as co-occurrence data were combined to generate a list of keyword-in-context combinations. These were compared with the groupings and categorical combinations from the Dugoni and Ilgen (1981) study and overlap as well as additional information (new categorizations not in the original study) were assessed.

Finally, at the third stage, statistical comparisons were made between the sets of classifications and further examinations were made to "unpack" the situations where the initial stages alone were unable to reproduce classifications of the original study.

Results and Discussion

Stage 1: Initial Categories based on highest word frequencies fell into two groups, those consistent with the original study and those that were not. Of the latter, there was only one group that constituted more than 10% of the responses. These groupings are shown in Table 1.

Table 1. Stage 1 Keywords

Consistent with original study	
Keyword	Proportion
BOSSES	0.16
JOB	0.19
MANAGER	0.16
PEOPLE	0.30
BOSSES	0.16
New Additional subset	
FEEL	0.18

The most frequently occurring words identified by the program overlapped quite well with those identified originally by the manual reviews done in 1981. The main exception involved occurrences of the word "feel." Possible explanations for this are explored in the following sections.

Stage 2: In the second stage, the co-occurrence of these keywords among themselves or with other words was examined. This evaluation developed the keyword-in-context combinations shown in Table 2. Again most of these overlapped nicely with the review done for the original study. The context specific information revealed in examining the co-occurrence patterns led to the classification of the last category as "Positive/Negative Feelings because the keyword "Feel" identified in stage 1 seems to indicate situations where the respondent is talking about how an experience he or she was describing made them feel. This categorization did not appear in the original study, hence the empty cell.

Table 2: Keyword in Context		
	Computer Assisted	Original Study
Category	Proportion	
Supervision: BOSSES/Manager	0.41	0.42
Work itself / tasks	0.15	0.39
Co-Workers	0.05	0.09
Customers	0.07	0.07
Recognition for work	0.04	0.02
Other	0.01	0.01
Positive / Negative feelings	0.18	-----

Stage 3: In stage 3, we conducted statistical tests of the findings at the previous stages and probed further into the lack of overlap. Table 3 shows z-tests for differences in proportions between the proportions observed for each classification with the computer assisted approach compared to that observed in the original study.

Table 3. Differences in proportions of categories from Table 2.		
Category	z	Sig.
1. Supervision	-0.07738	n.s.
2. The Work Itself	-1.48185	n.s.
3. Coworkers	0.60357	n.s.
4. Customers	-0.0913	n.s.
5. Recognition	0.310728	n.s.
6. other	0.116709	n.s.

None of these tests showed a significant difference between proportions indicating that despite the appearance of a lower occurrence of some keywords-in-context, the two evaluations seem substantially comparable. As a further assessment of this set of observations, a Chi-squared Goodness of Fit test was conducted to assess whether the observed classifications varied from that which would be expected using the original study proportions to compute expected frequencies. The result was Chi-squared=0.206352, n.s., indicating no significant deviation from fit.

Despite the fact that there were no significant differences between the original classifications and the newer computer assisted approach for the majority of categorizations, there is still the matter of the additional classification category not present in the earlier study and the fact that not all classifications reproduced exactly what was done before. In examining this further, the following observations were made with regard to confusions that were not straightforward for a computer driven classification to navigate.

Some confusions related to several categories originally grouped under “People” Here, there were a number of instances that talked about a “Manager” or “Boss” as a person vs. other text that made it clear that the concept of supervision was the key focus. Resolving this confusion is relatively straightforward, but others are more involved.

Differentiating positive vs. negative did not play a role in the earlier paper but even though included here it is not straightforward. For example, a clear interpretation of a verbal protocol often involves “unpacking” a respondent’s thought process which may be

somewhat convoluted or less than well focused. Here is an example. "I caught someone cashing a stolen check and held onto it I was thanked and given a five dollar reward by the management for saving the store money by not cashing the check. It started a new reward system for people who catch stolen checks." In this example, the respondent was actually relating an experience they found to be positive because they were rewarded and the event led to a positive outcome. However, because many of the words are negative, the correct classification required more review.

In other instances, confusions can result from a respondent's less than perfect command of grammar and syntax. Consider the following verbatim, "I am caught customer committing a crime and the people where I work said I was really doing my job." Again, this response was grouped with other negative responses, but the response actually appears to be relating a positive event.

Still other responses that are difficult to code and hence seem to wind up in an "other" category often involve jargon. For example, "The E.S.I.S. registers had to be coded on certain items PLU and a lot of the customers were unable to know the prices."

While in general, the computer driven tools such as described here seem to be worthwhile aids to assist coding and evaluation of text data, care must be taken in order to avoid confusions or missed categorizations. As with much of the natural language programming attempts in artificial intelligence work, the eccentricity and nuance of human language often require meticulous human review to derive accurate sense.

Future Directions

Drawing on the observations of our 2014 and 2015 studies, we plan further work to develop an approach for computer guided conduct of the sorting techniques such as q-sort, attempting to disambiguate confusing classifications, and to interface with approaches such as Delphi to provide a technique which draws on valuable new approaches without losing valuable past developments.

References

Dugoni, B.L., and D. R. Ilgen. 1981. Realistic job previews and the adjustment of new employees. *Journal of Applied Psychology,* 24: 579-591.

Dugoni, B.L., and Brown, K. 2014. Code-frame Development for Verbatim Responses Using a 3-stage Approach. Proceedings of the 2014 Joint Statistical Meetings, 403, Boston, MA.

Peladeau, N. 1996. SIMSTAT for Windows. Montreal: Provalis Research.