Qualitative Analysis of Oral Testimony Information Retrieval and Data Management of Electronic Transcripts from the Royal Commission on Aboriginal Peoples

Mary Jane Commanda
National Archives of Canada
and
Seymour Hamilton
Colophon Communications

Résumé: *Cet article rend compte de la méthode originale ainsi que de l'analyse qualitative élaborée nécessaire pour traiter les transcriptions des plus de 2 000 présentations orales faites devant la Commission Royale sur les peuples autochtones. À l'aide du logiciel Folio Views, on peut non seulement vérifier la fréquence avec laquelle certains sujets ont été abordés par divers groupes d'intervenants, mais aussi saisir qualitativement les perspectives communes de ces groupes. L'article explique la dimension interculturelle du matériel et décrit les instruments novateurs employés pour préserver les modes de discours des intervenants devant la Commission. Les auteurs abordent la question de l'inductivité dans l'analyse du contenu, qui est maintenant rendue possible grâce aux fichiers numérisés complets. La spécification explicite des présuppositions dans la catégorisation du matériel soumis, jumelée à l'analyse informatique, permet de réduire la possibilité de rendre des jugements de valeur importuns. L'article conclut en explorant des possibilités d'application dans d'autres contextes.*

The Royal Commission on Aboriginal Peoples

The Royal Commission on Aboriginal Peoples (RCAP) was established in 1991, and scheduled to submit a final report to the Government of Canada in 1995. The Commission was mandated to investigate and propose solutions to problems in the relationship among Aboriginal Peoples (First Nation, Inuit, Métis,) the Canadian government, and Canadian society as a whole. In practice, this included a broad range of issues including: the scope of Sec. 91(24) of the Constitution Act, the legal implications and status of Aboriginal treaties and modern-day agreements, the Indian Act, land and resources, economics, social and cultural issues, justice, education and health. The mandate placed emphasis on the Métis, off-reserve and Inuit peoples, including women, youth and the role of elders, as well as the

situation in the North. Co-chair Georges Erasmus noted that the mandate of the Royal Commission "covers virtually every issue of some importance to all Aboriginal people in Canada."[1]

Over three years, the Commissioners held four rounds of public hearings in Aboriginal communities, as well as round table discussions and special consultations. The Commission was particularly careful to include a balanced representation of views by First Nation, Métis, off- reserve, and Inuit peoples, as well as non-Aboriginal people. In addition, the Commissioners sought input from federal, provincial, territorial and municipal governments, as well as the views of individuals and organizations with a particular interest in Aboriginal issues.

The RCAP established an elaborate electronic and textual information system that includes more than 80,000 pages of verbatim transcripts, 700 summaries of written briefs, 120 intervenor participation projects, written summaries of past commissioned reports and studies, and almost 400 research contracts, projects and community studies relating to all aspects of the Commission's mandate. This paper deals only with the transcripts.

The entire collection was managed by the Commission's Information Management Unit (IMU). The information was stored, indexed, integrated and manipulated using a whole-text retrieval system, Folio Views 3.0. Folio is user-friendly software that allows novices to work productively with minimal training, while providing sophisticated users with ample search and manipulative power. The central feature of Folio is that it automatically indexes every word in the "infobase" or collection, making it possible to search for individual words or words in groups using Boolean ('and'/'or'/'not') logic. Folio can be used to code or "tag" the same material into any number of categories without affecting the basic text. It can also be used to group material that has been temporarily or permanently identified by whatever criteria (such as searches by word, word cluster, speaker demographics, etc.). This makes it possible, for example, to isolate in a separate sub-collection all the material that deals with a specific subject — such as governance or the natural environment.

The system allowed the Commissioners and their research and policy teams to access the material *as it was being gathered*. The information base became the steadily-increasing corporate memory of the Commission, and consequently was the information foundation on which policy options and recommendations were based.

Although this paper[2] is concerned only with the hearings transcripts, readers should know that Folio's "link" function (which takes the user automatically the reader to a pre-programmed destination elsewhere in the collection) integrated the different information subdivisions, helping to achieve a holistic view of all the different sources of fact and opinion under the Commission's mandate.

The staff of the IMU developed innovative approaches to cope with a variety of users, including research and policy teams, Aboriginal and non-Aboriginal communities, governments, as well as students, scholars, academics and national and international media people. A paramount concern in this process was that nothing should compromise the unique nature of the information, particularly in the case of the oral interventions.

The two approaches addressed in this essay involved:

1. documenting the cultural modes of discourse (occasions when the statements were charged with meaning not immediately apparent to a non-Aboriginal person); and
2. preparing a series of "Who, What, Where" demography-based papers.

Together, these two approaches relieved RCAP staff and Commissioners from having to rely on impressions based on a partial hearing (or reading) of the interventions. Using these approaches, it became possible to make valid generalizations about "issue ownership," for example, to discover that there are significant differences between male and female views about Aboriginal self-government — a realization that could only come from demographic examinations of the transcripts.

The information base documents the frustrations, hopes, and dreams of Aboriginal people who responded to the Commission's call for information and opinion. The collection provides the material by which it is possible to understand the Aboriginal world view in its wide-ranging scope and variety. The hearings in particular demonstrate the oral traditions preserved by elders, families and communities, and thereby show how the Aboriginal cultures and languages are essential to self-identity.

The Commission's collection will be an important resource for people from all disciplines and backgrounds who are interested in Canadian society. The CD-ROM version of the Commission's files, together with all their enhancements, will be available at the close of the Commission's activities.

Limitations of the Collection

Though impressive in its scope, the RCAP collection has its limitations.

While the views expressed in these hearings undoubtedly represent the spectrum of opinion among and about Aboriginal Peoples, the speakers are not necessarily statistically representative of the constituencies they belong to or represent. The sample of more than 2,000 intervenors was not chosen randomly. However, the Commissioners chose the communities they visited so that they would hear a balance among the various points of view; and once in the communities, the Commissioners continued each session until everyone who wished to speak had the opportunity to do so. Each successive round of hearings was analyzed in an Intervenor Characteristic

Report, based on the affiliations stated by the intervenors themselves. This information allowed the session organizers to redress any perceived imbalances in the next session. During the third and final rounds the time was increasingly taken by invited intervenors representing organizations.

There are several factors which must be taken into consideration when using the RCAP collection. First, the Intervenor Characteristic Reports themselves have limitations. Certainty is possible with respect to Aboriginality, gender and whether the intervenor spoke as an individual or an official spokesperson for an organization. Certainty is possible with respect to intervenors who spoke for an organization, because they stated, and/or were introduced by their position or title (chief, member of parliament, organization president, etc.). Reasonable certainty of at least 95% confidence applies to Aboriginal ancestry in terms of Inuit, First Nation and Métis, and among demographics such as youth, elder, veteran. The distinctions between status and non-status Indians, or among rural, urban and remote are not as certain because it was possible to count only those who explicitly identified themselves by these categories; that is, there may well be others who fit those categories who did not self-identify.

Second, the multilingual environment of the hearings challenged the resources of those who transcribed the tapes to a written form. Approximately 10% of the intervenors at the hearings spoke in their native Aboriginal languages, either in whole or in part. Some key phrases were recorded in their original languages. However, the integrity of longer statements in languages other than English or French is somewhat compromised by the fact that the translations were made by a different person at each session, some of which went on for many hours.

Consequently, the Commission's Information Management Unit had to find a balance between consistency in matters such as spelling and transcription of Aboriginal words, and the need to be scrupulously faithful to the text. The methodology and the innovations added to the Folio software were designed to preserve the essence and spirit of the contributions.

These caveats do not affect the validity of information gleaned by the Commission; however, they do limit the degree to which statistical generalizations can be made about the opinions of definable groups. Comparisons within the universe of statements to the Commission is possible and valid; projections to the general public must be made with caution. For instance, it is possible to make valid and useful comparisons between what Aboriginal women said when speaking for themselves, and what was said about Aboriginal women by leaders, politicians and both Aboriginal and Non-Aboriginal organizations; however, the numerical validity of these statements is limited to those who spoke to the Commission. Similarly, because a great many intervenors offered a list of issues facing Aboriginal people, even though they

spoke in detail about only one or two of those issues, it is not valid to use Folio to count instances of words or phrases — unless the material is read and checked by the protocols suggested in this paper.

For all these reasons, any projections from the hearings to the entire population should be made with caution.

The Methodological Challenge

First was the problem of managing a large collection of information. More than 2,000 interventions were made to the RCAP over the course of four rounds of hearings over nearly three years. (An intervention is defined as any statement at the oral hearings by an individual or group of individuals.) The testimony offered at the RCAP hearings totalled more than 40 megabytes even when electronically compressed by the Folio software. Another 60 megabytes of information flowed from the Commission's other sources, including research papers, solicited and unsolicited briefs, published research and statistics.

Second, the collection and access system had to be designed and in operation *before* all the evidence was accumulated. Pre-set, rigid categories necessary to traditional data bases or ethnographic software were not appropriate for so wide a mandate.

Third, the collection had to be accessible for the Commission's purposes. The Commission was mandated to propose "specific solutions, rooted in domestic and international experience, to the problems which have plagued those relationships and which confront Aboriginal Peoples today."[3] Accordingly, a flexible, non-judgmental, unbiased access system was necessary, even while the material was still being collected.

Fourth, the system had to be culturally sensitive. The intervenors (the people who spoke to the Commission) were mostly Aboriginal individuals and organizations with a wide variety of backgrounds and linguistic preferences. The oral interventions were heard in many languages: English, French, and several Aboriginal languages including Inuktitut, Cree, Ojibway and Chipewayan. Even when the intervenors used English in their presentations or submissions, they spoke in the context of their cultures and backgrounds.

One intervenor, Don Sax of Old Crow in the Yukon, captured the essence of the situation when he called language "a way of perceiving reality."[4] His thought was not unique. Other intervenors spoke of the differences among languages in terms of expressing a world view through words such as "a mind set"[5] or "a way of thinking."[6] Marguerite Cardin spoke of how close thought and behaviour can be: "Being Indian . . . is a way of thinking, a way of life — it is a philosophy."[7] As well as the diverse nature of Aboriginal cultures represented in the oral testimony, the infobase also included

testimony from lawyers, bureaucrats, academics and researchers. Each of these disciplines also has its own way of perceiving and expressing reality, and each has a characteristic mode of discourse.

Fifth, it was necessary to devise a neutral access system that did not rely on European-origin words and ideas. The Commissioners asked for an accessible record that did not filter, edit, paraphrase or re-state the words of the people and organizations who came to the public consultation process. At the same time, it was necessary to establish a thesaurus (discussed later as the WordFinder) that would offer a consistent vocabulary to people using the infobase so that they would not miss material because of lack of a shared vocabulary. For example, a phrase such as "insured health benefits" that would be clear to academics, researchers and public service employees, was rarely used by intervenors who dealt with this topic.

Finally, the system had to be instantly responsive to unpredictable questions. The Commissioners often wanted to know the answer to questions such as, "Who said what, and how many of them said it?"

How the IMU Used Folio to Meet These Requirements

Building on the strengths of the Folio software, the IMU prepared purpose-designed features that made the transcripts more accessible to report writers and policy analysts. These in-house users had expertise in aspects of Aboriginal matters, and many were Aboriginal people themselves who had personal experience of the issues under study; however, they were not a homogeneous group. From constitutional lawyers to social policy experts, each brought distinct linguistic preferences.

Some features were clearly necessary, such as collecting as much demographic information as possible on each intervenor or intervening organization. This information was searchable in the same way as the text itself, making it possible to list all the files of "elders" or "youth" or "non-Aboriginal organizations," and then to search those files for the incidence of topics such as "education" or "language."

The information had to be cross-culturally accessible *among* Aboriginal Peoples, recognizing and acknowledging not only Aboriginal diversity, but also all viewpoints. Each perspective had to be balanced and taken into consideration. The Commission was mandated to restructure relationships between Aboriginal and non-Aboriginals in Canadian society. It was therefore important that Aboriginal perspectives not be lost in the overall process of policy formulation.

On the one hand, Commissioners and policy experts wanted to know what groups of people — for example, the Métis — were saying; on the other, they also needed to know that the Métis are by no means unanimous in their concerns and interests. The same was true of other major groups.

For example, there were differences in the assumptions and concerns of the First Nation leaders (most of them male) and Aboriginal women both as individuals and in the context of their organizations.

For all these reasons, it was necessary to compile a project thesaurus, which became known as the "WordFinder." This thesaurus related the words of the intervenors to the Commission's mandate terms, and to the vocabulary of academic, governmental and other professional groups. For example, the WordFinder related the legal terms "in perpetuity" to the same concept expressed in the oral tradition — "As long as the grass grows and the rivers flow."

Because much of the testimony was delivered orally by people whose mother tongue was neither French nor English, and most of it was offered by people whose cultural background was not European, the testimony had to be read and tagged for occasions when people spoke on a topic without explicitly identifying it according to the European, bureaucratic, print-oriented conventions necessary to the writing of reports. Similarly, in some testimony, the intervenors used the language of spirituality or poetry to make statements "from the heart." These poignant evidences of concern had to be identified as belonging to a different mode of discourse from everyday language.

The Tagging Process

The data analysts read more than 80,000 pages of verbatim transcripts, and "tagged" issue groups, using an analytical framework based on the Commission's terms of reference. For example, if an intervenor discussed "band leadership," the excerpt was tagged as *issue: governance, trust relationship, relations federal/aboriginal relations and devolution.* Similarly, if an intervenor discussed the importance of "living off the land", or "living in the bush", the excerpt or presentation was tagged *issue: traditional economy, land and economy, land and resources, and traditional culture.* Tagging the same material into different issue groups ensured that the material would be considered in all relevant policy contexts. The analysts also tagged recommendations, solutions, strategies, models, pilot projects and suggestions, and in some cases, included dialogues between intervenors and the Commissioners.

Instances when intervenors used the heightened language of spirituality or of poetry to speak "from the heart" were tagged as a change in the mode of discourse. These were in the main occasions when the intervenors used concepts or symbolism to explain the interconnectedness of traditional spirituality, ceremonies, or the sacredness of Mother Earth. In some cases, where holistic world views were embodied and alluded to throughout, it was necessary to tag whole presentations.

The work required an in-depth knowledge of Aboriginal issues in both the historical and contemporary context, since many of the intervenors spoke about historical and political developments in federal/Aboriginal relations, constitutional and legislative processes, long-standing disputes over treaty rights, claims, resources, the result of colonization and other historical processes. The analysts all had a knowledge of research and policy needs, as well as an understanding of Aboriginal culture. The IMU team consulted together, making use of their experience and intuitive knowledge of Aboriginal concepts as expressed by elders and people with traditional indigenous knowledge.

Their job was to link the content of such passages to the Commission's themes and concerns for report- and policy-writers, some of whom might lack that special knowledge necessary to access the testimony. This work also made explicit links between the testimony and the equally voluminous collection of academic, legal, bureaucratic and community research papers, which were largely written according to European linguistic conventions.

Initially devised to make searching easier for the policy analysts, the features improved the strength of the searching process by making it specifically responsive to the oral testimony, the needs of the users and the nature of the information. These enhancements in no way affected the text itself. Any part of any intervention could be examined exactly as it was transcribed, with no identifying marks, changes or comments. The enhancements were all transparent; that is, there were explicit techniques by which the WordFinder was compiled, and for the criteria and methodology by which the mode of discourse changes were tagged. The links and tags empirically reflect recorded linguistic phenomena—they do not evaluate the content or opinions being expressed.

Together, the WordFinder or project thesaurus, the tagged groups and the identification of different modes of discourse expanded searches of the transcripts by approximately 20%. Added to a standard word search, the IMU's enhancements captured ideas, notions or concepts that might otherwise have been missed.

Analysis of Qualitative Material by Quantitative Means

Because they held simultaneous hearings in three teams, none of the Commissioners or Commission staff were present at all the hearings. The sheer bulk of the entire collection of transcripts prohibited a complete reading. The Commissioners therefore needed some objective way of checking their impressions about the overall flow of the testimony. Knowing that one poignant, well-expressed story might eclipse interventions that were equally important but less memorable, they essentially wanted to quantify the qualitative question, "What did we hear?" The answer was a series of "Who, What, Where?" papers.

The first step in preparing these papers was relatively simple: use the enhanced query system to collect sub-sets of the interventions which substantively discussed the Commission's mandated themes of investigation. The IMU team created a series of query statements to search for overlapping concepts that together added up to subject areas such as "Youth", "Urban Aboriginal", "Governance", and so on. Folio Views shows "records with hits", along with the *number* of hits, which allowed researchers to distinguish between a solitary use of the word "Youth" in a list of problems, and substantive considerations of the subject.[8]

An important feature of this approach is deliberate double-counting. Virtually all speakers addressed their subjects in a holistic manner, frequently touching on several of the Commission's mandated topics. Accordingly, each presentation was counted in several different categories. The resultant numbers have no absolute value, but may be used to compared among each other either as raw numbers or expressed in terms of percent of total interventions.

The next step was to examine each information sub-set, asking "When people talked about youth, in what *context* did they speak?" Essentially, this process was experimental. The IMU team compiled a list of possible contexts which, in the case of Youth, included language, education, culture, recreation, elders and so on. Then queries were framed for each of these contexts using the same techniques as in creating the Youth sub-set. Then each query was applied to the sub-set and the resultant hits winnowed, following the same rule as before: If three or fewer hits, read and (usually) discard. This process produced a total number of substantive presentations that discussed youth in each context. From these numbers, it was possible to rank in order the contextual material that was mentioned most, so to speak, "in the same breath."

This empirical methodology identified and counted instances of ideas (both singly and in groups), as opposed to words. The internal papers using this methodology therefore were a quantitative analysis *from the point of view and with the contextual weighting of the intervenors*.

Summary of the Methodology

1. The researchers first *identified* all the occasions when a subject was mentioned, and then reduced xFolio "hits" to substantive considerations of the topic as opposed to a passing mention of the key words.
2. The researchers examined and collated the *characteristics* of each intervenor who mentioned the subject.
3. The researchers identified the *contexts* in which the subject was examined by the intervenors. Essentially, this involved asking the question

"What else did the intervenors talk about?" and assessing the ten most frequently mentioned sub-topics.
4. The researchers examined *recommendations suggestions* in the interventions that considered the topic substantively, and ordered the contextual sub-topics by frequency. They then provided a summary overview of the recommendations.

Example: "Youth":
Youth was one of the Commission's designated concerns. It was clear that people talked about youth in the context of topics that included education, suicide, substance abuse and health, among others. What was not clear was the relative importance of these different contexts. The Youth paper therefore investigated how the intervenors understood these linkages and connections, supplanted the impressions and intuitions of those who attended some of the sessions or read parts of the transcripts.

Intervenors addressing the subject of youth were *most* likely to talk or write about youth in the context of education, culture and justice. Of the intervenors dealing with the topic of youth, 94% spoke or wrote in the context of education, 74% in the context of culture and 68% in the context of justice. Some of the information was counterintuitive. One might have imagined that any testimony dealing with youth would have recreation as one of the leading contexts. In fact, recreation was in a distant tenth place, mentioned by only 18% of those who spoke on the subject of youth.

The paper also analyzed the characteristics of the intervenors who held these opinions. Four out of five youth spoke about youth. Approximately two out of five elders spoke on youth, and the same proportion was true of non-Aboriginals. All youth organizations addressed the subject; however, somewhat less predictably, more than three-quarters (77%) of women's organizations were specifically concerned about youth. Roughly half the women who spoke or wrote to the Commission addressed the subject of youth, but only one third of the men. There were no significant differences when the material was considered geographically by provinces and territories.

The Consequences of this Approach

The foregoing information was available by no other means. It provided corroboration and correction to intuitions formed by Commissioners and analysts alike, and served to separate individualistic approaches from the overall trend of the interventions. In this way it provided highly-focussed experts (such as constitutional lawyers) with a wider context, and helped corroborate the experiential knowledge of Aboriginal policy- and report-writers by offering the evidence of people who shared the same background. The approach also helped distinguish between widely-held attitudes and the individualistic (and therefore sometimes more memorable) approaches

which could then be evaluated by policy analysts. Frequency of mention could be taken as a guide to the intervenors' priorities, and hence was at least an indication of opinions held by the constituencies to which they belonged.

The papers showed whether a subject was "owned" by a demographic group definable in terms such as sex, age, geographic location, ethnic group, organization and so on. Conversely, the papers also identified consensus among people from many different groups, locations and backgrounds. Counterintuitive discoveries such as the ranking of recreation by intervenors on youth tested research hypotheses about public perceptions.

Policy Implications

The approach used at the RCAP can be used to gain full value from any public participation process in at least three related ways:

1. Since public participation is rarely conducted on the basis of random surveys, process organizers can use the "Intervenor Characteristics" technique to ensure that there is balanced representation of different demographic and opinion groups.
2. Policy-makers can use the demographic "Who, What, Where" approach to gain an objective understanding of the opinions expressed by the intervenors. The statistical approach ensures that policy-makers discover what the most generally held opinions are, as distinct from those opinions that are most persuasively expressed, or those which accord with accepted theory.

   It is advantageous for policy-makers to know the demographics of people who share opinions. Policy-makers may build their policies on the opinion of one expert rather than on the convictions of many lay persons, and indeed this may be the best thing to do. However, policy-makers need to know when they are deliberately swimming against the tide of public opinion, even if it is only that of relatively small groups — especially when these opinions are held by articulate people who contribute to public process exercises.
3. Policy-makers can be helped by demographic analysis of opinions expressed in a public participation process. The "Who, What, Where" papers can distinguish between what a segment of society (youth, women, the aged, etc.,) say about themselves, and what is said about them. The results of this process can run counter to accepted dogma about the sub-groups of society and provide fresh insights that can subsequently be tested by more specific enquiry techniques (such as focus groups, interviews or studies). On occasion, demographic analysis can point to the need for factual inquiries into the special needs of

segments of society that are invisible when they are aggregated in larger, more conventional terms. For example, the needs of Aboriginal veterans could be overlooked if they were aggregated into larger and more conventional groupings such as "Indians" or "veterans."

With few exceptions, the contribution of most public participation exercises to policy-making is at best intuitive. When the author is an individual with specific training in listening to and analyzing verbal testimony, as was Mr. Justice Berger, Commissioner of the Mackenzie Valley Pipeline Inquiry. His final report, "Northern Frontier, Northern Homeland,"[9] reflects the consensus of those who spoke to him. However, it is essential to have an objective process for analyzing what the intervenors say when a commission of inquiry is faced with a broad mandate, a great many intervenors and a public participation process conducted by teams of chairpersons from different backgrounds. Otherwise, the public participation is reduced to an opportunity for the public to ventilate, and those charged with the task of listening can be forgiven for selectively remembering only opinions that corroborated (or conflicted) with their own. Without an objective, transparent means of analyzing the testimony, the transcripts of the public participation process can become merely a source of quotations to bolster opinions that existed before the entire, costly public participation process was even begun. Under such conditions, the public may feel bitter that public money was spent to allow them to speak, but that nobody listened to what they said.

In the past, Aboriginal Canadians have been studied to death, but have not been heard. Researchers and commissions have listened through their biases, and written reports that confirmed their pre- judgments. By examining verbatim testimony that has been made accessible through a transparent electronic technology, it is possible to hear what is being said by the people. Accountability comes from having such a record available not only to the policy-makers, but also to the people themselves.

Theory-Based vs. Data-Based Analysis

The most important distinction between this approach and traditional analysis through data bases lies in the difference between theory-based and data-based conceptual analysis. Theory-based analysis examines data empirically through pre-existing concepts. (Creating a traditional data base must start with choosing and defining categories.) Data-based analysis proceeds by first sampling the data and then examining frequently-occurring instances. Theory-based analysis is *a priori +*; that is, it is driven by assumptions that pre-exist the analysis. Theory-based analysis can be compared to a fishing net with mesh sized to the fish that early in the project seem to be desirable. However, it usually turns out that you also need the

small but important fish which are not caught by the mesh of your chosen categories.

Data-based analysis is *a postiori*; that is, it is driven by the data itself. Thanks to the Folio software, it is not necessary to create permanent categories (the "mesh" in the fishing analogy). Instead, it is possible to use Folio to look for concepts as and when they emerge in the Commission's expanding agendas. Until the advent of software with Folio's powerful indexing capability, *a priori* analysis by traditional data base was the only way to approach massive quantities of data. The sophisticated indexing and searching capability of the Folio software make it possible to create information collections that are accessible today, tomorrow and next year by whoever wishes to exercise curiosity. The information collection need no longer be a problem-specific data base, of interest only to the specialists who constructed it. Moreover, the success of the IMU in upgrading the RCAP's collection from Folio 2.1 to 3.0 are a positive indication that the Folio system will enjoy considerable longevity.

Limitations of this Approach

Strictly speaking, this approach is not analysis in the sense of "policy analysis", which usually leads to a specific interpretation and/or an optimum approach to the problem. Rather, this approach is an empirical and statistical contribution to the policy-formation process. It is analogous to a public opinion poll in that it offers a statistical statement of how frequently people address a subject. Like a poll, it records but does not evaluate the positions people take. It only notes *who* spoke, *what kind of* people hold the opinions, and *where* they are speaking and writing. Unlike a poll, it does not "lead" the participants with questions, nor are the respondents necessarily a structured or representative sample of their population. The data-based papers represent an exercise in concept testing — they ascertain whether and how subjects are treated by the intervenors.

Unlike a poll, this approach considers the *context* of other subjects that the intervenors relate to the subject under study. For instance, it can quantify the connections people made between suicide and youth or substance abuse. This examination of context gives a sense of how intervenors think about the subject. The frequency with which they address problems related to youth can be a guide to their priorities and concerns.

The "Who, What, Where" papers used quantitative methods to arrive at a statement of what was on the minds of most intervenors — a qualitative result. The papers quantified *in relative terms* how important certain topics and sub-topics were to the intervenors *considered as a group*. The papers did not attempt to capture "the best" or "the most insightful" statements, and they did not claim absolute accuracy for their generalizations. The

numbers offered by these papers should be viewed as "soft but significant". For example with respect to youth, statistics should be interpreted at the level of statements such as: "Approximately two out of five of all the intervenors mentioned youth," or "When intervenors made recommendations with respect to youth, roughly three out of four such intervenors (74%) linked the subject of youth to culture, whereas fewer than one in twenty (18%) linked it to recreation." The *relative* importance to the intervenors of culture and recreation is therefore strongly defensible, but to continue on to apply statistical tests would be inappropriate both in terms of the methodology and sample.

Throughout the "Who, What, Where" process, interpretation and judgment were minimized and potential for bias was admitted and minimized. At least one Aboriginal person double-checked every step. The Folio technology allowed the data to be mechanically disaggregated; that is, the clusters of ideas within the individual testimonies were broken down into their components and counted. This distanced the researchers from the material, reducing the likelihood of bias.

Possible Applications of this Methodology

Surveys and questionnaires usually conclude with an open-ended question inviting the respondent to write or speak unprompted by questions whose responses are recorded by some variant of checking off boxes with a yes/no or preference. In practice, these comments are usually given an intuitive and frequently cursory reading, and perhaps are featured in an addendum to the structured part of the questionnaire — which tends to be regarded as the "real" questionnaire. The transcripts of open-ended interviews, round tables or public hearings are usually examined unsystematically.

Until recently, there was no scientific way to approach this kind of testimony, especially when it is voluminous, unstructured and offered in a variety of linguistic styles and approaches. Folio Views is sold by its designers primarily as an electronic publishing system that can produce CD-ROM products characterized by this medium's ability to store and access very large quantities of information. Its search technology is designed to be used to find the least number and most specific of hits. The system can support queries from researchers who need to find and count *collections* of hits rather than individual instances.

Cross-cultural communication is increasingly important both in Canada and worldwide. The RCAP experience demonstrates that it is possible to collect information so that it is accessible to people from different cultural backgrounds, with none of them feeling that the collection has been skewed or biased by the conceptual background of the system's designers.

For people involved in commissions of enquiry, which characteristically receive large quantities of information in "plain language" (frequently in several languages and cultural traditions), the searching and analyzing techniques used at the RCAP can be invaluable:

1. during the gathering process to provide course-corrections so that those organizing the hearings can make sure that the process is fair and representative;
2. during the policy-formation stage to provide concept testing of hypotheses and tentative solutions against "what the people said"; and
3. during the writing process to provide instant access to quotable material with confidence that the quotes will be accurate and demonstrably representative.

For people in the communications and marketing fields, the whole-text system allows researchers to discover consensus among the members of focus groups, Delphi studies and the like, *without* either forcing the discussions down prescribed directions, or sacrificing access to individual "significant" statements. Similarly, academic surveys and studies can use transcripts of open-ended questionnaires and interviews to test hypotheses before launching on multivariate analysis of a specific research question. This could reduce the number of projects yielding null or ambiguous results.

A Note on the Folio Software

Because software has a relatively short life expectancy, there are two important measures of excellence: 1) its compatibility with other systems, and 2) its capability for upgrading. Folio passes both tests. Folio is supported by MS-DOS, Windows and Macintosh; that is, a specific collection can be run on all three systems. Moreover, text from Folio can be up- and downloaded easily from word-processing systems such as Microsoft Word and WordPerfect, and Folio can be linked to other data base systems such as Lotus. Folio Corporation also has a good track record with respect to upgrading: material originally composed on older versions of Folio can be transferred to newer versions with a minimum of difficulty. The power of the indexing and searching capability in the Folio software makes it likely that information collections created today will be useful tomorrow and next year by whoever wishes to examine the material.[10]

Considerations for the Future

The process described in this paper needs to be explicitly documented in detail. This will not only make the process more easily replicable by similar projects, it will also ensure the accountability of the RCAP collection.

Although the Commission did as much as it could to capture Aboriginal voices, ideally, an information system that is holistic and reflective of Aboriginal world views could make use of photographs, sound and visual recordings other than print media.

The electronic transcript of each round of hearings was made available to key Aboriginal organizations and regional libraries throughout Canada. The RCAP experience showed that when people know what their fellow-citizens have said to a commission, they begin to understand the extent and complexity of their situation. Simplistic solutions are voluntarily discarded, and people look for accommodation of different views. Sadly, the transcripts reached only relatively few people, and there was little time for reflection on what had been said. In the future, a policy exercise of this kind could use electronic transcripts more vigorously to involve more people at a greater depth of understanding.

NOTES

1. Georges Erasmus, Lethbridge, Alberta, RCAP Transcripts, May 24, 1993.
2. The authors wish to acknowledge the invaluable assistance of everyone in the Information Management Unit at the Royal Commission on Aboriginal Peoples, especially Karen Ginsberg (Director) and Gail Bradshaw. Special thanks to Bruce Wilson (National Archives of Canada) and Katherine Fletcher (Colophon Communications).
3. Right Honourable Brian Dickson, former Chief Justice of Canada. Opening statements on the occasion of the launch of the public hearings of the Royal Commission on Aboriginal Peoples (RCAP), October, 1991.
4. Don Sax, Old Crow, Yukon, RCAP Transcripts, Nov. 17, 1992.
5. Marinus Dieleman, Timmins, Ontario, RCAP Transcripts, Nov. 5, 1992.
6. Rosemarie Kuptana, Ottawa, Ontario, RCAP Transcripts, Nov. 3, 1993.
7. Marguerite Cardin, Montreal, Quebec, RCAP Transcripts, May 26, 1993.
8. This process followed a protocol: if a Folio search turns up fewer than three "hits" in an intervention — for example, instances of the word "youth" — the researcher read that intervention to see if it was substantive, or merely a tangential use of the search words. As always, this process only discovered instances when ideas were raised by intervenors; it in no way evaluated the speakers' opinions on those ideas.
9. Berger Report, *Northern Frontiers, Northern Homelands*, 1977.
10. The RCAP upgraded more than 60 megabytes of information from Folio 2.1 to 3.0, thereby gaining increased processing power as well as a larger range of useful features. Although the architecture of the collection might have been slightly different had it been begun on Folio 3.0, upgrading to industry standards for publication in CD-ROM was a relatively painless operation.