

David Durian
Northern Illinois University

Corpus-Based Text Analysis from a Qualitative Perspective: A Closer Look at NVivo

This review presents information on a powerful yet easy-to-use entry-level qualitative text analysis tool called NVivo. NVivo is quite different from analysis tools readers may be familiar with already, such as TACT and Wordsmith Tools, two of the most commonly used text analysis tools currently available, in that the latter are quantitative analysis programs. NVivo can be used either to complement these quantitative tools or to explore an alternative form of analysis readers may have previously not considered undertaking in their studies. I am assuming that most readers will already be familiar with programs such as TACT or Wordsmith Tools, but if they are not, they can learn more about them by reading Rochester's recently published review of both programs.

Released in a second edition in 2000, NVivo works much like its quantitative counterparts, although from a theoretical standpoint, it is meant to analyze data in a very different way. NVivo is the latest version of NUD-IST, a program originally created in 1991 by Tom and Lyn Richards, founders of Qualitative Solutions and Research Pty, Ltd (Richards 7). NVivo is available on CD-ROM with two different manuals from QSR (http://www.scolari.co.uk/qsr/qsr_nvivo.htm) for a fee that varies depending on the type of license purchased. At the time of publication, a single academic license for the program is available for £270 (about \$430 US), and a value package of five academic licenses is available for £892 (about \$1400 US). A comparable qualitative text analysis program, such as SPSS's TextSmart, is available for \$699 US for a single academic license.

Although the program includes two manuals, the help files contained within the program itself will be sufficient for most users. The CD-ROM also contains an interactive tutorial. NVivo can process and analyze ASC-II or rich-text files, and any results that users wish to output can be saved in ASC-II text, which can be opened in any word-processing application for further analysis or incorporation into reports. Additionally, the program allows users to export node diagrams for use in programs such as Inspiration and Decision Explorer. At present, NVivo is available only for the Windows platform.¹ The system requirements for the second edition are the following: 32 MB of available RAM (although 64 MB is recommended), an Intel Pentium 100 MHz or better processor, Windows NT4 or 95 and above, and 80 MB of hard drive space for installation of the program itself (Fraser 187).

Essentially taking the concept of KWIC (Key Word in Context) from programs such as TACT or Wordsmith Tools and developing it in a different direction, NVivo is designed to allow users to conduct sophisticated analyses of electronic text data contextually. Instead of allowing users to focus on micro-level frequency-based analysis of language at the word level, as does TACT or Wordsmith Tools, NVivo allows users to focus on language within the context of macro-level text unit relations, particularly at the sentence or paragraph level. This makes NVivo a useful tool for looking at symbolic or metaphorical relations between spans of text or for discovering trends within the data that might not necessarily be revealed by statistical text analysis methods. The program focuses not on the production of descriptive statistics, but instead on description of the data themselves.

In spite of the differing strengths, TACT, Wordsmith Tools, and NVivo do share some similarities. Namely, like its quantitative counterparts, NVivo can be used to search texts for particular words, phrases, or collocations, and it does so primarily through the use of two different types of searches: text-string searches and text-pattern searches. Text-string searching allows the user to perform simple word and phrase searches, provided that he or she is looking for exact matches of the words or phrases entered. For example, if a user wanted to look for the exact sequence of letters contained in the word *love* (but not *lover*, *glove*, *clover*, etc.), then he or she would want to use the string-search capability. Text-pattern searching allows users to perform queries as they would in programs like TACT: through the use of operators such as SIMIL, wildcards, and regular expressions to look for words with similar meanings; words with common stems but different endings; ranges of numbers or letters; fixed-phrase collocations; and alternate spellings in pre-modern texts without standardized spelling. Text-pattern searching allows users to search a text for repetition, variant word forms with similar rhyme schemes, or recurrent phrases. Text units can be set at the word, line (useful in poetry analysis), sentence, or paragraph level, and if the document is formatted properly, NVivo can provide useful statistics like individual word token counts, individual word form counts, and percentage occurrence of the individual word form or word token within the document. However, NVivo does not give total word-frequency distribution counts for a document.

Beyond the similarities, NVivo has a number of features that make it quite different from its quantitative counterparts. One major difference is that, unlike in TACT, text does not have to be preprocessed in NVivo before being brought into the program for analysis. The text enters the system simply as a text file. Users will find this a useful time-saving design feature, as they will not have to spend time correcting errors that might otherwise be introduced into their texts by a preprocessing approach (as is often the case when using TACT).

Another major difference is that NVivo can analyze a text as a series of documents, with individual documents created by the user reflecting different

analytic interests in the text. Users can create annotated memos about the text, divide the text into a number of sections that can be analyzed as a part of a larger document set or as an independent section of the larger set, and make visual comparisons between various sections of a text via the use of independent windows that can be left open at the same time within the program itself. For example, making use of the multiple display windows capability of the program, users can divide novels or plays into chapters or acts as a series of text documents and then compare portions of the text contained within two individual chapters or acts with one another side-by-side via a series of open windows, as well as write notes about their comparisons in a third open window. This capability provides both a convenient timesaving device for the analysis of texts and a useful method for the in-depth comparison of two disparate sections of a text.

A third major design feature that makes NVivo different from TACT and Wordsmith Tools is that NVivo is not comprised of modules (like TACT) or a series of sub-programs with one controller program (like Wordsmith Tools). Instead, all features of the program are accessible through just one program, using one interface with two sub-interfaces. One of the two sub-interfaces, the Document Explorer, is the actual interface that allows users to organize information via a series of text documents that can then be analyzed as a group of documents or independently, depending on user need. As well as organizing data for analysis, the Document Explorer is the portion of the program that allows users to create memos. Through the "data link" annotation function, these memos can be hyperlinked to documents containing the analyzed text so that the memos can be called up in context with that data by the user. Texts brought into the system can also be analyzed using this interface, and by using the report feature, the user can create statistical reports on analyzed data that can be printed or exported to other programs for later analysis or incorporation into published research.

The second sub-interface accessible through the main interface is the Node Explorer. This interface allows the user to create an organized index system that operates through the use of "index root tree nodes" and "free nodes," both of which can be displayed on screen visually through the Node Browser or as a series of text-based logical relations via the Attribute Explorer. Index root tree nodes allow the user to organize information hierarchically, so that independent documents can be referenced either independently or as a series of topic documents, in which information is related through major topics or headings, minor topics or headings, and sub-topics or sub-headings. Free nodes are not hierarchically organized, and these nodes allow users to explore the data in a "context free" environment, which is useful if users are still trying to determine what sorts of connections can be made between documents in a set. Both of these types of nodes can later be reorganized, so free nodes can be made into index root tree nodes and index tree root nodes can also be made into free nodes. Such reorganization is useful in the interpretation

phase of a research project, once connections (or a lack of connections) among a document set have been determined.

The Node Explorer is a useful interface for organizing notes and observation memos, as well as actual excerpts or quotations from the text under analysis, into major and minor categories. Doing so, users can begin to visualize the structure of the papers they are writing as they conduct their qualitative analyses. As they change their thoughts about how ideas relate to one another, or as the rhetorical structure of their arguments become more fully formed, they can easily reorganize their notes, memos, and text selections to fit these new structures by simply reorganizing the index root tree and free nodes. As well, the resulting overall structure at the end of the reorganization process is useful later as a master outline for writing up the results of the analysis.

One final difference between NVivo and programs such as TACT is that NVivo allows users to code data using SGML or XML schemata. Thus, NVivo can be used in analysis with documents formatted by the standards of the Text Encoding Initiative. This is an important feature of the program for users who want to use corpora created over the last several years, and in this way, NVivo may lend itself to analysis in contexts where TACT is simply outdated. As more and more texts become converted to these standards and made available to scholars, this feature alone will prove to be extremely valuable for users who wish to stay current with trends in the corpus-based text analysis field.

NVivo has fewer operating-system problems than many of the quantitative programs still in use by many users today, primarily because NUD-IST is simply more recent than most of these programs. Unlike TACT, which was originally created to be run in DOS 3 environments and has known operating issues in some operating system environments created after Windows 98, Windows NT3, or Mac OS7, or Wordsmith Tools, which was originally designed to be run in Windows 3.1, NVivo was created with modern operating systems like Windows 9x, NT, 2000, XP, and Mac OS 7 (or later) in mind. The program will run smoothly in these operating systems, and users can expect the program to run unhindered with a minimum of operating-system issues as long as the software is installed properly.

Ultimately, NVivo provides users with access to a cost-effective, sophisticated tool that they can learn to use relatively quickly and easily. As well, for users wishing to make use of more recently released texts encoded by the standards of the Text Encoding Initiative, NVivo's ability to read SGML or XML encoded data is an important feature that may prove to be extremely useful in the years to come. As some users have already discovered, a blended approach using NVivo with a quantitative program such as TACT or Wordsmith Tools can provide access to a level of analysis previously unavailable.

Notes

¹ According to the QSR Web site, some Mac users are able to run it successfully using PC emulator programs, but the software has not yet been fully tested in this mode of operation.

² Most standard operating systems in use by users today will no longer allow TACT to run smoothly and coherently. This means ultimately only portions of the program will run in these environments, with the internal integrity of the program severely compromised. However, with the release of Windows XP, Windows users have found that they can make the program run more effectively using the Program Compatibility Wizard, which allows the program to run as if it were operating under the Windows 95 operating system. In terms of the Mac operating system, at the time of my writing this review, I have not heard of any reported "fixes" for operating issues in the OS 7 or above operating environments.

Works Cited

- Fraser, Donald. *NVivo: Reference Guide* (3rd Ed.). Victoria, Australia: Qualitative Solutions and Research Pty, Ltd., 2000.
- Richards, Lyn. *Using NVivo in Qualitative Research* (2nd Ed.). Victoria, Australia: Qualitative Solutions and Research Pty, Ltd., 2000.
- Rochester, Eric. "New Tools for Analyzing Texts." *Language and Literature* 10 (2001): 187-92.

Book Reviews

Nick Lacey. *Narrative and Genre: Key Concepts in Media Studies*. New York: Palgrave, 2000. x + 268 pp. \$59.95 cloth; \$24.95 paper.

Narrative and Genre is the second installment of Nick Lacey's projected trilogy of practical classroom textbooks that respond to the fact that "all individuals experience media as a set of interrelated and interacting systems." The series is strongly oriented to a very specific rationale, and has a very specific audience and use in mind. It takes its lead from a British Film Institute report's "signpost questions":

WHO is communicating, and why? WHAT TYPE of text is it? HOW is it produced?
HOW do we know what it means? WHO receives it, and what sense do they make of it?
HOW does it PRESENT its subject?

These lead to the key concepts of media agencies, categories, technologies, languages, audiences, and representations (1, qtd. from Bowker 5-6). Lacey's first book, *Image and Representation* (1996), addresses language and representation; this one addresses categories and languages; the third, *Media Institutions and Audiences* (just published), addresses agencies and audiences. Technologies are dealt with in relation to the other categories.

Lacey aims to "give the student the basic skills they require for post-16 [in American terms, precollegiate] education," and also, by emphasizing key concepts, to make his book relevant to any future media syllabus. *Narrative and Genre* "differentiates between the pre- and post-16 student, a difference that represents not just 20 months of study but, for most, the rapid intellectual development of 16-18 year olds." It is organized accordingly: "each of the key concepts is dealt with, first, at a basic level, appropriate for students at the beginning of the course. The following chapters introduce more advanced theories, such as ideology and semiotics, which are then applied to the concepts" (1-3). The first half of the book covers the theory of narrative, the second half the theory of genre. Each subject has introductory and advanced chapters, and a "history of narrative" chapter is sandwiched in between the narrative theory parts.

Both the strengths and the weaknesses of the book derive from its very distinct practical aim. Its strengths are its clarity of organization and writing, and its coverage of a fair amount of ground in narrative and genre theory. Its style is straightforward, clear, and conversational. The wealth of examples, taken largely but not wholly from pop culture, helps articulate difficult ideas. Each chapter starts with a summary of aims and is clearly sectioned into topics and subtopics. All these