A NEW ENVIRONMENT FOR LITERARY ANALYSIS

by John B. Smith
University of North Carolina

A text retrieval and analysis system called ARRAS provides rapid access to textual data; it also may herald new ways of investigating and understanding literature

Humanists have always been explorers. They sail not on seas of water but on seas of color, sound, and, most especially, words. The treasures they seek are the deepest thoughts and feelings that all experience but few express. The tools and vessels they traditionally use in their explorations are books and libraries. Like their literal counterparts, they may take many years to complete their journeys.

ARRAS... emphasizes fast, flexible open-ended analysis of individual texts.

Humanists ask questions that require consideration of very large quantities of information, especially textual information. Consequently, many are discovering that the computer is a powerful colleague. It can help them find their way through thousands of pages of text quickly and easily. It can help them make comparisons, trace similarities, plot differences. It can help them record their thoughts and communicate those thoughts to others. In short, it can help them do their work faster and better.

One particular computer-based system designed especially for humanists is ARRAS (ARchive Retrieval and Analysis System), which provides fast, flexible access to long texts, such as novels or plays. ARRAS is not the first text analysis system. Several earlier systems offer some of the same capabilities. We can gain a sense of where ARRAS fits into the spectrum of text systems by briefly comparing it with STAIRS,[1] perhaps the most widely used and best known full-text system. The relationship between STAIRS and ARRAS is a bit like that between a large truck and a small car, perhaps a sports car. STAIRS is normally used to search massive data bases containing a number of separate texts, such as a data base of legal documents or a data base of the scientific literature in chemistry. It helps one locate individual texts that may be relevant and then electronically delivers those texts to be read. ARRAS, by contrast, emphasizes fast, flexible, open-ended *analysis* of individual texts. The user will normally know which text he or she wishes to consider and will use the more extensive analytic capabilities of ARRAS to study that text in depth. Thus the two systems complement rather than compete with one another.

John Smith is an associate professor in the Department of Computer Science, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina 27514.

ARRAS has been in use on a pilot basis for approximately two years. It resides on computers at the Triangle Universities Computation Center (which serves the University of North Carolina, Duke University, and North Carolina State University) as well as at Yale University, Brigham Young University, and Queens University in Canada. The most extensive use of ARRAS to date has been with the Archive for Research on the Treasury of the French Language (ARTFL) at the University of Chicago. Scholars there and elsewhere, through the EDUCOM network, are using ARRAS for studies involving a data base of French texts that eventually will number 1500.[2]

In this article I first describe ARRAS itself, discussing what it does and how it is used. Then I explain why ARRAS can be viewed not just as a computer system that provides data, but as an entrance to a new kind of working environment. Finally, I suggest that systems or environments such as ARRAS may subtly influence the intellectual point of view of the user. That is, they may alter the way in which the inquirer views a text, the questions posed, the notion of what constitutes a valid answer or argument, and perhaps the aesthetics of the analytic experience itself.

Though ARRAS is discussed here primarily as it is applied to literary studies, it can be used in areas ranging from other academic disciplines to technical and commercial applications. From a computational point of view, there is no difference between a 200-page novel and a 200-page environmental-impact statement, a legal brief, a corporate policy manual, or a set of government regulations. All are composed of words grouped into sentences, paragraphs, chapters, or sections. The tools ARRAS provides for literary studies can be applied to other kinds of text. For example, the functions that perform a thematic analysis of a literary work can also locate all passages in an impact statement in which specific combinations of ideas are discussed. While space does not permit further discussion of these other potential applications, the reader is invited to consider analogous uses throughout the discussion that follows.

ARRAS consists of three major components: a collection of texts, an internal self-instruction system, and the analytic system itself.

The system

ARRAS consists of three major components: a collection of texts, an

SCENE II]
So;"
I know.
otherwise
further?
walks four
he does indeed
loose my daughter
I behind an arras then;
love her not
reason fall'n thereon,
for a state,
We will try it.

internal self-instruction system, and the analytic system itself.

Collections of texts. Texts to be accessed by ARRAS must first be encoded into a form the computer can "read" and then processed by a background program to produce the file structure ARRAS requires. To date, the primary mode of data entry has been typing or "key-stroking" the text on a computer terminal—a laborious and relatively expensive process (the labor cost involved in encoding a 300-page novel can be of the order of a thousand dollars). Optical scansion promises to reduce this cost substantially, but the cost will remain significant.

ARRAS is intended to be used by individuals with minimal computer experience, from a home study or from an office not necessarily near the computer in which ARRAS resides.

Ideally, for literary analysis we would like to have data bases of standard texts. Unfortunately, significant collections for most literary areas do not exist, the one notable exception being the Archive for Research on the Treasury of the French Language (ARTFL), mentioned above. For the present, most scholars will have to encode their own texts for use with ARRAS or other systems.

Internal self-instruction. The second major component of ARRAS is the internal self-instruction system. ARRAS is intended to be used by individuals with minimal computer experience, from a home study or from an office not necessarily near the computer in which ARRAS resides. Written documentation exists for ARRAS, but the system also attempts to explain itself. That is, after the user begins a session, any response other than a correctly formatted command prompts ARRAS to offer help. The user is gently led into the self-instruction system and encouraged to look first at a general introduction to ARRAS. After presenting an overview of the system's design and functions, ARRAS offers a review of the general format of the system's command language. Finally, the user is shown a list of ARRAS command terms and asked to select a term to review.

The instructions for individual commands consist of three parts: the format of the command with multiple examples, a narrative description, and a complete list of all options for that command. Some interactive systems scroll through internal user documentation taken from earlier published manuals, but command descriptions in ARRAS are designed as independent modules, each logically complete and each fitting completely on the screen. The user can move easily among the terms as well as portions of the explanation, either from within the self-instruction system or from the command mode (explained below). The user is also offered help (unless the feature is disabled at the user's discretion) whenever a command is used incorrectly. Although experience can lead to more sophisticated uses of ARRAS, the novice user can expect to start getting useful information almost immediately.

Text retrieval and analysis. The third major component of ARRAS is the retrieval and analysis system itself. ARRAS should not be thought of as a "black box" into which one inserts a text along with a set of commands and out of which one receives a completed analysis. A better analogy is a toolbox containing a set of tools, each designed for a particular task. The ARRAS design always presumes a human inquirer at the center. Thus ARRAS amplifies, rather than replaces, specific perceptual and cognitive functions. It can provide immediate recall of any portion of a text; it can reveal subtle patterns that might be missed or only partially perceived while reading; it can help the user attain a sense of proportion, emphasis, and accuracy that is difficult or impossible to gain otherwise. The system can help manage an evolving interpretation, but it is the human being who decides what information is important, what directions the analysis should take, and what the output means.

The system can help manage an evolving interpretation, but it is the human being who decides what information is important, what directions the analysis should take, and what the output means.

Using ARRAS

To use ARRAS, one asks for information in a conversational manner. For example, with reference to a text that had been entered into the system, the user might type the following:

```
Please display a concordance for
the word: fire.
```

ARRAS would respond immediately, displaying each sentence in which the word *fire* appears. ARRAS does not recognize full idiomatic English; rather, it recognizes specific keywords and their three- or four-character abbreviations. Therefore equivalent expressions for the concordance request might be

work we gladly took.
into the fire.
And, with a glance of fire, he withered
So fire
th intensity of the fire, he doth obtain
and me, we was hired to cut firewood,
sooner we light the fire
the flickering light of the fire her face seemed softer.
the women of the fire
shall she
late June, the fireflies, with their slow cool light, appeared.
fire? By
gather fire wood
"You're fired," he told me.
cold as ice, never
of him he could
are you so sure the firesta are loc
What I
Yes, it's the fall burning and the fire
As the rain ad he ran for the shelter
it up, and don't
for? Fi
fire licked up the stairway and the
protecting the fire.
WILD ELEPHANT
FIRE CHIEF
FIRE CHIEF
little woman, but the firey brilliance of her diamond solitaire ignited Crandell's interest.
he raises the
crackle, then a roar, then thirsty tongue
he fires
re could
The little metal box

```
display concordance: fire.
```

or

```
disp conc: fire.
```

The colon separates the command words from the text word or words to which the command applies. Since commands end with a period, they may be quite long, extending across several terminal lines.

...one of the most important features of ARRAS is its ability to handle categories.

The order in which commands are used is largely determined by the analyst's intent. However, one might typically begin a study by examining the vocabulary or lexicon of a text. To do so, the user would type

```
display dictionary.
```

ARRAS would immediately begin listing each word that occurs in the text, in alphabetic order and with the number of occurrences. More often, the request would be for lexical information involving a specific word or a specific alphabetic range—for example, *fire* or *fire* – *firez.* The latter would produce a list of all words beginning with the characters *fire*, such as *fire*, *fired*, *fireplace* (Figure 1).

Concordances. Having examined the lexicon for a number of words or alphabetic intervals, the user might next examine their contexts. The CONCORDANCE command, explained above, displays each full sentence in which a given word occurs. Larger or smaller contexts may be specified, expressed in terms of *words*, *sentences*, *paragraphs*, *chapters*, or *volumes.* If more or less context is desired for only a few occurrences, one can also instruct ARRAS to display the text around any single occurrence of a word for any size context desired (for example, two sentences before and one sentence after).

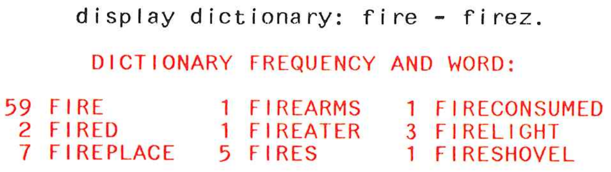
Distributions. The DISTRIBUTION command can help the inquirer gain a sense of proportion and emphasis. This command produces a bar graph in which the horizontal axis represents the text, extending left to right from beginning to end, like a tickertape; the vertical axis indicates the number of times a word occurs in each text interval. Such graphs also reveal the "rhythms" of a word or theme. When compared, they indicate the relative tendencies of words to occur together or in clusters.

Categories. The commands described thus far have been applied to single words. However, one of the most important features of ARRAS is its ability to handle categories. The user may specify a name to represent a collection of words; thereafter, reference to that name is interpreted to apply to the entire collection (Figure 2). The analyst may apply any principle of equivalence to establish categories. For example, a category might be the set of prepositions, the synonyms for a given word, or the words occurring in the same context with some designated word. Categories may also be defined to consist of sets of other categories. Thus one may have categories of words, categories of categories, categories of categories of categories, and so on (for example, the category *flora* may be defined to consist of the categories *trees*, *flowers*, and *shrubs*). Often the evolving hierarchy of categories constitutes the conceptual center of the study.

Most commands that apply to a word can also be applied to a category at any hierarchical level. For example, the DISTRIBUTION command applied to a category produces a graph for which the occurrences of all words in the category are accumulated to form a single distribution (Figure 3). Each column of the graph in Figure 3 represents two percent of the total text length; consequently, for a 250-page novel, each interval would correspond to approximately five pages.

There are two additional concepts of category in ARRAS. First, categories may consist of positions within a text. ARRAS maintains, as one perspective, a view of the text as an ordered sequence of words, numbered linearly 1, 2, 3, ..., *n.* A category may be a set of linear numbers.

Figure 1. Response, printed here in color, of ARRAS to a user's request for all words in a given document that begin with fire.



```
      display dictionary: fire - firez.

        DICTIONARY FREQUENCY AND WORD:

59 FIRE          1 FIREARMS      1 FIRECONSUMED
 2 FIRED         1 FIREATER      3 FIRELIGHT
 7 FIREPLACE     5 FIRES         1 FIRESHOVEL
```

```
         define category: fire fireconsumed fires heat heated hot hotly
CONTINUE: flame flamed flames flaming enflaming; name: firecat.

CATEGORY FIRECAT ESTABLISHED
```

Figure 2. Defining a category in ARRAS—in this case, a category of words having to do with fire. The user has named the category **firecat.**

```
         display distribution: firecat.

CUMULATIVE DISTRIBUTION REQUESTED FOR THE FOLLOWING:
          FIRECAT (CATEGORY)

FREQ. 20  |                       *
          |                       *
          |                       *
          |                       *
          |                       *
          |                       *
          |                       *
          |                       *
          |                       *
          |                       *
FREQ. 10  |                       *
          |                       *
          |  X                    *
          |  X                    *            X
          |X X                    *            X
          |X X X                  *XX          X
          |X X X   X  X           *XX      X   X  X
          |XXXXX   X  XX          *XX  XXX X   X  X       X
          |XXXXXX  X  XX         X*XX XXXX XX  XX XX XX   X
          |XXXXXXX X  XX  X  XXXXX*XX XXXX XXX XXXXX XX X XX
          +++++++++++++++++++++++++++++++++++++++++++++++++++
           2%                      50%                      100%

                          DISTRIBUTION: FIRECAT
```

Figure 3. The DISTRIBUTION command in ARRAS produces a bar graph showing the number of times a specified word or category occurs in a given interval of text. In this case, the occurrences of all words in the fire category are accumulated and plotted for each two-percent interval of text. The column of asterisks indicates more than 20 occurrences for that two-percent interval of text. Again, the program's response is printed here in color.

For example, one may use a linear category to separate homonyms, such as *rose* (the verb) from *rose* (the noun). Linear categories are particularly important for the CONFIGURATION command, described below.

From one perspective, ARRAS is simply a computer program; but it can also be thought of as a doorway leading to an expanded working environment.

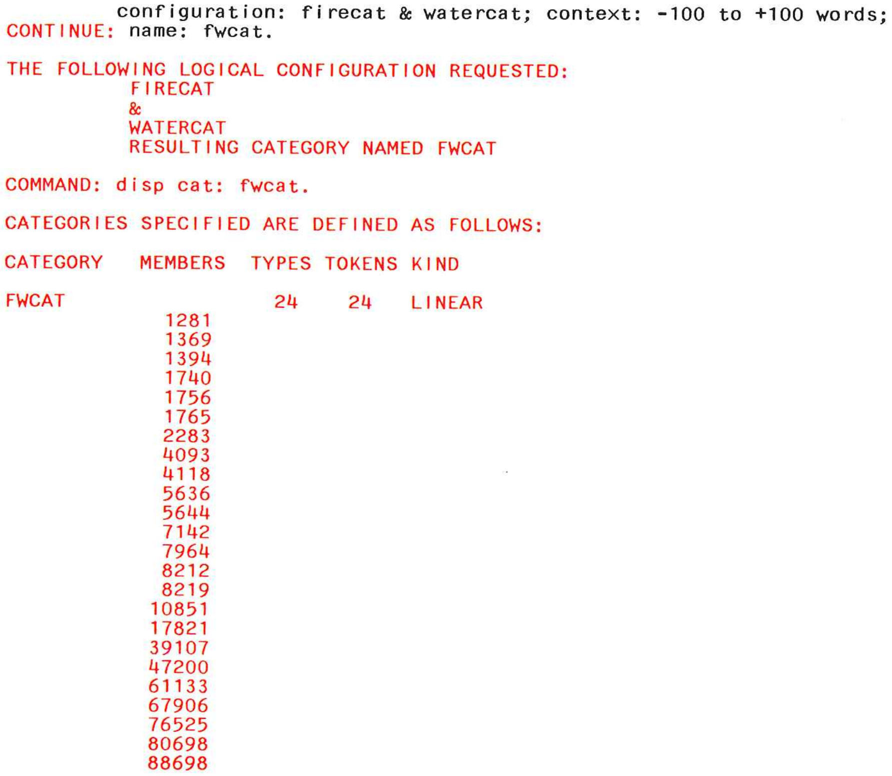
A second variation of the category concept is that each word or linear number contained within a category may have a numeric value or weight associated with it. For example, if one were considering the text of a play, one might note and scale the responses of the audience—they laughed (1), they applauded (2), they got up and left (3).

ARRAS provides additional commands to maintain the categories. These commands permit one to add and delete individual items, change the name of a category, store categories in a file system, and retrieve them at a subsequent session. The ability to store and retrieve categories can also be used to apply a standard thesaurus to an individual text as a generalized strategy for defining categories or for considering the same category structure across several texts.

Configuration. When used with a hierarchy of categories, the CONFIGURATION command is especially powerful. The user can specify any Boolean combination of words or categories; ARRAS then determines all locations in the text where that pattern occurs. The expression, by default, is evaluated within the context of a single sentence, but longer or shorter contexts can be specified, as well as alternative contexts for specific portions of the expression. Thus, for example, one might ask for all the sentences in which any word in the *fire* category occurs with any word in the *water* category. The CONFIGURATION command produces the set of locations (linear numbers) where the pattern occurs (Figure 4). ARRAS considers this set to be a category that can be named, saved, and used like any other category—for example, in a CONCORDANCE command, in a DISTRIBUTION, or as an element in another, more abstract CONFIGURATION.

An important group of commands currently being developed extends the use of the underlying function that produces the data for the graphic distribution. These commands will

Figure 4. The CONFIGURATION command in ARRAS produces a list of all locations in a text where a specified word or combination of words occurs. In this case the user has specified, within a context of 200 words, all configurations in which words in the fire and water categories occur together.



```
          configuration: firecat & watercat; context: -100 to +100 words;
CONTINUE: name: fwcat.

THE FOLLOWING LOGICAL CONFIGURATION REQUESTED:
          FIRECAT
          &
          WATERCAT
          RESULTING CATEGORY NAMED FWCAT

COMMAND: disp cat: fwcat.

CATEGORIES SPECIFIED ARE DEFINED AS FOLLOWS:

CATEGORY   MEMBERS  TYPES TOKENS KIND

FWCAT                 24     24    LINEAR
              1281
              1369
              1394
              1740
              1756
              1765
              2283
              4093
              4118
              5636
              5644
              7142
              7964
              8212
              8219
             10851
             17821
             39107
             47200
             61133
             67906
             76525
             80698
             88698
```

FLORA
The Jungle Books
Kipling
Katherine Anne Porter
Flowering Judas
Romeo and Juliet
Act II Scene II

95
A Writer's Diary
pp vii-xi
check on letters
The Cracked Servant
relates to Nighttown
Molly's
UL
D. H. LAWR
ORANGE TEA
LES DEUX MAGOTS
St GERMAIN des PRES
The Com
Letters:
pp 118

store numeric values (such as the list of numbers representing the distribution of a category over the text) that can then be analyzed using an interactive statistical system such as SCSS. Fourier analysis, for example, can characterize the underlying pattern of a distribution and indicate the relative complexity of its inherent rhythmic structure.[3] Similarly, nonmetric multidimensional scaling can be used to analyze the structure and complexity of interrelations among a group of categories.[4] Other statistical and mathematical models can be applied as well.

...will ARRAS the tool influence the intellectual point of view of the tool user?

ARRAS recognizes a number of other commands and can provide other forms of data and other services not discussed in this article. However, the sample presented here should give the reader a "feel" for the system and its potential application in language studies of all kinds.

ARRAS as environment

In a medium as flexible and dynamic as a computer system, how we think of a thing is often the most important factor in determining the nature or reality of that thing. From one perspective, ARRAS is simply a computer program; but it can also be thought of as a doorway leading to an expanded working environment.

ARRAS operates within the CMS environment.[5] It is linked to CMS through an assembler-language function developed by Russell Miller and generalized by William Verity of the Pennsylvania State University Computation Center. ARRAS uses that function to change the names of the files in which categories are stored and from which they are retrieved, thus permitting ARRAS to manipulate groups of categories much as a computer editor manipulates text files. In a similar fashion, the user may instruct ARRAS to change the text being examined. Thus, one may begin a session by accessing the text of a novel by James Joyce but change at will to a novel by Virginia Wolfe or D. H. Lawrence, or to any other text in the file system. ARRAS typically would make the change more quickly than the user could go to a shelf and take down a printed copy of the text.

The capability of changing the files on which ARRAS is working while the program is operating gives the system considerable versatility. Moreover, the user can leave ARRAS completely and work within CMS itself. If the user enters the command, CMS, control is transferred by ARRAS to the CMS system, where one can do most of the things one normally does in that system. For example, one may do word processing, or format a document; one may communicate with a colleague, or even find out what time it is. The experience is a little like swallowing the potion in *Alice in Wonderland*—it is as if one walked through a doorway, leaving ARRAS running next door while wandering around in a room labeled *CMS.* Unlike Lewis Carroll's world, however, this one permits the user to go back to ARRAS at any time simply by typing *return.*

What good is this? Suppose one were writing a journal article based in part on data obtained from ARRAS. Using the CMS editor XEDIT, the user could type in and edit the article under CMS in the same way one would use any other word processing facility. ARRAS, if instructed to do so, would make a copy of the information on the screen and store it in a CMS file. For example, a distribution or concordance listing could be stored as well as viewed on the screen. When control was transferred to CMS, the user could call up the text of the article, add a paragraph describing a new insight developed as a result of the information produced by ARRAS, and then insert directly into the text of the article the distribution, the concordance listing, or any other information supplied by ARRAS. Simply by typing *return*, then, one could go back to ARRAS and begin where one left off, with all categories intact, with access to the same text.

...ARRAS can give the user a different sense of analytic precision.

CMS also permits users to communicate with one another. Communications can be in the form of "one-liners" delivered immediately to any recipient who has access to a terminal and is logged on, or they can be much longer—such as a draft of the article being written. And, of course, the colleague who receives the communication can send back comments ranging from brief messages to a complete, edited version of the article. Such communication is routine today among those who use the same computer system, and many universities and colleges are joining computer networks. For example, BITNET[6] currently connects some 65 institutions in the United States and has links to a number of other countries. With BITNET and similar networks, scholars can communicate with one another in a greatly expanded academic community. Obviously one can mail drafts of a working document to a colleague without using a computer network, but the opportunity to exchange three or four versions a day, compared with three or four a month, makes possible a form of collaboration not possible otherwise.

ARRAS and other computer systems can lead the way to new concepts of relational structure.

Thus ARRAS, used within the context of CMS, provides a working environment that includes quick retrieval and display of texts, a range of analytic functions, and word processing and formatting services, as well as facilities for communication and document transmission. Each component is just a program or a computing service, but collectively they add up to more than the sum of the parts. Together, they constitute a new working environment.

Point of view

When scholars use ARRAS within the enriched working environment described, will they be doing what they have always done in a different way, or will they be doing something different in kind? That is, will ARRAS the tool influence the intellectual point of view of the tool user? The question is a slippery one. We tend to believe that when we look at the world we see it as it really is. A moment's reflection, however, is sufficient for most to realize that one's view is colored and shaped by one's background, education, politics, conceptual categories, and perhaps even the abstract models of structure one knows and can apply. In what specific ways, then, might ARRAS alter one's conceptual point of view?

When one reads a text, the words and images evoke memories that enrich the experience. A primary component of the experience, however, is the temporal nature of reading: the slow unfolding of events and patterns, the anticipation, the confirmation of what is expected or the surprise of the unexpected. When one uses a system such as ARRAS to consider a text or group of texts that one has already read, one's experience is outside that linear continuum of time inherent in reading. One can "look down" on the text and see it "whole." One can perceive patterns that could not be seen while reading. One can see one's largest impressions confirmed, modified, or denied by textual reality.

The computer...can be an instrument of perception and cognition, a fine as well as powerful lens for the mind.

To be more specific, ARRAS can give the user a different sense of analytic precision. Since ARRAS knows every word in a text, it can help sharpen the analysis. If, for example, one is interested in thematic structure, ARRAS can help with the processes of defining themes, checking for concepts omitted, confirming that all important words and phrases are accounted for. One can examine each occurrence of a theme in a text with the assurance that the examination is complete. One can locate all instances of a pattern or combination of themes with the same assurance.

ARRAS can also provide a different sense of proportion and rhythm. While reading a text, one may sense that a given theme, grammatical form, or stylistic feature is prevalent in one section but not in another. However, to gain a true sense of proportion from impression alone is difficult. ARRAS can show exactly where a feature appears and with what frequency; it can also produce materials that make comparisons simple, direct, and accurate.

When researchers are able to use ARRAS with large data bases of texts, they may gain a different sense of historical development. Scholars will be able to examine patterns or themes, stylistic traits, grammatical forms, and other features in a longitudinal way that is not practical without the computer. This possibility will permit the asking of questions that could not be asked until now, except in a very general sense. If ARRAS and other systems can compress the detailed examination of such questions into a period of weeks, perhaps days, the intellectual experience of examining the material will also be compressed. What will be the nature of a critical or analytic encounter of this sort? Will such accelerations of thought "feel" the same or will the information/time compression produce a different kind of intellectual experience?

ARRAS and other computer systems can lead the way to new concepts of relational structure. Language and literary studies currently use a very limited set of abstract models. ARRAS can greatly expand the repertoire of relational models by preparing data for a variety of statistical and mathematical tools. We noted earlier that Fourier analysis can tease out the predominant rhythm of a complex distributional pattern and indicate its relative complexity,[3] and that principal component analysis and nonmetric multidimensional scaling can, respectively, help the user locate clusters of analytic factors and characterize patterns of association.[4] These tools will expand the analyst's perception. The analyst will see patterns too subtle to be perceived directly while reading, but there nevertheless. Not only will one be able to see structures and relations not seen before, one will be able to characterize those structures—to say that this thematic structure is more complex than that thematic structure, or that the rhythm for one stylistic trait is more regular than the rhythm for another trait.

Thus, ARRAS may begin to influence the perceptual and analytic perspectives of those who use it. Scholars may come to see texts in a different way. They may clearly perceive patterns that have not been seen before. They may gain a different sense of the completeness or adequacy of a particular analytic approach. They may see proportion and rhythm more clearly. And they may be able to answer as well as ask questions that, without the help of the computer, would forever remain speculations.

Conclusion

The computer, at times, can be more than a purveyor of cold, isolated facts. It can be an instrument of perception and cognition, a fine as well as powerful lens for the mind. This realization is both more and less surprising than we might expect. We are not surprised that the nineteenth-century revolution in transportation enabled people to see, literally, things never before seen by members of a given culture. We should not be surprised that the current computer revolution will enable us to see, abstractly, things not seen before. We may someday explore vast continents of literature or history or other realms of information, much as our ancestors explored new lands. We can foresee those explorations as the logical extension of the present; and when they come, the experiences of discovery will be extraordinary. Moments of surprise! Moments of joy!

References and notes

1. For more information on STAIRS, see *STAIRS/CMS General Information Manual*, Order Number GH12-5147, IBM Corporation (obtainable through IBM marketing offices).
   To gain a sense of the range of other full-text systems, see the following publications:
   For humanities applications, see S. Hockey and I. Marriott, *Oxford Concordance Program Version 1.0 User's Guide*, Oxford University Computing Service, Oxford (1982); D. Ross, Jr., and R. H. Rasche, "EYEBALL: a computer program for description of style," *Computers and the Humanities* **6**, No. 4, pages 213-221 (March 1972); A. van Dam *et al.*, "An experiment in computer-based education using Hypertext," final report on NEH Grant No. EH-22258-75-95, Brown University Division of Applied Mathematics, Providence, Rhode Island (1976); and A. S. Frankel, "All about the Responsa Retrieval Project," *Jurimetrics* **16**, pages 149-156 (1976).
   For office automation applications, see M. Hammer *et al.*, "Etude: an integrated document processing system," Office Automation Group Memo OAM-028, MIT Laboratory for Computer Science, Cambridge, Massachusetts (1981).
   For an ambitious full-text data base approach to publishing, see *Publishing From a Full-Text Data Base* (Second Edition), U. S. Government Printing Office, Washington, DC (1983).
2. For a description of the analytic capability of ARRAS when used with a large data base, see R. Morrissey and C. Del Vigna, "A large natural language data base: American and French research on the treasure of the French language," *EDUCOM Bulletin* **18**, No. 1, pages 10-13 (Spring 1983).
3. J. B. Smith and B. A. Rosenberg, "Rhythms in speech," *Computer Studies in the Humanities and Verbal Behavior* **4**, No. 3/4, pages 166-173 (1975).
4. J. B. Smith, "Thematic structure and complexity," *Style* **9**, No. 1, pages 32-54 (Winter 1975).
5. CMS (the Conversational Monitor System) is a component of IBM's VM/370 operating system. It presents to each user of a large computer system a virtual machine, or software replica of a real computing system, for his or her personal use. The user may store information in files, use an editor to insert and modify information, and execute programs belonging to the user or supplied by the real system to perform a broad range of tasks. Moreover, since each user's virtual machine is part of a real, larger computer, users may also communicate with one another, exchange files of information, even share their virtual machines. When a single large installation is part of a network of such installations, the group that one can communicate with and share with grows enormously. For more information on CMS and VM/370, see L. H. Seawright and R. A. MacKinnon, "VM/370—a study of multiplicity and usefulness," *IBM Systems Journal* **18**, No. 1, pages 4-17 (1979); also see *IBM Virtual Machine System Product: CMS User's Guide*, Order Number SC19-6210, IBM Corporation (obtainable through IBM marketing offices). For information on network access, see Reference 6.
6. I. H. Fuchs, "BITNET—because it's time," *Perspectives in Computing* **3**, No. 1, pages 16-27 (March 1983). ■