A Computer-Aided Discovery of the Tech Review Genre

Cheryl Geisler
Department of Language, Literature, and Communication
Renneselaer Polytechnic Institute
Troy, New York 12180
geislc@rpi.edu

David Kaufer
Department of English
Carnegie Mellon University
5000 Forbes Ave., Pittsburgh, PA 15213
kaufer@andrew.cmu.edu

Tech reviews represent a genre whose function is closely related to genres in technical communications like the scientific article, popular science articles, the grant proposal, and the manual. Like these other genres of scientific and technical communication, tech reviews are part of the circuit of culture by which products of science and technology are diffused in the culture yet, unlike these other articles, they had not yet been studied.

Tech reviews are also important in the diffusion of technology. In fact, even though the Palm Pilot was the most successful consumer IT product introduction in history, it was several years before any advertising was done for it. It thus fell to more informal routes, like word of mouth and tech reviews, to spread the word about this new product.

Like an increasing number of texts, tech reviews are relatively easy to come by the digital age. In particular, for tech reviews, the ProQuest's ABI/INFORM Global database seemed an appropriate source since it provides access to 1800 business periodicals, many of them with full-text access. This abundance is a mixed blessing: Between 1996, the year the Palm Pilot was introduced and 2002, ProQuest referenced 1,988 articles containing the term "personal digital assistant" in citations and abstracts; 1.472 of these were available with full text. It was not feasible for ordinary readers to seriously review and analyze this number of texts. Furthermore, a few hours of perusal are enough to show that only some of these articles are actually tech reviews.

According to ProQuest, between 1996 and 2003, Stuart Alsop, a well known reviewer of technologies, published 242 articles. Reviewing these by hand, only 34 actually dealt with reviews of specific technologies and for the reader interested in isolating these 34 from the 242, the task would be difficult. How can we take advantage of the amazing digital libraries for genre-based search when their very size makes the task daunting?

We have explored an answer to this question through a new text analysis tool called DocuScope (Kaufer, Ishizaki, Butler, Collins, 2004). To build the sets of texts from which Docuscope would learn how to discriminate tech reviews from non-reviews, 34 Stuart Alsop tech reviews were combined with the reviews of Stephen Wildstrom. Like Alsop who wrote the tech review column for *Fortune* from 1996 to 2003, Wildstrom served as the tech columnist for *Business Week* from 1995 through 2003. In the 626 articles he wrote during that time, we used intuition to identify 78 of them as reviews of mobile technologies. Combined, this gave Docuscope a set of 112 tech reviews on which to learn.

To provide a comparable set of articles that were not tech reviews, we selected 52 articles across the same time period, that our intuitions told us were not tech reviews. To insure a wide range of features in these so-called non-reviews, we made sure to include some of all the major types of articles we had seen mentioning PDAs, in particular:

* articles about specific companies associated with personal digital assistants (17),
* articles about industry trends in personal digital assistants (15),
* articles discussing an application of the technology of personal digital assistants (8),
* articles profiling specific CEO's in companies associated with personal digital assistants (5),
* articles reporting academic research on personal digital assistants (4), and
* articles providing social commentary on phenonmena like time management using personal digital assistants (3).

Docuscope build a discriminant classifier that separated reviews from non-reviewers with 91% accuracy (using an unbiased cross-validation method). We then asked it to classify a third set of articles it had never seen that included a mix of both reviews and non-reviews. This set consisted of 112 articles published during 1996 that mentioned personal digital assistants in their titles or abstracts minus ones that had been included in the training sets. These thus represented the equivalent of the results simple search in the ProQuest database for the year 1996, about one tenth of the set of potential reviews with which we had been faced originally. This seemed, then, like a good test of the power of Docuscope to aid with the rhetorical analysis using large-scale textual databases.

Before studying how Docuscope had classified the text, one of us read and classified the 112 new texts. Docuscope managed to match a human readers' judgment most but not all of the time. For those 48 texts judged to be reviews, it picked out 35 of them, 73%. For the 64 thought to be non-reviews, it picked out 42 or 89%. Docuscope also added another 7 texts to the review pile that we would not have put there; and would have left out 13 texts from the pile that we would have put there. This produced a total of 20 apparently misclassified texts from the original set of 112, an apparent error rate of 18%.

Manual \ Docuscope	**Review**	**Non-Review**	**Total**
review	35	13	48
Non Review	7	57	64
Total	42	70	112

There are two other issues to consider: First, in what sense might Docuscope's judgments might actually be better than the human reader? To get a sense of how well I had done under these conditions, then, we went back and took a second look at the twenty texts on which Docuscope and we had disagreed. We read more carefully, looking for elements of mixed genres, and classifying each text a second time according to its predominant elements.

This secondary review resulted in my changing my mind in 9 out of 20 cases. Two of the texts that we had classified as non-reviews, we decided were actually reviews; seven of those we had classified as reviews, we decided were non-reviews. When we factored these corrections in, the agreement with Docuscope rose from 82% to 90%, with Docuscope and ourselves disagreeing in the final analysis on only 11 texts, less than 10% of the original sample. Thus, not only may Docuscope be a more efficient way for a rhetorician to pick out specific texts for genre analysis in a large-scale database, it may also be a more accurate way.

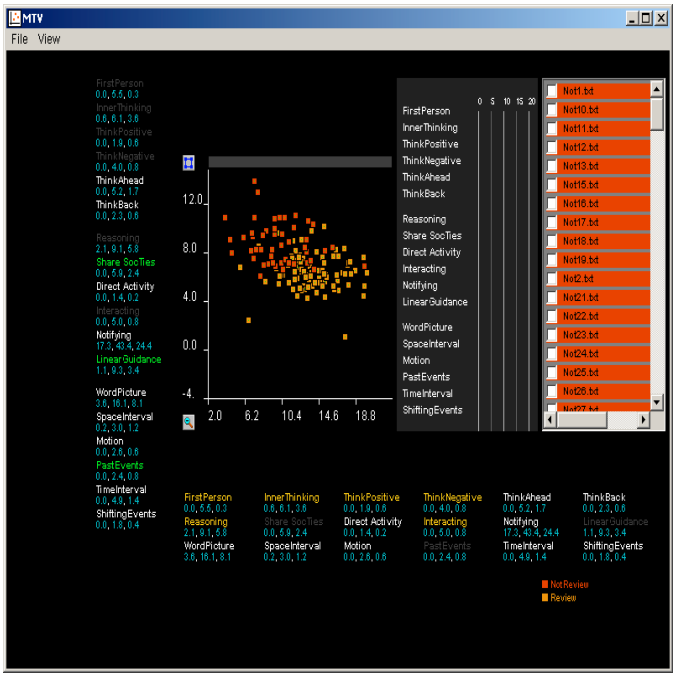

Fig.1 This is gross map of split because we are only mapping at the dimension level (18 dimensions), not at the string class level, where we did the statistics.

Reference

Kaufer, D., Ishizaki S. Butler, B. & Collins J. The Power of Words: Unveiling the Writer and Speaker's Hidden Craft.