Mapping Techno-Literary Spaces:
Adapting Multiple Correspondence Analysis for Literature and Art Informatics

Stephen Paling

School of Library and Information Studies, University of Wisconsin-Madison, 4251 Helen C. White Hall, 600 N. Park St., Madison, WI 53706-1380. E-mail: paling@wisc.edu.

Introduction

This paper describes the use of multiple correspondence analysis (MCA) for data exploration as part of a recently completed study of the use of information technology (IT) by literary authors. The study (Paling, 2005) discussed in this paper constitutes part of an ongoing effort to establish *Literature and Art Informatics* (LAI), *the interdisciplinary study of the design, uses and consequences of information technologies that takes into account their role in the creative efforts of writers and artists* (Paling, 2006; Kling, 1999). This paper is primarily methodological in nature, and has two goals:

1. To provide an introduction to (MCA), a statistical method for exploring and visualizing data derived from multiple scales or measurements.

2. To provide an example of how MCA can be used to explore and visualize empirical data about the use of information technology.

A full mathematical description of the technique is beyond the scope of a developmental paper such as this one, so the paper will focus on interpretation and data preparation in MCA use.

Correspondence analysis is "an exploratory multivariate technique that converts a matrix of nonnegative data into a particular type of graphic display," and *Multiple correspondence analysis* is "concerned with displaying the categories of more than two discrete variables" (Greenacre & Hastie, 1987, p. 437). MCA takes data from multiple scales and establishes maps in non-Euclidean, low-dimensional vector space that look superficially like scatter plots. The maps are abstract vector spaces which depict chi-square distances between points. The points can represent various entities such as individual respondents or descriptive categories.

A simplified example will give an introduction to part of the current study, and to some of the basic principles behind MCA. The mathematics for the example are considerably simplified, and more involved instances of MCA are meant to construct much richer exploratory maps for more complex data sets.

As part of the current study respondents were asked how positively or negatively they viewed the use of computer technology to produce innovative literature. They were asked the question twice, once in the context of evaluating their own success, and again in evaluating the work of others. They were asked to give a response on a 5-point Likert scale (5 highly positive, 1 highly negative). Figure 1 shows a scatter plot of hypothetical responses for four respondents meant to demonstrate the basic principles of the technique, with the data table superimposed (Distances in the figures for this example are approximate).
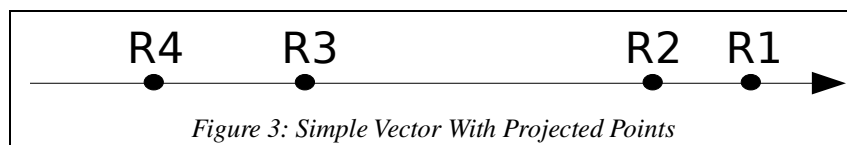
Figure 2 shows the same scatter plot, but with the optimal subspace based on the least-squares distance to each point included. A line has been drawn orthogonally from the optimal subspace to each point to represent the chi-square distance to each point (Minium, King, & Bear, 1993, pp. 175, 454). This is a similar procedure to the chi-square test, but no test of significance is involved in MCA. Instead, the chi-square distances are used to project each point onto a lower-dimensional space.

Figure 3 shows the rotated vector for this simplified example with the two-dimensional Euclidean mapped mapped onto a one-dimensional space. It is worth noting the differences between Figures 1, 2, and 3. On one hand, Figure 3 contains less information than Figure 1 or Figure 2. For example, because the pair of scores for R1 and R2 are reversed (4,5 and 5,4 respectively), they appear on opposite sides of the optimal subspace in Figure 2. But Figure 3 offers an interesting exploratory perspective. Because the scores for R1 and R2 are so similar, their points lie close together on the optimal subspace. In fact, the averages of their respective scores are identical at 4.5. Figures 1 and 2 do not show that similarity as readily. In contrast, R3 and R4 lie farther along the optimal subspace, close to one another, but not as close as R1 and R2 to each other. This adds an interesting exploratory perspective that becomes more useful with data richer than the data in this example.

	Self	Other
R1	4	5
R2	5	4
R3	2	3
R4	1	3

Figure 1: Conventional Scatter Plot with Data Table
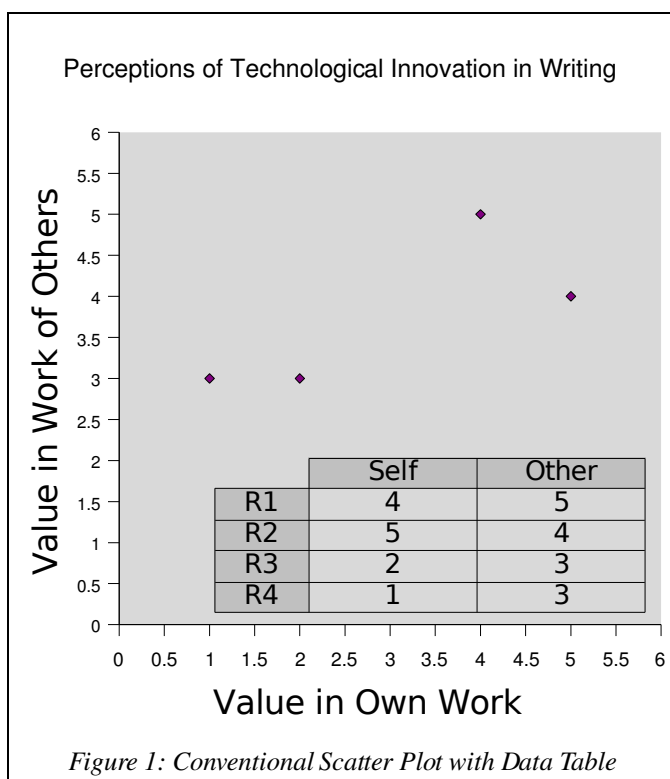
MCA has an extensive history in French sociological scholarship (Greenacre, 1984, p. v), but has been used less extensively elsewhere. The method is also an extension of dual scaling, and Nishisato (1994) points out that it is known by other names in disparate disciplines (p. xi). It has also been used in the study of information. Doré and Ojasso (2001) used correspondence factor analysis to analyze publication trends. Shen, Li, and Shen (2006) used correspondence analysis to evaluate university library Web sites in China.

The current study follows work (Paling and Nilan, 2006) which posited a synthetic conceptual framework made up of four key values, adapted from Bourdieu (1996), to describe the motivations of little magazine editors in pursuing their creative goals. The study found evidence of support for three of the four key values, and posited the idea of *intensifying use of technology* (IUT) to describe the use of information technology to pursue a value more strongly than before. The current study focused primarily on one of those key values, Positive Regard for Avant-garde-ism (AvG), in order to validate part of the original findings in greater detail.

Figure 2: Scatter Plot with Optimal Subspace and Orthogonal Distances.

Figure 3: Simple Vector With Projected Points
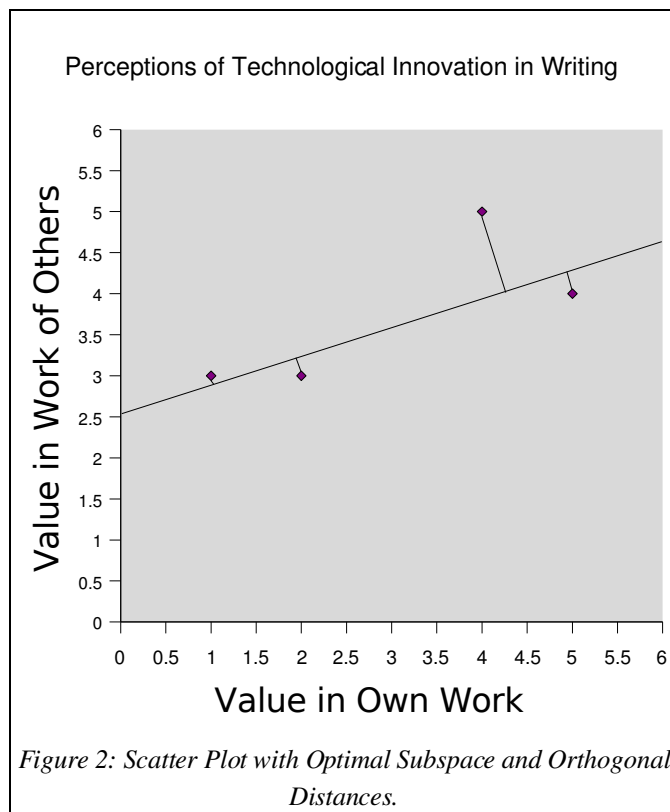
Bourdieu (1996) wrote extensively about the literary community, but did not specifically address the role of information technology (IT) in the creative lives of authors or artists. This paper will focus on two conceptual constructs, consecration and wiredness, that are being developed for mapping techno-literary spaces. *Consecration*, as conceived by Bourdieu (1996), refers to the achievement of decorations, prizes, and awards of various kinds (p. 123), including appointments to academic positions. Those achievements are directly germane to understanding the creative efforts of literary writers since writers will often define themselves in terms of those achievements, whether through participation or opposition. *Wiredness* was defined in the current study as the use of information technology in an author's efforts to create and publish literary work. These elements serve to establish a link between the broader fields of social informatics and the study of literature and art. *Techno-literary spaces* were defined as maps, created through MCA, of the associations between authors based on similarities and differences in their approaches to using IT in their creative efforts, the degree of consecration attained by each author, and the authors' views of AvG.

Bourdieu, (1996, 1984), made use of various social maps. There are several key differences that set the current study apart from Bourdieu's, however. First, Bourdieu did not always use empirical data in constructing his maps, instead creating maps which were conceptual in nature. Figure 4 shows one such map (Bourdieu, 1996, p. 122) based on degrees of consecration and economic profits. No explicit empirical data were provided with the map, though. Second, when Bourdieu did incorporate

empirical data into the construction of his social maps, he did not focus on IT use by authors or other artists.
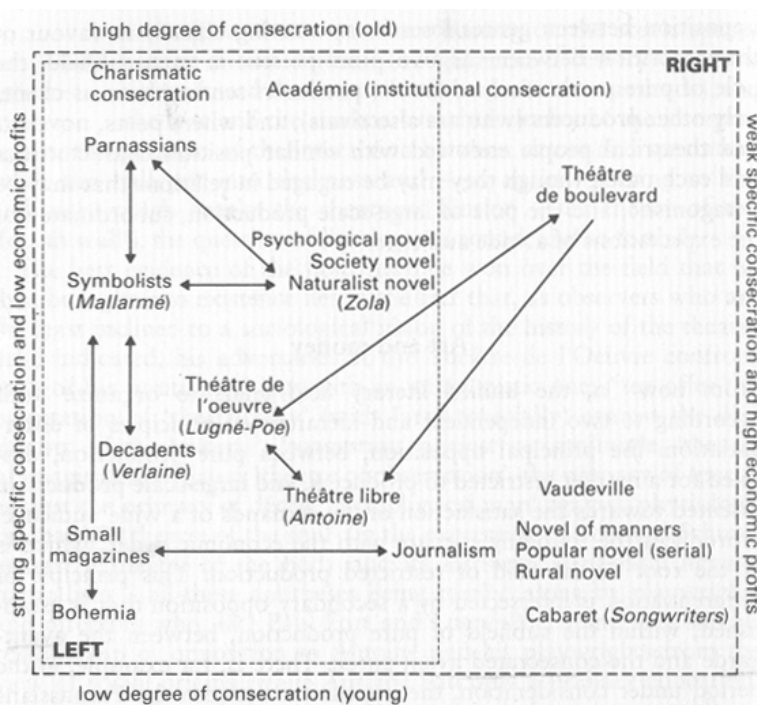

Figure 4: Conceptual Map of Social Space from Bourdieu.

From Pierre Bourdieu, translated by Susan Emmanuel, *The Rules of Art: Genesis and Structure of the Literary Field.* Copyright © 1996 Polity Press.

Methodological Background

The current study made use of multiple scales to operationalize both wiredness and consecration. Because the number of scales exceeded the number that could comfortably be depicted in a conventional two- or three-dimensional scatter plot, MCA offered an appropriate set of techniques for mapping the associations between the respondents in the study.

The study operationalized consecration with the following variables:

1. Professional Status: based on factors such as whether the respondent held an academic position (high professional status), was a student (lower professional status).
2. Type of Publication(-s) Achieved: based on whether the respondent had individual stories or poems published, had work anthologized, or had a book published.
3. Awards Received: based on the receipt of major or minor literary awards.
4. Membership in Honorary Literary Organizations: based on membership in societies not open for general membership.
5. Honorary Directories: based on inclusion in honorary directories such as *Who's Who*.

The details of the variable scales have been omitted for the brief paper format. The total scores for consecration ranged from a low of 5 to a high of 90.

Wiredness was operationalized with a second set of variables:

1. Place of Publication: based on whether a respondent published work primarily on paper, online, or some combination in between.
2. Past and Current Use of IT: based on whether a respondent has always used IT in his or her writing.
3. Technologies Used: based on which technologies (chosen from a list) each respondent used, along with an Other category for unanticipated technologies.
4. Primary Technology: based on which technology on the previous list a respondent considered to be her or his primary technology.
5. Media Used For Submission: based on whether the respondent used paper or electronic means of submission to publishers.
6. Primary Media Used For Submission: based on which technology on the previous list a respondent considered to be her or his primary technology for submitting work.

The details of the variable scales have been omitted for the brief paper format. The total scores for wiredness ranged from a low of 2 to a high of 20.

The respondents were also asked to provide responses on 5-point Likert scales that measured how positively or negatively each respondent viewed various factors for assessing the success of their own efforts, as well as the efforts of others. The focus here is on two questions that bear directly on AvG:

- How positively or negatively each correspondent viewed technological innovation in his or her own writing (variable abbreviated FT_s, short for FreshTechSelf).
- How positively or negatively each correspondent viewed technological innovation in the work of other writers (variable abbreviated FT_o, short for FreshTechOther).

The respondents' scores for wiredness and consecration (with the scores from the constituent variables for each scale being combined) were originally used to build a simple scatter plot with which to track the purposive sample as it was built. Figure 5a shows the conceptualization of the sample space, and Figure 5b shows the actual scatter of respondents across that space.

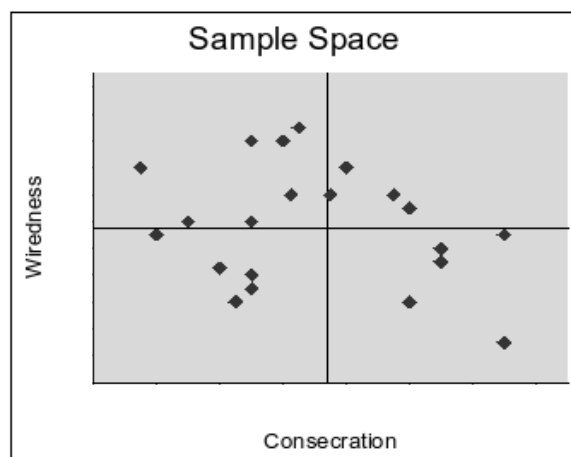| A	B
High Wiredness Low Consecration	High Wiredness High Consecration
C	D
Low Wiredness Low Consecration	Low Wiredness High Consecration

Figure 5a: Conceptualization of Sample Space

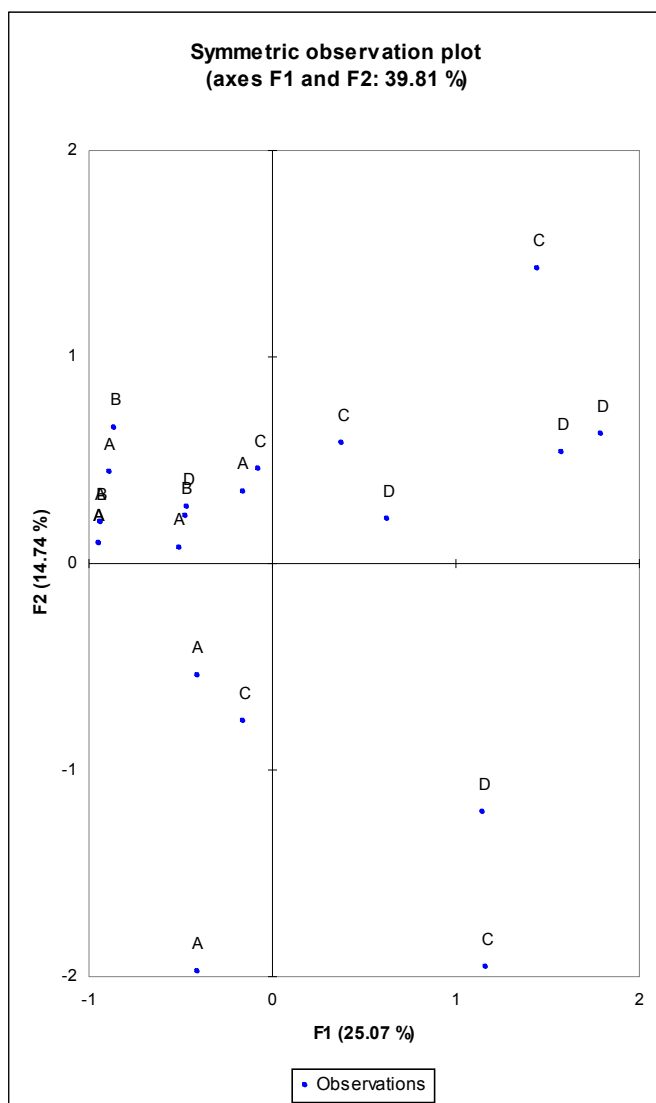*Figure 5b: Scatter Plot of Sample Space*

Figure 6: MCA Map of the Empirical Results

The map of the sample space turned out to be unexpectedly successful. A respondent's presence in a particular quadrant of the sample space provided useful information about, for example, how positively or negatively she or he viewed the use of computer technology to produce innovative writing through the introduction of animation, etc. The respondents in quadrants A and B viewed that type of innovation much more positively than did the respondents in quadrants C and D.

A third variable, such as the score for each member of a quadrant on one of the Likert-scale questions, could be added as a third dimension. However, such additional comparisons became unwieldy on a conventional scatter plot if more than three scales were involved. In contrast, MCA allowed description of the associations between the various respondents, as well as a depiction of where each respondent fell in relationship to each group and variable.

Figure 6 is an MCA map showing a space in which the respondents (represented by single letters indicating the group to which each belonged) are grouped along with the other variables that went into the map: wiredness (W), consecration (C), FT_s, FT_o. A label such as C-90 in the lower right part of the figure represents a score of 90 on the consecration scale. The map thus represents 5 dimensions in low-dimensional space. We can look at the lower right portion of the map to see a sample interpretation. A respondent from Group C and a respondent from Group D (the less-wired groups) fall on the map near the lowest scores (1) for FT_s and FT_o. They also fall near both high and low scores for consecration, indicating that both highly consecrated and relatively unconsecrated writers in those two groups often had a low opinion of technological innovation in literary writing.

The middle left-hand portion of the map in Figure 6 is difficult to read because of the close proximity of the respondents and the variables, which reflects the relative homogeneity of the data from respondents in groups A and B. A simpler map will often prove easier to read. Figure 7 is identical to the map in Figure 6, but shows only the respondents, with each respondent labeled according to the quadrant in the original sample space to which that respondent belonged.

Notice in Figure 7 that the correspondents in A and B (the more-wired groups) cluster relatively close together. The scores for FT_s and FT_o for respondents in those quadrants tended to be almost uniformly positive. In contrast, the respondents in C and D tended to provide lower scores on those questions, though not uniformly so. Thus they are scattered more widely across the map.

Figure 7: MCA Map Showing Respondents Labeled by Original Sample Quadrant

Because the map in this case compressed 5 dimensions into a two-dimensional map, some of the variance was lost. The cumulative variance for the two axes in Figure 7 account for 39.81% of the observed variance. In contrast, the correspondence map in Figure 8 accounts for an appreciably higher percentage of the variance (93.35% vs. 39.81%). The correspondence map in Figure 8 only measures the association between group membership, FT_s, and FT_o, and as a result the map can account for a very high percentage of the variance. A conventional scatter plot could have shown the scatter of scores by quadrant, but would not have been as helpful in showing associations between whole subgroups and the score for a particular variable, e.g., FT_o-4.

Data Preparation and Software

Preparing data for MCA is surprisingly easy. Figure 9 shows spreadsheet columns containing the data used in generating the maps in Figures 6-8.

The cells contain the ID number for each respondent, the group to which the respondent belonged, the respondent's score on FT_S and FT_O, as well as the respondent's scores on the consecration and wiredness scales. Microsoft Excel was used in the current study. Excel in its native form cannot perform MCA or produce the maps, so the XLStat add-on from Addinsoft was used. The scales used to produce data for MCA need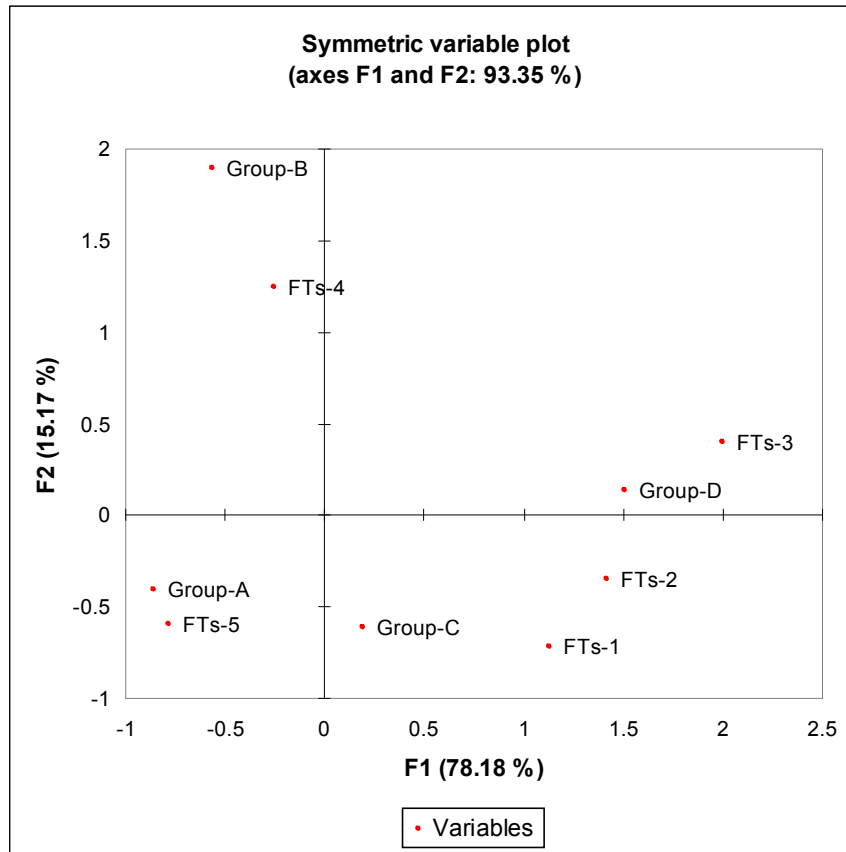 not use the same units of measurement. Greenacre provides an example in which the height, weight, shoulder width, and waist circumference of respondents were combined into vectors (1984, p. 15). This makes MCA ideal for exploring complex data sets in which disparate variables are used.

Summary and Conclusions

As an exploratory technique MCA offers rich possibilities. In the current study, it allowed multiple scales to be mapped onto two-dimensional spaces that depicted the various associations between individual respondents, subgroups of those respondents, and the variables that have been shown to describe their motivations as authors. Traditional tables and scatter plots provide a more limited set of data exploration tools. The maps used in the current study and other create enhanced visualizations that show the data from perspectives not possible with simpler tools. In an area such as the study of IT use in literature or art, where multiple factors contribute to the establishment of a writer's or artist's position in the field, MCA can transform tabular data into coherent spaces that demonstrate the associations between those various factors in rich and and useful ways.

Figure 8: MCA Map with Reduced Number of Variables.

#	Group	FT$_s$	FT$_o$	C	W
R01B	B	5	5	33.5	15
R02A	A	5	5	4.5	11
R03C	C	5	5	23	2
R04D	D	5	4	61	11
R05A	A	5	5	30.5	11
R06D	D	3	3	33	11
R07B	B	5	4	43.5	10
R08C	C	2	2	4.5	2
R09C	C	3	4	27	2
R11D	D	1	2	70.5	2
R12A	A	5	5	15.5	20
R13C	C	1	1	5	6
R14A	A	5	5	30	20
R15A	A	5	5	31	15
R16A	A	5	4	31.5	15
R17C	C	3	5	14	6
R20B	B	5	4	30.5	11
R21A	A	5	4	14	6
R23D	D	1	2	33	2
R25A	A	4	5	12	6
R26D	D	1	1	90	11

Figure 9: Spreadsheet Cells Used to Create MCA Maps.

References

Bourdieu, P. (1996). *The rules of art: Genesis and structure of the literary field*. Stanford, CA: Stanford University.

Doré, S., & Ojasso, T. (2001). How to analyze publication trends by correspondence factor analysis: Analysis of Publications by 48 countries in 18 disciplines over 12 years. *Journal of the American Society for Information Science and Technology, 52*(9), 763-769.

Greenacre, M.J. (1984). *Theory and applications of correspondence analysis*. Orlando, FL: Academic Press.

Greenacre, M.J. and Hastie, T. (1987). The geometric interpretation of correspondence analysis. *Journal of the American Statistical Association, 82*(398), 337-447.

Kling, R. (1999, January). What is Social Informatics and why does it matter? *D-Lib*. Retrieved September 12, 2006, from http://www.dlib.org/dlib/january99/kling/01kling.html.

Minium, E.W., King, B.M., & Bear, G. (1993). Statistical Reasoning in Psychology and Education (3rd ed.). New York: Wiley.

Nishisato, S. (1994). *Elements of dual scaling: An introduction and analysis*. Hillsdale, NJ: Lawrence Erlbaum.

Paling, S. (2005, October). *Technology, tradition, and the individual talent: Literary authors and aesthetic use of information technology*. Poster presented at the ASIS&T annual meeting, Charlotte, NC.

Paling, S. (2006, November). *Artistic use of information technology: Toward a definition of Literature and Art Informatics*. Paper presented at the Annual Meeting of the American Society for Information Science and Technology Social Informatics Symposium, Austin, TX.

Paling, S., & Nilan, M. (2006). Technology, values, and genre change: The case of little magazines. J*ournal of the American Society for Information Science and Technology, 57*(7), 862-872.

Paling, S. (2005, October). *Technology, tradition, and the individual talent: Literary authors and aesthetic use of information technology*. Poster presented at the ASIS&T annual meeting, Charlotte, NC.

Shen, X., Li, D., & Shen, C. (2006). Evaluating China's university library Web sites using correspondence analysis. *Journal of the American Society for Information Science and Technology, 57*(4), 493-500.