Running Head: SECRET LIFE OF PRONOUNS

The Secret Life of Pronouns:

Linking Latent Semantic Analysis of Writing Samples to Physical Health

R. Sherlock Campbell and James W. Pennebaker

The University of Texas at Austin

Correspondence should be addressed to James W. Pennebaker, Department of Psychology, The University of Texas, Austin, Texas 78712 (email: Pennebaker@psy.utexas.edu). Preparation of this paper was made possible by a grant from the National Institutes of Health (MH52391). We are indebted to Kyle Pierce, Darrell Laham, Katie Galloway, and Andrea Garcia for their help on this project.

Abstract

Numerous disclosure studies have demonstrated that individuals randomly assigned to write about emotional topics evidence improved physical health compared to those who write about superficial topics. The writing samples from three previously published studies using 74 1st year students, 50 upper division students, and 59 maximum security prisoners were reanalyzed using Latent Semantic Analysis (LSA) to explore possible relationships of writing content and style to changes in physician visits following the disclosure intervention. LSA revealed that flexibility in the use of common words when writing about traumatic memories was related to positive health outcomes. More specifically, changes in the usage of personal pronouns were driving this effect. The findings point to the importance of the role of discussing the self and social relationships in therapy and, at the same time, the remarkable potential of techniques such as LSA. Flexibility in the reconstruction of social realities when writing about traumatic memories is thus linked to improved health.

The Secret Life of Pronouns:

Linking Latent Semantic Analysis of Writing Samples to Physical Health

Beginning in the 1960s, several studies demonstrated the adverse health effects of traumatic experiences. Holmes and Rahe (1967), for example, developed a comprehensive trauma survey that found that the more traumas that individuals had experienced in the previous year, the greater the probability that they would die or be hospitalized due to a number of health problems. By the late 1970s, it was clear that there were large individual differences in how people dealt with traumas and that some coping strategies were better than others. For example, traumatized individuals who had a social support network were less likely to experience illness episodes than those lacking support (e.g., Cobb, 1977). In addition, emotional upheavals that were kept secret – such as sexual traumas – were more likely to lead to mental and physical health problems than those experiences that could be openly discussed (Silver, Boon, & Stones, 1983).

Drawing on this early work about the secrecy of traumas, we developed a technique wherein people were randomly assigned to write about either traumatic experiences or superficial topics for 15-20 min per day for 3-5 consecutive days. At the time, it was assumed that merely writing or talking about emotional upheavals would have positive effects on people's health. Indeed, that was what was found. In the first studies, college students who were asked to write about traumatic experiences were less likely to visit the student health center for illness in the months after writing compared with controls who wrote about superficial topics (Pennebaker & Beall, 1986). Since the original writing studies, dozens of replications have been published demonstrating that emotional writing can influence physician visits, immune function, stress hormones, blood pressure, as well as a variety of social, academic, and cognitive variables.

These effects hold up across cultures, ages, and diverse samples (cf., Smyth, 1998; Lepore & Smyth, 2001; Pennebaker & Graybeal, 2001).

Although putting emotional experiences into words is apparently healthy, one of the biggest challenges has been to find a good explanation for the phenomenon. In the last decade, several mediating factors have been proposed and tested. There is some evidence to suggest that individuals who write clear, coherent stories about their emotional upheavals benefit more than those who do not construct “good” stories (e.g., Pennebaker, Mayne, & Francis, 1997). Others have argued that confronting emotional upheavals allows for habituation to powerful emotional experiences (Greenberg, Wortman, & Stone, 1996). More recently, Klein and Boals (2001) have suggested that writing about an emotional experience helps to bring closure to it thereby freeing up working memory. Other studies are now hinting that one of the effects of emotional writing is that it brings about changes to people’s social lives. That is, after writing about a traumatic secret, the person is now able to talk more openly with friends (Pennebaker & Graybeal, 2001).

One consistent finding in the writing studies is that individuals who have been randomly assigned to write about emotional topics report that the experiment made them think differently about their experiences (Pennebaker, 1989). Is this change in thinking reflected in the ways people write? That is, do people who show health improvements after they write about traumatic or emotional topics write in ways different from those who do not improve? If so, what is the best approach to analyze writing samples to determine healthy from unhealthy writing? The purpose of the current investigation is to address these questions using the text analytic approach Latent Semantic Analysis (LSA). Before detailing our approach, however, it is instructive to provide some background concerning prior attempts to use features of people’s writing samples to predict health changes.

Deductive Strategies to Text Analysis

One of the first strategies used to analyze traumatic writing samples involved subjective ratings by groups of judges. Judges were given the writing samples of several participants whose health either did or did not improve after writing about emotional topics and asked to rate them on several categories. In general, judges blind to health status rated the essays of people who improved as being more self-reflective, emotionally open, and, according to some judges, even "smarter." In fact, there were no differences in the intelligence of the writers based on standardized measures of aptitude or grades. Unfortunately, inter-judge agreement was typically low and not consistent from study to study (Pennebaker, 1997).

To bypass some of the rating problems, we explored some standard text analysis programs. At the time, the only available strategies had been developed in the 1950s and reflected either psychoanalytic traditions (e.g., Gottschalk & Gleser, 1969) or need- or motive-based thinking (Stone et al., 1966). In order to better reflect our thinking about emotional and cognitive processes, we developed our own text analysis program called Linguistic Inquiry and Word Count, or LIWC (the most recent version is Pennebaker, Francis, & Booth, 2001).

The LIWC program uses a dictionary of 2,200 words and word stems that represent approximately 72 text dimensions, the majority tapping a variety of psychologically relevant themes such as causal thinking, use of emotion terms, etc. The words within each of the dictionaries were agreed upon by groups of independent judges over the course of the program's development. For any given text file, then, the LIWC program processes all of the words and calculates the percentage of words that are reflected in each of the dictionaries. For example, the negative emotion dictionary searches for words such as hate, sad, and ugly and calculates the percentage of total words in the text that have been categorized as negative emotion words. Note that this strategy is context free and is ultimately based on the judges' estimates of what words are viewed as negative emotion.

Despite the obvious limitations of the LIWC program, it has proven to be a useful system in analyzing people's emotional writing samples. For example, in a reanalysis of emotion essays from 6 previous writing studies, it was found that the use of positive emotion words relative to negative emotion words together with an increasing use of causal and insight-related words over the days of writing was associated with modest improvements in health (Pennebaker & Francis, 1996; Pennebaker, Mayne, & Francis, 1997). Other studies have found that the use of particular LIWC categories that tap people's ability to make distinctions is associated with better health habits among large groups of people writing about college-related topics (Pennebaker & King, 1999).

Although promising, the LIWC strategy is limited in providing us with information about the relationships among the various writing samples provided by participants. In many ways, the interpretation of any LIWC findings is based on judges' conceptions of what words are intuitively related to particular broader categories. By the same token, the LIWC system does not take into account how categories of words or even the words within any given category are inter-related. Because there are so many judge-related categories, it is difficult to predict which categories should be related to any given outcome. Because of these limitations – together with fact that the LIWC approach accounted for a relatively small amount of variance related to health changes – we adopted a very different strategy associated with the use of Latent Semantic Analysis, or LSA.

An Inductive Approach to Text Analysis: LSA

LSA is set of computerized text analysis tools. It is useful to consider LSA as a two stage process: first, a training stage, then a comparison stage. For a more detailed introduction, see Landauer, Foltz, & Laham, 1998, Landauer & Dumais, 1997, and Foltz, Kintsch, & Landauer, 1997. The training stage begins with the assumption that words are grouped into writing samples based on some underlying structure. That is, words do not appear in writing samples at random. For this paper, the writing samples referred to are participant essays. Given a large number of writing samples, it should be possible to identify the underlying dimensions. The first step in the training stage is to assemble a large number of appropriate writing samples, a training corpus. Because people use words differently in differing contexts and across different topics, the choice of writing samples to analyze can have a tremendous impact on the underlying dimensions that are computed. Imagine two groups of writing samples, one consisting of writings about experiences from grade school, and another consisting of writings about classical conditioning. The words included in these two groups of writing are likely to be very different. In the essays about grade school, some people might focus on their teachers' hard work or indifference. Others might emphasize the features of their normal class day. Across writing samples, however, it is likely that a group of common words may emerge. People who use the word "pencil" might also use words like "arithmetic," "desk," "recess," or "bell."

Imagine the different word patterns that are likely to emerge across writing samples in the classical conditioning group. People who use the word "bell" are likely to use the word "dog" or "saliva." Not only are the groups of writings likely to include different words, the words that are common to both groups are likely to be associated differently. The word usage patterns surrounding the word "bell" in the writings about grade school are hardly likely to include the word "saliva." These differing patterns of word usage are presumably the result of different underlying dimensions, so training LSA on different kinds of corpora will result in different computed underlying dimensions.

Once an appropriate corpus is assembled, a co-occurrence matrix is constructed. A co-occurrence matrix consists of one row for each word used anywhere in the training corpus and one column for each writing sample in the corpus. The cells of the matrix contain frequencies, or how many occurrences of each word there were in each of the writing samples. This co-occurrence matrix consists mostly of zeroes. After all, each writing sample contains only a few hundred words, with many repetitions, while it does not contain all the other words used in the rest of the corpus. LSA inductively derives the underlying dimensions from this co-occurrence matrix using Singular Value Decomposition (SVD), a powerful mathematical tool closely related to Factor Analysis. The SVD derives the underlying dimensions and numerically indicates the strength of the relationship of each word and writing sample to each of the underlying dimensions. The resulting information is collectively referred to as a semantic space.

The second stage of LSA uses the semantic space to make comparisons between items of interest based on the inductively derived dimensions. A writing sample that was not included in the original training corpus can be treated as simply a group of words. Recall that the information distilled about an individual word is represented as numeric relationships to underlying dimensions. LSA represents new groups of words (writing samples) using information from the semantic space. Averages are calculated for the relationships to each underlying dimension for all the words in a new writing sample. That is, based on all the words found in a writing sample, an average relationship to each underlying dimension is calculated, and an entire writing sample is represented as numeric relationships to the underlying dimensions. In this way, any two writing samples can be compared to one another based on their relationships to inductively derived dimensions from a particular semantic space.

Consider the implications of an approach like LSA. First, it is not dependent on predetermined language categories, or for that matter, any particular language. Second, it allows a researcher to mathematically compute the degree to which a writing sample is similar to any other sample. It even allows the comparison of two parts of the same sample or, for that matter, adjacent sentences within the same sample. In short, LSA can allow the researcher to mathematically assess the degree to which sentences or entire writing samples are coherent – that is, using words that naturally covary. In other words, LSA makes it possible to compare the overall similarity of writing samples within subjects.

As with factor analysis, decisions regarding the raw data for analysis can have a significant impact on the results using LSA. Two of the major decisions are 1) the choice of writing samples to include in the training corpus and 2) choice of words to focus on.

<u>Defining semantic spaces.</u> The writing samples included in a training corpus can have a profound effect on the underlying dimensions computed. The analyses presented in this investigation are all based on a large group of writing samples chosen to represent informal, diary-style writing. Because the participants' essays to be studied included traumatic writings (experimental condition) and control writings, a large number of similar writing samples were collected. Also added to the collection were a number of stream of consciousness writing samples (cf., Pennebaker & King, 1999). The authors of these samples were not restricted to any particular topic, and tended to write in ways that varied between traumatic topics (upsetting experiences that happened to be in their thoughts at that moment) to more "control-like" samples that described their surroundings or thoughts during the assignment. These writing samples were

gathered from labs across the United States, England, and New Zealand, from a variety of investigators, spanning approximately the last 10 years. Overall, 7,501 writing samples were collected for a total count of 3,445,940 words. There were 31,320 unique words (used at least once), in the corpus.

Exploring writing content versus style. The second major decision involves word choice. Different kinds of words carry different kinds of information. LSA provides facilities to choose words based on their frequency in the original training corpus. LSA investigations commonly focus on the topic of writing samples, or the content. This default approach could be used to address the question, “Is the topic across a participant’s essays related to health changes?” It is conceivable that a person who writes several times about the same or similar topics would have more opportunity to deeply process their traumatic memory and could show more health improvement than a person who changed topics more often. Content, or topic words are words that are relatively uncommon in the training corpus.

LSA has been used in this content-focused manner in a variety of interesting psychological applications (Landauer, Foltz & Laham, 1998). There are presently several projects using information from LSA analyses to automatically assign grades to essays. LSA is remarkably good at grading essays; across several studies, LSA grades correlated with expert graders as well as expert graders did with one another (also in Landauer, Foltz & Laham, 1998). Another application is a computerized tutoring program called Auto Tutor (Graesser, et al, 2000, Graesser, et al, in press.) Auto Tutor is designed to mimic the tutoring style of non-expert tutors, such as students, while training participants in a particular subject. This interactive approach uses LSA to analyze participant responses and match them to a database of knowledge. LSA allows participants to use their own words when communicating information while still assessing the similarity to the knowledge database.

It should be emphasized that very common words (pronouns, prepositions, articles) are actively excluded from traditional LSA projects. Since these words occur in most every paragraph, researchers have assumed that these high frequency commonly-used words cannot help differentiate samples by topic, so are treated as "noise" and excluded in the creation of the co-occurrence matrix. In a sense, the use of these uncommon LSA words reflect the content of writing. Pronouns, prepositions, and other very common words can be considered to carry style information. This would include information about time (tense of common verbs), whether a sample is self or other focused (pronouns), or the level of concreteness (use of articles). An important feature of the present project, then, will be to compare LSA approaches for traditional content-oriented semantic spaces with style-oriented spaces.

Overview of the Current Project

The current project sought to explore LSA as a tool to determine if features of writing about emotional topics could predict health changes. Based on earlier studies, we hypothesized that the ways people changed in their writing from day to day could predict health improvements. The first question was simply to learn if a change in writing content from day to day was related to health markers. As outlined below, this question evolved into an exploration of changes in writing styles. Assuming that writing styles were indeed important, we gradually honed in on which of several features of writing styles accounted for most of the variance in health improvements.

Methods

Writing samples from three previously published studies were reanalyzed using a variety of LSA approaches. These studies were chosen to provide a variety of ages, backgrounds, geographic locations and experimental instructions. Overall, the participants in the experimental conditions demonstrated health improvements relative to participants in the control conditions. Participants in all studies were randomly assigned to condition.

Participants

First year students (Pennebaker, Colder & Sharp, 1990). Seventy-four first year undergraduate students enrolled in Introductory Psychology (35 male, 39 female, mean age 17.9 years, SD 0.4) were instructed to write for 3 days about coming to college (n=35) or, for control participants, about descriptions of their day or of a social event recently attended (n=39). The sample was based on those participants who completed all questionnaires at baseline and at follow-up. Health center visits for illness were collected from the university health center and were compared for the two months prior to writing with the four months after writing. Visits were converted to mean number of visits per month for the two time periods.

Upper division students (Pennebaker, Kiecolt-Glaser, & Glaser, 1988). Fifty undergraduates (14 males, 36 females, mean age 19.8 years, SD 2.6) were randomly assigned to write for 4 days about the most traumatic events of their lives (n=25) or about superficial topics (n=25) for 20 minutes each day. Health center visits for illness for the two months before and two months after writing were converted to mean number of visits per month for the two time periods.

Psychiatric prison inmates (Richards, Beal, Seagal, & Pennebaker, 2000). Fifty-nine male maximum security psychiatric inmates (mean age 35.4 years, SD 9.5) from a prison in the Midwest were randomly assigned to write for three days about either traumatic experiences (n=33) or superficial topics (n=26) for 15 minutes per day. Participants were screened to ensure a minimum of a sixth grade education (mean education 12.3 years, SD 2.4). Number of infirmary visits in the two months before and after writing served as the dependent measure. Of the participants, 19 of the inmates were serving terms for sex-related crimes (e.g., rape, child molestation) whereas the rest were being held on non-sex-related violent crimes (e.g., murder, robbery). Relevant to the current study, only those trauma participants who had committed sex-related crimes evidenced improvements in health.

Analytic Procedure

Recall that LSA represents a particular “bag of words” as a set of numeric relationships to underlying concepts. Any two “bags of words” (essays, in this case) can be compared, since both essays are represented in relation to the same underlying concepts, much in the same way that means can be compared by representing them as standardized scores. When means are compared based on the same metric, it is a simple matter to assess similarity. LSA builds a semantic space to be used as a common metric, to allow different essays to be compared based upon that metric.

Semantic spaces commonly contain several hundred underlying concepts or dimensions. Comparisons in such spaces are accomplished by computing the cosine of the two sets of numeric relationships. A cosine is conceptually related to a correlation, in that both are measures of pattern similarity that vary from –1 to 1. A correlation measures degree of relationship, or pattern similarity across subjects, even if the two sets of numbers are originally measured using different scales. A cosine can be used if both sets of numbers share a common metric, and is insensitive to the overall variability of those numbers. A cosine is usually used to compare pattern similarity across variables, in this case, between two essays. LSA uses a cosine to compare how similar two essays are across all the underlying dimensions of a particular semantic space. As mentioned above, each semantic space computes different underlying dimensions, depending on the training corpus and the choices of words to include in the space.

A variety of semantic spaces were created using the training corpus of 7501 essays to explore the relationships of different groups of words to health outcomes. Writing samples were analyzed within subjects. The first sample was compared to the second, the second to the third, and, for the upper division students, the third essay was compared to the fourth. The resulting LSA similarity ratings were averaged to provide an overall similarity rating for each participant. As described below, this process was repeated for each of the various semantic spaces created, i.e. Content, Style, or Particles. Note that none of the essays analyzed in this investigation were used in the training corpus (7501 essays) to create the semantic spaces.

It is important to appreciate the nature of the health outcome measures for the three studies. In all three experiments, researchers obtained permission from participants to track medical visits for illness from the student health center (studies 1 and 2) or the prison infirmary (study 3). An illness visit was based on the medical facility diagnosis of a complaint that had a presumed illness cause. Injuries, check-ups, and non-infectious OB/Gyn visits were not counted. For each participant, the mean monthly physician visits after writing was subtracted from the mean visits in the months prior to writing. The change in doctor visits, then, was ultimately correlated with the similarity ratings between essays. It is important to note, then, that the illness difference measure is scored so that the higher the number, the more health visits after writing.

Results

Because of the exploratory nature of the project, the results of the data analyses are presented as a function of separate LSA-related questions.

<u>Do similarities in the content of writing predict health change?</u>

Recall that the default approach for LSA focuses on content words that are relatively uncommon. Indeed, the default approach weights words such that the less frequently they are used, the higher their weights. Figure 1A includes part of a writing sample from a participant in the emotion writing condition from the first year student study. Directly below, as part of Figure 1B, the words that are recognized as part of the Content semantic space are highlighted.

Insert Figure 1 about here

The Content semantic space recognizes 19,013 unique words. Recall that the entire training corpus contains 31,320 unique words. A Content (default) space ignores words that appear only one time in the training corpus. More important, words that are very common are also excluded. The Content space is capable of recognizing 60.7% of the unique words (the vocabulary) in the training corpus. Overall, this approach recognizes 30% of the over 3.4 million total words in the training corpus. As is apparent in Figure 1B, the Content semantic space recognizes words that communicate the gist of the writer's topic fairly well.

The initial LSA strategy involved reducing the training corpus of 7501 essays to 276 independent factors, or dimensions. This step of reducing the co-occurrence matrix to an appropriate number of underlying dimensions is important. Previous work with LSA often chooses this appropriate number in an empirical manner, by reducing the matrix many times, using a range of dimensions, and choosing the number of dimensions that is most effective for the task at hand. This number is usually around 300 dimensions. These dimensions served as the basis by which we were able to compute similarity ratings between pairs of essays for each of the three studies. So, for example, to determine the degree of similarity or coherence among the possible pairs of adjacent essays (e.g., essay one with essay two and essay two with essay three), separate similarity coefficients were computed and averaged for each participant. The Content similarity coefficient, then, served as an index of the degree to which the content words of the essays were mathematically similar over the days of writing.

For each study, the content similarity coefficient for each trauma writing participant was correlated with changes in physician visits in the months after writing compared with the months prior to writing. As can be seen in the top row of simple correlations in Table 1, the Content similarity coefficients were statistically unrelated to health changes in all three studies. The results from the Content semantic space, then, indicate that what individuals are writing about from day to day is unrelated to health. That is, participants were equally likely to benefit from writing if they wrote about the same general topics or very different topics each day of the study.

Insert Table 1 about here

Do changes in writing style affect health?

Although LSA was developed to study linguistic content, it can, in theory, be modified to explore linguistic styles. By linguistic style, we mean the ways people express themselves rather than what they are saying per se (cf., Pennebaker & King, 1999). Recall that words carrying style information (such as prepositions and pronouns) are among the most commonly used words. Our first approach to exploring the linguistic bases of style was to develop a semantic space that recognized the most common words in the training corpus. A word was chosen for inclusion in this Style semantic space if it occurred 500 times or more in the training corpus.

Overall, 611 words met this criterion, accounting for only 2% of the total vocabulary. This 2% of the possible vocabulary accounts for 83.8% of the total words used.

As with the Content space, the global Style space was reduced in the 7501 essays to 26 factors which were then used as the basis by which to compare adjacent essays for participants in each of the three writing studies. This space was reduced to considerably fewer dimensions than the Content space. One way of thinking about dimensions is to consider each as a different type or dimension of style. It seemed appropriate that there should be an order of magnitude fewer "styles" than there are "kinds" of content. As can be seen in the second row of Table 1, the mean similarity ratings from essay to essay were consistently and positively related to the physician change measures across all three studies. That is, the more similar that people's writing styles were from day to day, the more likely they were to visit physicians for illness in the months after writing. Conversely, those participants who switched in their writing styles demonstrated improvements in health after writing.

What accounts for style?

The results from the style semantic space are striking from several perspectives. First, contrary to a coherence argument, the more that people changed in their writing styles, the more their health improved. Second, the magnitude of effects across all three studies was striking and far more impressive than any other strategy we have ever undertaken. Third, although the results are impressive, it is not entirely clear what the words that we have deemed style-relevant truly are.

The style semantic space was constructed based on the most commonly used words in the corpus. Closer inspection of these 611 words suggested that the words with the highest loadings were a broad class of words referred to as particles (cf., Miller, 1996). Particles consist of prepositions, conjunctions, articles, and pronouns. Although not officially listed as particles, some researchers (including us) include auxiliary verbs in this category. Particles, or function words, are important because they link phrases, clauses, and other grammatical structures together. They can also serve as linguistic shortcuts to help identify relationships between the speaker and other individuals and objects.

Particles are intriguing for several reasons. They are among the most commonly used words in English. Although there are fewer than 200 common particles, they account for over 55% of the words we have found in our archive of essays, natural speech, and literature of over 13 million words (Pennebaker, Francis, & Booth, 2001). Given the disproportionate influence of particles in the Style semantic space, we embarked on a series of analyses on our writing samples beginning with a Particle semantic space and subsequently created separate spaces consisting of subsets of words from the Particle semantic space.

Particle semantic space. The Particle semantic space recognizes 172 words made up of the most common prepositions, conjunctions, articles, pronouns, and auxiliary verbs in the training corpus. These 172 words accounted for 0.5% of the vocabulary of the training corpus, but account for 59.6% of the total words used. The particle semantic space was reduced to 19 factors that served as the basis of the similarity coefficients. The paragraph in Figure 1C highlights the words recognized by the Particle semantic space.

As can be seen in the third row of Table 1, the Particle semantic space was indistinguishable in both direction and magnitude of results found earlier with the overall Style semantic space. That is, for all three studies, the more that individuals varied in their patterning of use of particles from essay to essay, the more their health improved from before to after writing. Our next goal was simply to distinguish which category of particles may have been accounting for most of the variance.

Prepositions, Articles, Conjunctions, and Auxiliary verbs. Separate semantic spaces were created for the most commonly used prepositions (based on 31 words), the combined group of articles and conjunctions (29 words), and auxiliary or common helper verbs (44 words). Similarity coefficients were again computed for the written essays. As is apparent in Table 1, none of these dimensions were consistently related to health improvements in the studies.

Pronouns. The pronoun semantic space consists of the 19 most common pronouns in the training corpus. These words, in decreasing order of frequency, are: I, my, it, you, me, she, he, her, we, they, your, him, his, them, our, myself, their, us, its. Pronouns account for only 0.06% of all the vocabulary words in the training corpus, but account for 14.9% of the total words used.

In the experimental conditions, the pattern of association between similarity coefficients and change in doctor visits in the pronoun semantic space is almost identical to that displayed by the larger style and particle semantic spaces. Traumatic writing (experimental) participants with low similarity ratings went to the doctor less after the writing intervention than did participants with high similarity ratings.

What do pronouns tell us?

Consider the findings so far. The ways individuals use 19 pronouns over the three to four days of writing about emotional experiences predict ultimate changes in their physical health. More specifically, those participants who change in the ways they use pronouns from essay to essay tend to visit physicians less. That the effect is consistent and significant across three studies with very different samples is, in itself, surprising.

Interestingly, most LSA experts do not explore the factors that make up the vectors that underlie their semantic spaces. Indeed, in traditional LSA analyses, it is not uncommon to have over 300 factors each with a smattering of words with moderate to high loadings. With our pronoun semantic space, however, it is possible to explore differing factor structures to see what the minimum number would be that would continue to predict health changes. To do this, a series of pronoun semantic spaces were computed that were based on 19 (the original Pronoun semantic space), 8, 6, and 2 factors. Separate analyses computed similarity coefficients for each space which were then correlated with physician visits separately for each study.

Remarkably comparable results were found for all pronoun semantic spaces. Most striking were the findings for the two-dimensional semantic space. The two factor solution displayed strong correlations for the first semester and upper division samples ($\underline{r}(35) = .44$, $\underline{p} = .01$; and $\underline{r}(25) = .47$, $\underline{p} = .02$, respectively), but failed to yield significant results for the prison sample. However, when we look separately at the sex offenders in that sample (who were the only ones to show health improvements in the original study), the two-factor solution does approach significance, $\underline{r}(19) = .34$, $\underline{p} = .15$.

Because of the surprising success of the two-factor solution, we are able to examine the two dimensions of factor loadings to get a sense of what dimensions are relevant for health improvement. In LSA, the items on the first factor are always positively loaded and closely mirror the usage frequency of the words. For the two factor semantic space, the factor weights for the most heavily weighted words were: I (.92), my (.28), it (.20), and me (.12), with the remaining 15 pronouns loading less than .07. The second factor is more intriguing in that only one item was modestly positive loaded, I (.17) and the following were negatively loaded: she (-.58), her (-.53), he (-.35), it (-.30), you (-.23), his (-.17), him (-.14), they (-.14), with the remaining pronouns loading from -.08 to + .03. As is apparent, then, the first factor is essentially controlled by first person singular and the second by references to other people with the effects of "I" subtracted out.

Keep in mind that LSA determines the similarity of two essays by comparing the relative scores on the (two) factors for the two essays[1]. Further, this difference is not expressed in direction – only in the degree of similarity/change between the essays of interest. Interestingly, separate analyses on the two factor scores themselves for each essay did not reveal simple correlations with health changes. That is, the loadings on an individual factor were not related to the observed health changes, both factor loadings need to be taken into account. To repeat, the results do not indicate any directional change in use of pronouns, such as moving from a self oriented perspective to another more inclusive orientation. Nor do these findings suggest that changes in the sheer number of pronouns used are related to health changes.

An alternative strategy has been to simply go back to the original essays and to look at the essays associated with individuals' health improving versus not improving. Figure 2 includes the writing samples of adjacent days of writing for two participants -- with Figure 2A reflecting the writing of a student who did not show health benefits and Figure 2B of someone from the prison study who did improve. Note that the pronouns for the two essays are highlighted. Although the person in Figure 2A appears to be writing in a self-reflective mode on both days, it is clear that the person's use of pronouns (especially first person singular) is indistinguishable for the two days. The Figure 2B person, on the other hand, evidences a striking shift from day one to day two, with the second essay reflecting a very different way of writing and thinking.

Insert Figure 2 about here

Other Analyses

In addition to the results described above, a large number of additional analyses were conducted as well. The results are summarized below:

Control group analyses. Recall that in each of the writing studies, control groups wrote for three to four consecutive days about superficial, non-emotional topics. Separate LSA analyses were conducted using each of the semantic spaces summarized in Table 1. In no case did the similarity scores correlate significantly with changes in health center visits.

Sequential versus global measures of similarity. The results reported in this paper correlated a measure of overall similarity of essays with a health improvement measure. This overall similarity measure was computed by averaging the similarity ratings of sequential essays, that is, essays adjacent to one another in time. Another measure of similarity was also computed, including all possible comparisons of essays, or a global similarity measure. The sequential measure averaged the similarity of essay one to two, and two to three. The global measure averaged the similarity of one to two, two to three, and one to three. This global similarity measure was computed for all of the above semantic spaces. Overall, the pattern of correlations was virtually identical to the sequential measure (although the correlation for the first year sample was only marginally significant, $r(35) = .26$, $p = .13$).

Linear change model. Previous analyses of essays have taken an approach testing a linear change model. In these analyses, descriptors of essays from LIWC were tabulated and difference scores reported. For example, subtracting the level of emotionality in the first essay from the emotionality score for the last essay tries to assess a linear change in these essays. A conceptually similar measure using an LSA approach would be to correlate the similarity coefficient of first to last essay with the health assessments. This comparison did not reveal any significant correlations.

Meta-analyses of the three studies. Meta-analytic procedures were conducted to assess the consistency of effects across the differing participant groups. Across the three studies, the effect sizes for the Pronoun semantic space were not significantly different from one another, failing a test for heterogeneity of effect size, $\chi^2(2) = 0.46$, $p = 0.80$. Additionally, the significance levels of effects across study were not significantly different from one another, failing a test of heterogeneity of significance levels, $\chi^2(2) = 0.16$, $p = 0.92$. If the null hypothesis (no association between similarity of pronoun use and health benefits) were true, the probability of observing this pattern of results is $p < .0001$. Overall, remarkably consistent results were observed. The combined effect size of these three studies, weighted by degrees of freedom, is (Cohen's) $d = 1.15$.

Discussion

A traditional LSA analysis revealed that the content of participants' essays was unrelated to observed health changes. However, an LSA analysis focusing on words that have been traditionally ignored revealed a strong and stable relationship to health. Further analyses using more and more restricted groups of words revealed that health could be predicted by a two factor solution analyzing only 19 common words, personal pronouns.

The results of the present investigation are important for two reasons. First, the analyses help to clarify the role of pronouns and, indirectly, people's thinking about their social worlds over the course of their writing about emotional topics. Across the three to four days of writing, the change in references to self and others predicts health improvements. Second, the study points to the potential power of LSA and comparable techniques in providing insight into the ways people think. Although LSA has traditionally been used as a method by which to construct networks of nouns and regular verbs, the current study provides compelling evidence that the "junk" words that we typically use in writing and speech reveal a tremendous amount about how we are thinking.

That changes in pronoun use predict health is a rather bold and, at the same time, enigmatic statement. The LSA approach is based on the assumption that constellations of pronouns vary in their usage from writing sample to writing sample. Unlike traditional correlational or analysis of variance strategies, LSA is an idiographic approach. This is important to note because the natural tendency of many researchers is to search for a common pattern of effects across participants. For example, the author of the essays in Figure 2B wrote about self in the first essay, and changed to writing much more about others in the second. It is tempting to conclude that it is healthy to move from writing about the self to writing about others. However, the overall pattern of data tells us that this author would likely have benefited just as much if he had begun writing about others and shifted to writing about himself. The LSA results suggest that a flexible pattern of pronoun use is healthy and that a rigid style is less healthy. How the pronouns change is not addressed by this kind of LSA analysis.

On both a theoretical and clinical level, it is interesting to speculate why changes in pronoun use may ultimately be beneficial. Translating a traumatic event into language calls on cognitive, emotional, and linguistic processes, among them, introducing and describing main characters, contexts and events (people, places, and things). Once introduced, the interrelationships of these components need to be described. Pronoun choice communicates this relational information. Pronouns are generally seen as placeholders in language, simply referring to components previously introduced (people, places, and things). A quick glance at the essays in Figure 2 reveals that attending to pronouns in the absence of their referents allows this relational information to come to the foreground. Using the Pronoun semantic space in LSA analyzes essays in just this decontextualized way, comparing a particular "bag of pronouns" with another. The relational world communicated by essay B in Figure 2B is clearly different than the relational world in essay A.

Consider the following (simplified) bags of pronouns, each referring to an author and two other people: (I, he, she), (I, they), (I, we), (we, him), (we, us). Even without any other information, each cluster of pronouns defines a definite relational/social world, grouping self and others in a variety of ways. This delineation of social relationships is a basic part of translating experience into language that occurs, for the most part, outside of conscious awareness. An essay using the first cluster of pronouns above (I, he, she), describes a social world in which self and others are all separate, contrasting with the last cluster (we, us), evoking a unitary world in which not even the self is differentiated.

It is not coincidental that virtually all traumatic experiences that are written about in our studies are ultimately social. In the rare instances where the trauma was not caused by other people (e.g., an isolated swimming accident), the event still had tremendous social consequences. Coming to terms with a traumatic experience appears to be intimately linked to thinking about oneself in relation to others (cf., Kohut, 1971; Ogilvie & Ashmore, 1991; Swann, 1997). The LSA analyses have starkly demonstrated that different clusters of pronouns describe different social realities or different lenses through which our participants see their worlds.

It is also interesting that health improvements are associated with changes in pronoun constellations rather than fixed constellations per se. Further, these changes are not linear in the sense of emerging from one dimension on the first day to a different one on the last day of writing. Rather, the mere shifting of perspectives from day to day is most predictive. A clear prediction from these findings is that explicitly encouraging participants to change the ways they are thinking about their topics in relation to who they are may produce stronger health effects.

This project has also been important in demonstrating another use for LSA. To date, many researchers have tended to think of LSA as a system that represented the way cognitive and/or semantic information was organized in the mind (Landauer & Dumais, 1997). This approach has assumed that the interesting linguistic action associated with thinking and knowledge resides in nouns, regular verbs, and exotic modifiers. This study demands that more respect be given to particles in general and pronouns in particular.

As a class of words, particles convey virtually no content but define the ways people express themselves. These words are used to communicate perspective and are an important factor in studying conversation (for examples see Schober, 1998, and Schober, 1993.) Indeed, in some of our research relying on LIWC, we have consistently found that articles, prepositions, pronouns, conjunctions, and auxiliary verbs have been related to individual differences such as self-esteem and other trait markers (Pennebaker & King, 1999), depression and suicide proneness (Stirman & Pennebaker, in press), as well as demographic variables such as sex and age (Pennebaker, 2001). Only through the LSA approach have we begun to appreciate the potency of writing style in addition to content.

Footnotes

1. This description is somewhat misleading. In fact, the similarity score with the two factor solution is actually computed by determining the angular difference between the essays in two dimensional space. By way of example, consider a two-dimensional plot where scores on factor one are along the x-axis and scores for factor two are long the y-axis. Any given essay, then, can be plotted as a point in this space using its two factor scores. LSA judges similarity based on the angle formed between each essay point and the (0, 0) point on the graph, also considering the distance of the two essays (points) from the origin. Similarity is the cosine of the angle, which ranges from zero to one in this example (since factor one is always positive). In general, the greater the angle separating essay one and two, the lower the cosine, and the less similar the essays.

References

Cobb, S. (1976). Social support as a moderator of life stress. Psychosomatic Medicine, 38, 300-314.

Foltz, P. W. Kintsch, W. & Landauer, T. K. (1998). The measurement of textual coherence using latent semantic analysis. Discourse Processes, 25, 285-307.

Gottschalk, L. A., & Gleser, G. C. (1969). The measurement of psychological states through the content analysis of verbal behavior. Berkeley: University of California Press.

Graesser, A., Wiemer-Hastings, K., Wiemer-Hastings, P., Kreuz, R., & the Tutoring Research Group (2000). AutoTutor: A simulation of a human tutor. Journal of Cognitive Systems Research, 1, 35-51.

Graesser, A., Wiemer-Hastings, P., Wiemer-Hastings, K., Harter, D., Person, N., & the Tutoring Research Group. (in press). Using Latent Semantic Analysis to evaluate the contributions of students in AutoTutor. Interactive Learning Environments.

Greenberg, M. A., Wortman, C. B. Stone, A. A. (1996). Emotional expression and physical heath: Revising traumatic memories or fostering self-regulation? Journal of Personality & Social Psychology, 71, 588-602.

Holmes, T. H. & Rahe, R. H. (1967). The Social Readjustment Rating Scale. Journal of Psychosomatic Research, 11, 213-218.

Klein, K., & Boalls, A. (in press). Expressive writing can increase working memory capacity. Journal of Experimental Psychology: General.

Kohut, H. (1971). The analysis of the self. New York: International Universities Press.

Landauer, T. K., & Dumais, S. T. (1997). A solution to Plato's problem: The latent semantic analysis theory of acquisition, induction and representation of knowledge. Psychological Review, 104, 211-240.

Landauer, T. K., Foltz, P. W., & Laham, D. (1998). An introduction to latent semantic analysis. Discourse Processes, 25, 259-284.

Lepore, S. & Smyth, J.M. (Eds.) (2001). The writing cure: How expressive writing promotes health and emotional well-being. American Psychological Association press: Washington, DC.

Miller, G. A. (1996). The science of words. Scientific American Library: NY, NY.

Ogilvie, D.M., & Ashmore, R.D. (1991). Self-with-other representation as a unit of analysis in self-concept research. In R.C. Curtis (Ed.), Theoretical convergences in psychoanalysis and social psychology (pp 282-314). New York: Guilford Press.

Pennebaker, J.W. (1989). Confession, inhibition, and disease. In L. Berkowitz (Ed.), Advances in experimental social psychology (Vol. 22, pp. 211-244). New York: Academic Press.

Pennebaker, J. W. (1997). Writing about emotional experiences as a therapeutic process. Psychological Science, 8, 162-166.

Pennebaker, J.W. (2001). Words of wisdom: Language use over the lifespan. Manuscript submitted for publication. University of Texas, Austin.

Pennebaker, J. W., Beall, S. K. (1986). Confronting a traumatic event: Toward an understanding of inhibition and disease. Journal of Abnormal Psychology, 95, 274-281.

Pennebaker, J. W., Colder, M., & Sharp, L. K., (1990). Accelerating the coping process. Journal of Personality and Social Psychology, 58, 528-537.

Pennebaker, J. W., & Francis, M. E. (1996). Cognitive, emotional, and language processes in disclosure. Cognition and Emotion, 10, 601-626.

Pennebaker, J. W., Francis, M. E., & Booth, R. J. (2001). Linguistic Inquiry and Word Count (LIWC): LIWC2001. Mahwah, NJ, Erlbaum Publishers.

Pennebaker, J. W. & Graybeal, A. (2001). Patterns of natural language use: Disclosure, personality and social integration. Current Directions, 10, 90-93.

Pennebaker, J. W., Kiecolt-Glaser, J. K., & Glaser, R. (1988). Disclosure of traumas and immune function: Health implications for psychotherapy. Journal of Consulting and Clinical Psychology, 56, 239-245.

Pennebaker, J. W., & King, L. A. (1999). Linguistic styles: Language use as an individual difference. Journal of Personality and Social Psychology, 77, 1296-1312.

Pennebaker, J. W., Mayne, T. J. & Francis, M. E. (1997). Linguistic predictors of adaptive bereavement. Journal of Personality and Social Psychology, 72, 863-871.

Richards, J. M., Beal, W. E., Seagal, J. D., & Pennebaker, J. W., (2000). Effects of disclosure of traumatic events on illness behavior among psychiatric prison inmates. Journal of Abnormal Psychology, 109, 156-160.

Schober, M.F. (1993). Spatial perspective-taking in conversation. Cognition, 47, 1-24.

Schober, M.F. (1998). Different kinds of conversational perspective-taking. In S.R. Fussell & R.J. Kreuz (eds.), Social and cognitive psychological approaches to interpersonal communication, p. 145-174. Mahwah, NJ: Lawrence Erlbaum.

Silver, R. L., Boon, C., & Stones, M. (1983). Searching for meaning in misfortune: Making sense of incest. Journal of Social Issues, 39, 81-102.

Smyth, J. M. (1998). Written emotional expression: Effect sizes, outcome types, and moderating variables. Journal of Consulting and Clinical Psychology, 66, 174-184.

Stirman, S. W. & Pennebaker, J. W. (in press). Word use in the poetry of suicidal and non-suicidal poets. Psychosomatic Medicine.

Stone, P. J., Dunphy, D. C., Smith, M. S., & Ogilvy, D. M. (1966). The General Inquirer: A computer approach to content analysis. Cambridge, MA: MIT Press.

Swann, W.B. Jr. (1997). The trouble with change: Self-verification and allegiance to the self. Psychological Science, 8, 177-180.

Table 1

Correlation of essay similarity to change in doctor visits for experimental (traumatic writing) participants.

	first semester n = 35	upper division n = 25	psychiatric inmates n = 33
Content	-.05	.08	.25
Style	**.34 ***	**.51 ****	**.43 ***
Particles	**.38 ***	**.51 ****	**.41 ***
Prepositions	.20	.32	.14
Conjunctions, articles	.18	.38	.08
Irregular verbs	-.22	.10	.22
Pronouns	**.35 ***	**.50 ****	**.43 ****

* $p \leq .05$

** $p \leq .01$

Figure 1

A. Original

Coming to college conjured up these feelings. Excitement, anxiety, happiness, worry, anticipation, glee, nervousness, sadness, grief, energetic and many others. Most of all I felt very excited I could not wait to get away from anything having to do with high school. I hated high school. Any possible way I could get away from my high school I would do it. I waited to go to another high school all throughout my high school career. Mine never satisfied me. Going to college was the ultimate escape for me. It gave me a fresh start in a new world where I could make a name for myself. A new name which would be mine for the rest of my life. I was somewhat anxious to find out exactly what this name would be for myself.

B. Content Semantic Space

Coming to college conjured up these feelings. Excitement, anxiety, happiness, worry, anticipation, glee, nervousness, sadness, grief, energetic and many others. Most of all I felt very excited I could not wait to get away from anything having to do with high school. I hated high school. Any possible way I could get away from my high school I would do it. I waited to go to another high school all throughout my high school career. Mine never satisfied me. Going to college was the ultimate escape for me. It gave me a fresh start in a new world where I could make a name for myself. A new name which would be mine for the rest of my life. I was somewhat anxious to find out exactly what this name would be for myself.

C. Particle Semantic Space

Coming to college conjured up these feelings. Excitement, anxiety, happiness, worry, anticipation, glee, nervousness, sadness, grief, energetic and many others. Most of all I felt very excited I could not wait to get away from anything having to do with high school. I hated high school. Any possible way I could get away from my high school I would do it. I waited to go to another high school all throughout my high school career. Mine never satisfied me. Going to college was the ultimate escape for me. It gave me a fresh start in a new world where I could make a name for myself. A new name which would be mine for the rest of my life. I was somewhat anxious to find out exactly what this name would be for myself.

Figure 2

A. More visits

Essay A

Last night **I** had 50 thousand things on **my** mind, so many that **I** couldn't even focus on one. **It** was strange. **I** couldn't sleep so **I** got up, turned on the light, and smoked a cigarette thinking **it** would relax **me** so **I** could sleep although **I** know nicotine is a stimulant. **I** was so mad that **I** couldn't sleep because **I** knew **I** was only going to get 4 hours of sleep. **I** ended up getting 3. **My** major is pre-med and **I** spent 4 hours today looking through a microscope at a cell of the intestine. Usually **I**'m really interested in this type of thing, but today **I** was so frustrated **I** couldn't think straight. **I** think sleep would do wonders. **I** probably sound stressed too. **I**'ve been very pessimistic lately, and **I**'m usually a very optimistic person. **I** didn't use to get very depressed.

Essay B

I know or **I** feel that **I** can do anything **I** put **my** mind to if **I** really try and really dedicate **myself**, but with so many obstacles in the way, ie stress, lack of sleep, **it** sometimes seems so unreachable. **I** want probably what everyone else in the world wants. **I** want a good career (physician), to be financially independent (**I** learned early that 1 out of 2 marriages do not work), friends and eventually a happy, healthy family. **I** have to admit **I**'m scared to get married though. **I** mean **I** love boyfriends, **I**'ve had a boyfriend boyfriends since the 7th grade, and **I** enjoy the attention, companionship, love, but marriage is an abstraction to **me** in a way. **It** is such an enormous thought to actually give a part of yourself wholly, all the time, to another. That didn't come out right. **I** really can't explain **it** on paper.

B. Fewer visits

Essay A

I can remember as a small boy **I** was always afraid of the dark and no matter how hard **I** tried not to be **I** just couldn't help **myself**. **I** would go to **my** mother and father's room looking for love and comfort only to be beaten and thrown into a dark closet. **I**'d cry and be even more afraid because **I** couldn't come out for fear of being beat worse, but sometimes if **I** stayed quiet just long enough **I** could crawl out and sleep at the end of **their** bed where **I** would feel safe. **I**'m not sure why **I** was always so afraid. **I** do know there are things in **my** past that **I** can remember up to a point and **I** freak out and lose **it**. **I** can remember being real small **I** was in a dark room laying face down on the floor, **I** remember the floor being made of wood and **I** was scared. **I** laid real still hoping the person in the room with **me** would go away but as the other person came up behind **me I** turn to look back and **I** just lose **it**.

Essay B

I remember being happy most when **I** would run away from home, there was always fighting and arguing going on at **my** house, either **I** was being beat or one of **my** brothers and sisters or **my** mom would be getting jumped on. **My** dad would always be drunk when **he** came home and most of the time **we** went hungry because **he** would spend **his** check on booze or lose **it** at a card table. There was even times when **my** mother would tell **us** to leave the house just so **we** wouldn't get beat on. **We** all hated **him** to **his** or **her** own level. **It** was like **I**'d always get **it** the worst though, **I** fought back, **I**'d step in when one of **my** sisters or mother was being beaten which meant **I** got **it** just that much worse. **I** put a fork in **his** leg once when **I** was about six just trying to get **him** off **my** mother, then **my** mother would always turn around and threaten **me** with **him**.