Research Article

THE SECRET LIFE OF PRONOUNS: Flexibility in Writing Style and Physical Health

R. Sherlock Campbell and James W. Pennebaker
The University of Texas at Austin

Abstract—Numerous disclosure studies have demonstrated that individuals randomly assigned to write about emotional topics exhibit improved physical health compared with those who write about superficial topics. The writing samples from three previously published studies of 74 first-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 frequency of physician visits following the disclosure intervention. LSA revealed that flexibility in the use of common words—particularly personal pronouns—when writing about traumatic memories was related to positive health outcomes. The findings point to the importance of the role of discussing the self and social relationships in writing and, at the same time, to the remarkable potential of techniques such as LSA.

Psychological factors such as stress have long been linked to health markers and behaviors (e.g., Sapolsky, 1998). Multiple laboratories have demonstrated that people who are asked to write about traumatic experiences subsequently exhibit better physical health than people who are asked to write about superficial topics. In these studies, individuals are randomly assigned to write about either emotional or nonemotional topics for 15 to 20 min per day for 3 to 5 consecutive days. In the past 15 years, dozens of replications have demonstrated that emotional writing can influence frequency of physician visits, immune function, stress hormones, blood pressure, and a host of social, academic, and cognitive variables. These effects hold up across cultures, ages, and diverse samples (Pennebaker & Graybeal, 2001; Smyth, 1998).

Mediating factors that have been proposed to explain why emotional writing produces its effects include cognitive change (Pennebaker, Mayne, & Francis, 1997), habituation (Greenberg, Wortman, & Stone, 1996), and changes in working memory (Klein & Boals, 2001). In all cases, the proposed mechanisms have proven to be inconsistent or not directly tied to health changes per se. One consistent finding is that individuals who have written 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, does the writing of people who show health improvements differ from the writing of those who do not improve? If so, what is the best approach to analyze writing samples to determine healthy from unhealthy writing?

Various strategies have been used to attempt to answer these questions. Human raters have judged essays for degree of emotional expression, magnitude of trauma, and even participants’ writing styles. No consistent effects have emerged in predicting health or other outcome variables (Lepore & Smyth, 2002). Using a text-analysis program called Linguistic Inquiry and Word Count, or LIWC (Pennebaker, Francis, & Booth, 2001), researchers have found that individuals are more likely to benefit from writing if they use a relatively high rate of positive emotion words or an increasing number of causal or insightful words across their days of writing (Klein & Boals, 2001; Pennebaker et al., 1997). Although this word-counting approach is promising, the effects reported have been modest and often inconsistent.

These approaches to the assessment of essays have lacked a methodology that tapped the overall consistency of the essays in a reliable way. This consistency issue is relevant to psychotherapy because it is assumed that for individuals to improve, they must exhibit a change in the ways they are thinking (Beck, Rush, Shaw, & Emery, 1979). It follows that if the writing paradigm works on similar principles, people who benefit would also exhibit changes in their writing across essays. The problem is in defining and measuring essay similarity or consistency.

In recent years, researchers working on artificial intelligence have devised a number of mathematical ways to think about similarity that do not rely on human judges. One particularly promising strategy, Latent Semantic Analysis (LSA), computes a coefficient of similarity between any two text samples in a reliable and multidimensional way. Text samples can be adjacent sentences or separate essays written by the same person (see Foltz, Kintsch, & Landauer, 1998; Landauer & Dumais, 1997; and Landauer, Foltz, & Laham, 1998). LSA, then, can be a tool to measure the similarity of participants’ writings across essays. These markers of similarity can then be correlated with health outcome measures.

The similarity rating between any two essays will depend to a great extent on the basis for comparison. For example, two marbles may look quite similar if they are being compared with kangaroos but may look very different if they are being compared with other similar marbles. LSA develops a basis for comparison from a training corpus of text samples. The separate words within the various training samples are subjected to an analytic strategy called singular value decomposition, which is theoretically akin to factor analysis. The result is a series of word “factors” that represent the underlying linguistic structure of the training corpus. Collectively, this information is called a semantic space. The similarity of any new pairs of text samples can be assessed by applying the factor loadings from the semantic space to their respective words.

Traditionally, LSA has been used to compare the content of text samples (Graesser, Wiemer-Hastings, Wiemer-Hastings, Kreuz, & the Tutoring Research Group, 2000). The most common words—articles, auxiliary (helper) verbs, prepositions, and pronouns (collectively known as particles or function words)—are excluded. Whereas the more distinct and less frequently used content words tell what the author is saying, the more common particle words tell how the author is speaking. Indeed, this content-versus-style distinction is quite relevant in the present context. For example, in common speech, personal or professional writing, or stream-of-consciousness writing, words that...
reflect linguistic styles tend to be related to health behaviors, personality, and social processes. However, the more content-oriented linguistic domains are generally unrelated to psychological processes (Pennebaker & King, 1999; Stittman & Pennebaker, 2001).

In the present study, then, we sought to learn if the similarity of people’s writings—as measured by LSA—could predict health improvements. Text samples from three previously published written-disclosure studies were subjected to LSA, and the various markers of text similarity were then correlated with the objective health outcomes from the studies—number of health-center or infirmary visits. Because each of the studies randomly assigned participants to write about either emotional or nonemotional topics, we were also able to determine if essay similarity across the days of writing was differentially predictive of health changes as a function of experimental condition.

**METHOD**

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

**Participants**

*Study 1: First-year students (Pennebaker, Colder, & Sharp, 1990)*

For 3 days, 74 first-year undergraduates enrolled in introductory psychology (35 male, 39 female; mean age = 17.9 years, SD = 0.4) wrote about their deepest emotions about coming to college (n = 35) or, in the control condition, wrote nonemotional descriptions of their daily activities (n = 39). The sample consisted of participants who completed all questionnaires at baseline and at follow-up. Data on health-center visits for illness were collected from the university health center, and the mean number of visits per month was calculated for the 2 months prior to writing and the 4 months after writing.

*Study 2: 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 min each day. Data on health-center visits for illness were collected, and the mean number of visits per month was determined for the 2 months before and the 2 months after writing.

*Study 3: 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 3 days about either traumatic experiences (n = 33) or superficial topics (n = 26) for 15 min per day. Participants had a minimum of a sixth-grade education (mean education = 12.3 years, SD = 2.4). Number of infirmary visits in the 2 months before and after writing served as the dependent measure. Of the participants, 19 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).

**Analytic Procedure**

LSA represents a particular group of words as a set of numeric relationships to underlying concepts. Any two groups of words (essays, in this case) can be compared, because both 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. The analyses we report here are all based on a training corpus of writing samples chosen to represent informal, diary-style writing. Because the essays to be studied included traumatic 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. The training corpus was gathered from laboratories across the United States, England, and New Zealand, from a variety of investigators. 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.

To explore the relationships of different groups of words to health outcomes, we created several semantic spaces using the training corpus. For each participant, these semantic spaces were used to compare the first writing sample with the second and the second with the third; in addition, for the upper-division students, the third essay was compared with the fourth. The resulting LSA similarity ratings were averaged to provide an overall similarity rating for each participant. As described in the Results section, this process was repeated for each of the various semantic spaces created (e.g., Content, Style, and Particles). None of the essays analyzed in this investigation were used in the training corpus to create the semantic spaces.

In all the experiments, researchers obtained permission from participants to track their medical visits for illness at the student health center (Studies 1 and 2) or the prison infirmary (Study 3). A visit was considered to be an illness visit if the medical facility assigned a diagnosis that had a presumed illness cause. Visits for injuries, checkups, and noninfectious gynecological concerns were not counted. For each participant, the mean number of monthly physician visits prior to the writing intervention was subtracted from the mean number of monthly visits after the intervention. Then the change in frequency of doctor visits was correlated with the similarity ratings between essays. The illness difference measure was scored so that higher values mean more health visits after writing than before writing.

**RESULTS**

Because of the exploratory nature of the project, we discuss the results of the analyses in the context of separate LSA-related questions.

**Do Similarities in the Content of Writing Predict Health Change?**

The default approach for LSA focuses on content words that are relatively uncommon. The more frequently a word is used, the less important it is in the semantic space. 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.

Of the 3.4 million words in the training corpus, 19,013 unique words were part of the Content semantic space. Because of their multiple usage, these different words represented 60.7% of the total words used in the
Secret Life of Pronouns

The entire corpus. Overall, 30% of the words in the entire training corpus were recognized by this approach. Table 1 shows a writing sample from a participant in the emotional-writing condition of Study 1 with the words that were recognized as part of the Content semantic space highlighted. As this example makes apparent, the Content semantic space captures words that communicate the gist of the writer’s topic.

The initial LSA strategy involved distilling 276 independent factors, or dimensions, from the 7,501 essays in the training corpus. (The number of dimensions has varied in LSA studies depending on performance for the task at hand, but has usually been around 300.) 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. For each participant, we computed a Content similarity coefficient (the average of the coefficients for adjacent pairs of essays) to index the degree to which the content words of the essays were mathematically similar over the days of writing.

As shown in Table 2, the Content similarity coefficients were statistically unrelated to health changes among emotional-writing participants in all three studies. That is, participants were equally likely to benefit from writing whether they wrote about the same general topics or very different topics across the days of the study.

Do Changes in Writing Style Affect Health?

Although LSA was developed to study linguistic content, it can be modified to explore linguistic styles. 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 vocabulary accounted for 63.8% of the total words used.

The global Style space was reduced to 26 factors, considerably fewer dimensions than the Content space had. Indeed, there should be an order of magnitude fewer "styles" than there are "kinds" of content, partly because so few words are associated with style compared with content. The Style semantic space was then used to compare adjacent essays for participants in each of the three writing studies. The mean similarity ratings from essay to essay were consistently and positively related to the change in physician visits across all three studies (see Table 2). 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 more in the months after writing than in the months before writing. Conversely, those participants who changed their writing styles demonstrated improvements in health after writing.

What Accounts for Style?

The results from the Style semantic space were striking. The more that people changed in their writing styles, the more their health improved. Across all three studies, the effect size was far greater than the effects we had found with any other analytic strategy. However, although the results were impressive, it was not entirely clear what kind of words was truly responsible for the effect.

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 particles, or function words (cf. Miller, 1996). Particles consist of prepositions,

---

Table 1. Test sample comparing content words and style words (particles)

<table>
<thead>
<tr>
<th>Original text</th>
<th>Text with content words highlighted</th>
<th>Text with particles highlighted</th>
</tr>
</thead>
<tbody>
<tr>
<td>Coming to college conjured up these feelings.</td>
<td>Coming to college conjured up these feelings.</td>
<td>Coming to college conjured up these feelings.</td>
</tr>
<tr>
<td>Excitement, anxiety, happiness, worry, anticipation,</td>
<td>Excitement, anxiety, happiness, worry, anticipation,</td>
<td>Excitement, anxiety, happiness, worry, anticipation,</td>
</tr>
<tr>
<td>glee, nervousness, sadness, grief, energetic and</td>
<td>glee, nervousness, sadness, grief, energetic and many others.</td>
<td>glee, nervousness, sadness, grief, energetic and many others.</td>
</tr>
<tr>
<td>others. Most of all I felt very excited I could</td>
<td>I could not wait to get away from anything having to do with high school.</td>
<td>I could not wait to get away from anything having to do with high</td>
</tr>
<tr>
<td>not wait to get away from anything having to do</td>
<td>I hated high school.</td>
<td>school.</td>
</tr>
<tr>
<td>with high school.</td>
<td>Any possible way I could get away from my high school I would do it.</td>
<td>Any possible way I could get away from my high school I would do it.</td>
</tr>
<tr>
<td>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.</td>
<td>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.</td>
<td></td>
</tr>
</tbody>
</table>
Table 2. Correlation of essay similarity with change in frequency of doctor visits for experimental (emotional writing) participants

<table>
<thead>
<tr>
<th>Semantic space</th>
<th>Sample</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>First-year students (n = 35)</td>
</tr>
<tr>
<td>Content</td>
<td>-.05</td>
</tr>
<tr>
<td>Style</td>
<td>.34*</td>
</tr>
<tr>
<td>Particle</td>
<td>.38*</td>
</tr>
<tr>
<td>Preposition</td>
<td>.20</td>
</tr>
<tr>
<td>Conjunction, article</td>
<td>.18</td>
</tr>
<tr>
<td>Auxiliary verb</td>
<td>-.22</td>
</tr>
<tr>
<td>Pronoun</td>
<td>.35*</td>
</tr>
</tbody>
</table>

*p ≤ .05, **p ≤ .01.

conjunctions, articles, auxiliary verbs, and pronouns. Particles link phrases, clauses, and other grammatical structures together and can also serve as linguistic shortcuts to help identify relationships between the speaker and other individuals and objects.

Particles are among the most commonly used words in English. Although there are fewer than 200 common particles, they account for more than 55% of the words in our archive of essays, natural speech, and literature, which contains more than 13 million words (Pennebaker et al., 2001). Given the disproportionate influence of particles in the Style semantic space, we analyzed our writing samples using a Particle semantic space and then a series of separate spaces consisting of subsets of words from the Particle semantic space.

**Particle semantic space**

The Particle semantic space recognized 172 words comprising 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 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 in our analysis. Table 1 shows an example of which words in a writing sample were recognized by the Particle semantic space.

As shown in Table 2, the results for the Particle space were indistinguishable in both direction and magnitude from the results found earlier for the overall Style semantic space. That is, for all three studies, the more that individuals varied in their patterns of particle use from essay to essay, the more their health improved from before to after writing. Our next goal was to distinguish which category of particles accounted 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 verbs (44 words). Similarity coefficients were again computed for the written essays. As is apparent in Table 2, none of these dimensions were consistently related to health improvements in the studies.

**Pronouns**

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

The pattern of association between similarity coefficients in the Pronoun semantic space and change in doctor visits was almost identical to the pattern found for similarity coefficients in the larger Style and Particle semantic spaces. Among emotional-writing participants, the lower the similarity ratings of the essays, the more the health outcomes improved.

**Other Analyses**

Additional analyses were conducted as well, and we summarize the results here:

**Analyses of control groups**

In each of the writing studies, control groups wrote for 3 to 4 consecutive days about superficial, nonemotional topics. Separate LSA analyses for the control groups were conducted using each of the semantic spaces summarized in Table 2. In no case did the similarity scores correlate significantly with changes in the frequency of health-center visits.

**Sequential versus first-last measures of similarity**

Thus far, we have reported correlations between a health-improvement measure and a measure of overall similarity of essays. 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, comparing each participant's first essay with his or her last essay, under the assumption that first-to-last-day similarity (or change) in writing style might predict maximal health improvement. In fact, no similarity coefficients comparing the first and last days of writing were significantly correlated with health change. In other words, day-to-day changes in writing were more closely associated with health improvement than overall change from the first to the last day of writing.

**Meta-analyses of the three studies**

We conducted a meta-analysis 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, \( \chi^2(2) = 0.46, p = .80 \). Additionally, the significance levels of effects across studies were not different from one another, failing a test of heterogeneity of significance levels, \( \chi^2(2) = 0.16, p = .92 \). If the null hypothesis (no association between similarity of pronoun use and health benefits) were true, the probability of observing this pattern of results would be less than .0001. Overall, remarkably consistent results were
observed. The combined effect size of these three studies, weighted by degrees of freedom, was 1.15 (Cohen’s d).

**DISCUSSION**

Across the three writing studies, the way individuals wrote about emotional topics was related to a measure of health improvement. The LSA analyses demonstrated that changes in writing styles were consistently associated with better health, whereas similarity in the content of writing was unrelated to health outcome. Closer analyses of the factors that defined writing styles indicated that particles, and in particular pronouns, predicted the health changes. Individuals who altered their individual and social perspectives from day to day were the participants most likely to benefit from the disclosure exercise.

The results are important for two reasons. First, the analyses suggest the potential role of pronouns as markers of psychological and physical health, and, indirectly, of people’s thinking about their social worlds over the course of their writing. Second, the study underscores the promise of LSA and comparable techniques in providing insight into the ways people think. Although LSA has traditionally been used as a method for constructing networks of nouns and content verbs, the current study provides compelling evidence that the “junk” words that people use in writing and speech reveal a tremendous amount about how they are thinking.

Our conclusion that changes in pronoun use predict health is a rather bold and, at the same time, enigmatic statement. Our analysis compared constellations of pronoun usage between writing samples. The natural tendency of many researchers is to search for a common pattern of effects across participants; in this case, the common pattern is flexibility, not directional change. For example, it is tempting to speculate that it is ideally healthy for people to first write about themselves and then the next day to write about other people’s perspective. Statistically, however, this conclusion is not warranted. How the pronouns changed was not addressed by LSA.

Why might changes in pronoun use 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 (i.e., people, places, and things). Once these components are introduced, their interrelationships need to be described. Pronoun choice communicates this relational information, even though pronouns are generally seen only as placeholders in language, simply referring to components previously introduced (see Mulhauusler & Harre, 1990). Pronoun choice is based on perspective. For example, “us and them” betrays a very different perspective than “you and them” by highlighting whether the author is identified with a group, and which one.

Virtually all the traumatic experiences that were written about in our studies were ultimately social, and this is not coincidental. In the instances in which 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 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 saw their worlds.

In many ways, the current project raises more questions than it answers. Because the LSA approach is ultimately correlational, it is impossible to know if pronoun use is causally related to health changes or merely reflects an underlying emergent property in healthy people’s writing. Experimental studies are needed to see if flexibility in pronoun use can be manipulated and if health changes result. Similarly, we have not examined how pronoun flexibility may be linked to other linguistic changes within and across essays. For example, do shifts in pronoun use covary with judge-rated or LIWC-based markers of story construction? Indeed, preliminary analyses suggest that increases in the use of causal and insight words across the days of writing (which have been found to be related to health improvements—Pennebaker et al., 1997) are independent of changes in pronoun use.

Social and psychological processes are related to a wide range of health and illness measures (e.g., Cohen & Herbert, 1996). A recurring difficulty in health psychology is that there are no gold standards of illness or health. Physician visits may reflect malingering, hypochondriasis, or a predilection for seeking help. Interpretation of immune measures, physicians’ ratings, and a host of diagnostic tests is equally ambiguous. The outcome measure of number of physician visits must be viewed as merely a correlate of true illness, but with an appreciation of its financial and health-resource implications.

Finally, many researchers have tended to think of LSA as a system that represents the way cognitive or semantic information is organized in the mind (Landauer & Dumais, 1997). Underlying this approach is the assumption that the interesting linguistic action associated with thinking and knowledge resides in nouns, content verbs, and exotic modifiers. Our study shows that more attention must be given to particles in general and pronouns in particular.

**Acknowledgments**—Preparation of this article 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 and to Art Graesser, Josh Holahan, Bob Josephs, Matthias Mehl, and Lori Stone for comments on earlier drafts.

**REFERENCES**


(Received 11/12/01; Revision accepted 2/26/02)