

Linguistic Analysis of College Aged Smokers and Never Smokers

Sadie Alexander-Emery,¹ Lee M. Cohen,^{1,2} and Eric H. Prensky¹

Accepted March 19, 2004

College-aged adults represent the largest age group reporting current use of cigarettes. One factor that may account for these elevated smoking rates is tobacco advertisements targeting this population. Given the detrimental influence of smoking advertisements, it is essential to focus on counter measures. A unique and promising approach that may prove useful in preventing individuals from engaging in smoking behavior is to examine the manner in which individuals view smoking as well as how they view themselves. This study presents preliminary data on the use of Linguistic Inquiry and Word Count analysis to examine differences in language use among 48 college-aged smokers and never smokers when writing about the way they view themselves with regard to the dangers of smoking. Results revealed significant differences across a number of linguistic domains, indicating that the use of language may have implications for public health messages designed to prevent the onset of smoking, as well as those messages that target smoking cessation.

KEY WORDS: nicotine; college-aged smokers; linguistic analysis.

An estimated 430,700 Americans die each year from smoking-related illnesses, making smoking the leading cause of preventable death in the United States (American Lung Association [ALA], 2000). Smoking is a direct factor in 87% of all cases of lung cancer and has been implicated in cancers occurring throughout the body. Additionally, smoking has been linked to a number of other medical problems including emphysema, chronic bronchitis, coronary heart disease, stroke, infertility, peptic ulcer disease, and increased healing times for wounds (ALA, 2000). The compromised health status of individuals who smoke is not the only concern, because research indicates that second-hand smoke is responsible for approximately 3,000 deaths associated with lung cancer each year (ALA, 2000).

Alarming, college-aged adults (e.g., individuals aged 18–25) represent the largest age group reporting current use of cigarettes (Substance Abuse and Mental Health Services Administration [SAMHSA], 2002). Specifically,

40.8% of the 18–25 year olds surveyed reported smoking cigarettes in the past month compared to 13% among those individuals aged 12–17 and 25.2% among those aged 26 or older (SAMHSA, 2002). Although level of education appears to play a role in the smoking rates of these individuals (i.e., the higher the level of education, the less likely one is to smoke), those individuals who reported being full-time college students also reported higher smoking rates over the past month compared to the national average across all age groups (e.g., 32.6% vs. 26%, respectively; SAMHSA, 2002).

One factor that may account for the elevated smoking rate observed among this age group is tobacco advertisements targeting this population. Specifically, advertisements for cigarettes have been found to be effective in attracting the attention of adolescents (Gilpin, Pierce, & Rosbrook, 1997), and a positive correlation between exposure to such advertisements and adolescent smoking behavior is well established in the literature (National Cancer Institute, 2001; Altman, Levine, Coeytaux, Slade, & Jaffe, 1996; Botvin, Goldberg, Botvin, & Dusenbury, 1993).

The influence of advertisements on smoking behavior has become an area of concern among researchers

¹Texas Tech University, Lubbock, Texas.

²To whom correspondence should be addressed at Department of Psychology, Texas Tech University, Box 42051, Lubbock, Texas 79409-2051; e-mail: lee.cohen@ttu.edu.

interested in determining the best way to promote a message. The bulk of the research available in this field has focused on the outcome of positively or negatively framing messages and their impact on related health behaviors. A positively framed message focuses on the benefits of performing a health behavior, whereas a negatively framed message focuses on the losses associated with not performing a behavior (Smith & Petty, 1996). Previous research on this topic suggests that positively framed health messages are more likely to result in preventive behaviors, whereas negatively framed health messages are more likely to result in detection behaviors (Rothman & Salovey, 1997; Rothman, Salovey, Antone, Keough, & Martin, 1993; Banks et al., 1995). Specifically, positively framed health messages have been shown to increase requests for sunscreen among people at the beach, a behavior intended to prevent skin cancer (Detweiler, Bedell, Salovey, Pronin, & Rothman, 1999), whereas negatively framed health messages have been shown to increase mammography utilization, a behavior intended for detection of breast cancer (Banks et al., 1995). More recently, this area of research has been extended to include the influence of advertisements and warnings designed to promote smoking-cessation (Krugman, Fox, & Fischer, 1999). An initial study found that individuals who smoke cigarettes were more likely to respond favorably to smoking-cessation appeals if the message focused on the gains associated with quitting smoking (e.g., improved health) rather than the detriments associated with continued tobacco use (Schneider et al., 2001).

Given the detrimental influence of smoking advertisements on adolescent smoking behavior, it is essential to focus on counter measures. A unique and promising approach that may prove useful in preventing individuals from engaging in maladaptive health behaviors (e.g., smoking cigarettes) is to examine the manner in which they view specific health-compromising behaviors as well as how they view themselves. One method in which to obtain this information is through the utilization of linguistic analyses. Early on, such analyses were identified as a functional means to differentiate individuals experiencing positive health outcomes (e.g., less health care visits) after writing an essay from those who did not (Pennebaker, 1990). Specifically, Pennebaker (1990) observed that there was an association between positive health outcomes and increased use of positive emotion words (e.g., happy, pretty), causal thinking words (e.g., because, hence), insight words (e.g., think, consider), and negative emotion words (e.g., hate, enemy). Linguistic analyses have also been shown to differentiate between depressed and nondepressed individuals based on

differential use of negative emotion words, body-oriented words (e.g., heart, cough) and first person singular pronouns (e.g., I, my), with depressed individuals applying more of these types of words in their essays compared to those who were nondepressed (Rude, Gortner, & Pennebaker, 2002). Finally, in a brief review, Pennebaker (2002) suggested that the way an individual uses language is more predictive of individual psychological states than the specific topic.

This study examined specific linguistic patterns among college smokers and never smokers. It was hypothesized that the linguistic patterns of college smokers would differ from the linguistic patterns of college never smokers when asked to write about the dangers of smoking. Given the results from prior research examining message promotion, it was hypothesized that college-aged never smokers would use language consistent with prevention and detection behaviors, because they are writing about a behavior that has negative health consequences in which they do not engage. Thus, identifying potential differences in linguistic style could be helpful in determining how smokers and never smokers may differ in the views they hold of themselves. This information could be applied in the development of public health messages promoting prevention and cessation of cigarette smoking.

METHOD

Participants

Forty-eight undergraduates (25 males and 23 females) enrolled in Introductory Psychology classes participated in this study. The mean age of those who participated was 19.42 years (range 18–26), and the sample was primarily Caucasian (89.6%). Of the 48 participants, half were classified as smokers, reporting smoking 16 or more cigarettes per day over the past year, whereas the remaining half reported never smoking a cigarette.

Materials

Essay on Smoking

Participants were asked to write for 10 min about the effects of smoking on their health. They were told that the essays could be used in a statewide campaign to encourage youth not to smoke. In particular, participants were asked to consider the risk of contracting lung cancer, emphysema, chronic coughing, and coronary heart disease. In addition to the physical costs, participants were asked to discuss the increasing financial costs of smoking.

Tobacco Use and Mood Symptom Questionnaire (TUMSQ)

The TUMSQ is a scale developed for purposes of this study to screen potential participants to determine daily tobacco use, as well as any tobacco-related mood changes. The Chronbach's alpha coefficient for the scale is 0.97.

Procedure

Participants were recruited using a survey administered to 1,953 Introductory Psychology students during the 2001–02 academic year. The survey contained the TUMSQ to screen for smokers who reported 16 or more cigarettes per day ($n = 60$) and never smokers ($n = 811$). Of the 871 eligible respondents, 24 students were randomly selected from each group and were recruited using a telephone interview. From this number, four individuals (i.e., three smokers and one never smoker) did not show for their scheduled time, and one individual (i.e., a smoker) refused to participate. Four more smokers and one never smoker were randomly selected from the eligible respondents to fill each cell. All individuals who agreed to participate were asked to come to the laboratory at a specified date and time. Upon arrival, informed consent was obtained, and participants were provided an envelope containing the essay instructions. After writing the essay for 10 min, participants returned their essays to the experimenter in the envelope provided. Once the envelope was returned, participants were debriefed and remaining questions were answered. All participants received course credit for their participation.

Statistical Analyses

Essays were entered into a text file for analysis using Linguistic Inquiry and Word Count (LIWC). LIWC is a program that analyzes written text across 84 dimensions (Pennebaker, Francis, & Booth, 2001). These 84 dimensions include linguistic categories, psychological construct categories, relativity categories, and personal concerns categories. One-way analyses of variance calculations (ANOVAs) were performed on each of the dimensions of interest among smokers and never smokers.

RESULTS

In view of the exploratory and descriptive nature of this study, the dimensions of interest were numerous (see Tables I and II). Table I includes those variables that were found to be statistically different between smokers and never smokers, and Table II summarizes those variables of interest that were not found to be statistically different across groups. Consistent with study hypotheses, there was a significant difference between smokers and never smokers in the number of pronouns utilized, $F(1, 46) = 15.79$, $p < 0.001$, the use of insight-related words, $F(1, 46) = 8.858$, $p < 0.01$, and the number of inhibition words, $F(1, 46) = 8.511$, $p < 0.01$. Smokers used more pronouns and more inhibition words but fewer insight words than never smokers. There were, however, no significant differences between groups in the use of positive emotion words, $F(1, 46) = 0.161$, *ns*, causally related words, $F(1, 46) = 3.544$, *ns*, or negative emotion words, $F(1, 46) = 1.219$, *ns*.

Table I. Significant Mean Differences in Linguistic Categories Between College Smokers and Never Smokers

	College smokers	College never smokers	
Total pronouns	13.60	9.91	$F(1, 46) = 15.79$, $p < 0.001$
First person	10.26	4.20	$F(1, 46) = 36.55$, $p < 0.001$
First person singular	10.14	3.79	$F(1, 46) = 45.54$, $p < 0.001$
Second person	0.67	1.76	$F(1, 46) = 4.38$, $p < 0.05$
Third person	0.74	2.11	$F(1, 46) = 8.315$, $p < 0.01$
Death words	0.30	0.72	$F(1, 46) = 6.35$, $p = 0.015$
Time words	6.98	3.98	$F(1, 46) = 19.70$, $p < 0.001$
Past tense verbs	3.13	1.50	$F(1, 46) = 11.16$, $p = 0.002$
Social words	3.70	6.86	$F(1, 46) = 16.88$, $p < 0.001$
Communication words	0.64	0.25	$F(1, 46) = 5.82$, $p = 0.02$
Other references	1.74	4.41	$F(1, 46) = 17.69$, $p < 0.001$
Human words	0.43	1.08	$F(1, 46) = 10.67$, $p = 0.002$
Insight words	1.33	2.14	$F(1, 46) = 8.86$, $p < 0.01$
Inhibition words	0.96	0.41	$F(1, 46) = 8.51$, $p < 0.01$

Table II. List of Nonsignificant Mean Differences in Linguistic Categories Between College Smokers and Never Smokers

	College smokers	College never smokers	
Anxiety-related words	0.13	0.17	$F(1, 46) = 0.150, ns$
Sadness/depression words	0.26	0.33	$F(1, 46) = 0.178, ns$
Anger-related word	0.52	0.65	$F(1, 46) = 0.370, ns$
Achievement-related words	0.42	0.40	$F(1, 46) = 0.019, ns$
Religion-related words	0.22	0.24	$F(1, 46) = 0.001, ns$
Body-related words	4.33	5.30	$F(1, 46) = 2.998, ns$
Physical states words	4.60	5.45	$F(1, 46) = 2.277, ns$
Money-related words	2.37	2.64	$F(1, 46) = 0.462, ns$
Present tense verbs	13.22	13.15	$F(1, 46) = 0.009, ns$
Future tense verbs	1.54	1.55	$F(1, 46) = 0.0, ns$
Certainty-related words	1.29	1.57	$F(1, 46) = 0.652, ns$
Senses-related words	1.03	1.13	$F(1, 46) = 0.082, ns$
Friends-related words	0.07	0.15	$F(1, 46) = 0.863, ns$
Family-related words	0.45	0.44	$F(1, 46) = 0.004, ns$

A one-way multivariate analysis of variance (MANOVA) was conducted to explore further the observed difference in pronoun use across groups. Results revealed significant differences in total first person pronouns, first person singular pronouns, first person plural pronouns, total second person pronouns, and total third person pronouns, $\Lambda(5, 42) = 7.981, p < 0.000, \eta^2 = 0.487$. Subsequent univariate analyses revealed that smokers used more first person, singular pronouns, $F(1, 46) = 40.544, p < 0.001, \eta^2 = 0.468$, as well as more first person pronouns total, $F(1, 46) = 36.545, p < 0.001, \eta^2 = 0.443$, than did never smokers. Compared to never smokers, smokers less frequently used second person pronouns, $F(1, 46) = 4.377, p < 0.05, \eta^2 = 0.087$, and third person pronouns, $F(1, 46) = 8.315, p < 0.01, \eta^2 = 0.153$.

There were unanticipated differences observed between college smokers and never smokers. Specifically, significant differences were found in the use of words related to time, $F(1, 46) = 19.702, p < 0.001$, with college smokers writing significantly more past tense verbs compared to their never smoking counterparts, $F(1, 46) = 11.155, p = 0.002$. There were also significant differences in the use of socially related words, $F(1, 46) = 16.879, p < 0.001$, with smokers using significantly more communication-related words than never smokers, $F(1, 46) = 5.816, p = 0.02$. However, never smokers used more references to other people, $F(1, 46) = 17.690, p < 0.001$, and used more human-related words, $F(1, 46) = 10.671, p = 0.002$. Finally, never smokers used significantly more death- and dying-related words in their essays compared to their smoking peers, $F(1, 46) = 6.349, p = 0.015$.

DISCUSSION

The aim of this study was to perform a preliminary examination of the linguistic differences that characterize college smoker and never smoker writing styles when describing who they are as individuals and the health-compromising behavior of smoking cigarettes. This assignment was seen as a unique and potentially useful exercise that could be used to uncover differences between smokers and never smokers, thus leading to more effective public health messages as well as enhanced smoking cessation treatment programs.

Differences were observed in the manner in which smokers and never smokers used pronouns. Among the linguistic differences observed when participants were asked to write about the health effects of smoking, it was found that smokers used significantly more pronouns overall compared to never smokers. Further inspection of this finding revealed that first person pronouns, specifically, first person singular pronouns, were used more often among cigarette smokers. Although smokers produced more pronouns overall, never smokers were significantly more likely to use second person and third person pronouns. Thus, the initial hypothesis that significant differences would be observed in the way smokers and never smokers apply language when considering smoking and in relation to themselves was supported. Specifically, it appears that smokers wrote more about their personal experiences with smoking cigarettes compared to never smokers. This finding is not surprising, because the subject of the assigned essay was personally relevant to those individuals who were smokers.

Linguistic differences were also observed in the use of insight-related, and inhibition-related words. Specifically, never smokers were significantly more likely to write about the dangers associated with smoking cigarettes (i.e., insight words), whereas smokers were significantly more likely to write about their difficulty in refraining from this unhealthy, addictive behavior (i.e., inhibition words). Moreover, college smokers were significantly more likely to use references to time, particularly past tense verbs, compared to their counterparts who never smoked. Never smokers, however, were significantly more likely to write words that referenced other people, perhaps indicating that college never smokers are more likely to consider the effects of second-hand smoke compared to their smoking peers.

One of the more noteworthy findings was the observed difference in the use of words related to death and dying across groups, with never smokers being significantly more likely to use such words. This finding is consistent with previous research examining how public health messages are framed (Rothman et al., 1993; Rothman & Salovey, 1997; Schneider et al., 2001; Smith & Petty, 1996). With respect to cigarette smokers, Schneider et al. (2001) examined the effects of public health messages regarding tobacco use framed in terms of loss (i.e., fear-related messages) or gain (i.e., incentive-related message). Results from this study indicated that individuals who smoke cigarettes responded more positively to messages focusing on the gains associated with smoking cessation. The current study lends further support to this notion given that cigarette smokers in the current sample were less likely to use language focusing on death and dying. This finding suggests that loss-related messages are less likely to influence college-aged individuals who already smoke cigarettes. However, such messages may prove useful among college-aged individuals who have not yet initiated cigarette smoking.

Surprisingly, there were no differences found between smokers and never smokers in the use of words related to anxiety and depression. It is well documented that, at least for some individuals, tobacco products are used to regulate negative affect (Cohen, McCarthy, Brown, & Myers, 2002; Hall, Muñoz, Reus, & Sees, 1993). Perhaps the lack of difference observed between groups in terms of such word application can be attributed to the fact that smokers find in cigarettes a means to regulate affect, thereby negating the reflection of negative mood states in their writing samples.

This study was preliminary in nature and as such, results are considered in the light of several limitations. One weakness of the work is the limited demographic makeup of the sample, as well as its small size. Studies incorporat-

ing greater individual diversity, as well as increased sample sizes, may be helpful to increase the generalizability of findings. Second, it is necessary to examine the linguistic patterns of individuals in writing assignments involving different topics. Specifically, future studies could randomly assign both college-aged smokers and never smokers to writing conditions where they are asked to describe the health consequences of smoking or some other health-compromising behavior versus a neutral topic. Finally, because the present effort did not find differences in words related to negative affect, other experimental designs may be devised to target potential writing outcomes related to negative affect and its relationship to smoking.

The application of linguistic analysis as a means of understanding a variety of psychological states is relatively new, and present results hold promise for innovations in addictions research, specifically for assessment and treatment of nicotine dependence and withdrawal. Because psychological processes appear to play an integral part in the maintenance of nicotine use, linguistic analysis may serve an important function in identifying differences in the way that smokers and never smokers think of themselves with regard to smoking behaviors. Understanding such differences could be useful for development of public health messages designed to prevent smoking, as well as to improve upon existing smoking cessation treatment packages. There is no doubt that the influence of tobacco industry advertising has been correlated strongly with cigarette use among adolescent smokers (Altman et al., 1996; Botvin et al., 1993; National Cancer Institute, 2001). Therefore, information regarding the way college smokers and never smokers think of smoking, as well as describe this behavior, may be used to good advantage for formulating health-related messages targeting both prevention of smoking and smoking cessation among this particularly vulnerable population.

ACKNOWLEDGMENTS

Appreciation is extended to Stephanie Lewis Harter, PhD, and James W. Pennebaker, PhD, for their reviews of an earlier version of this manuscript. Appreciation is also extended to Gregory H. Mumma, PhD, for advice regarding statistical analyses.

REFERENCES

- Altman, D. G., Levine, D. W., Coeytaux, R., Slade, J., & Jaffe, R. (1996). Tobacco promotion and susceptibility to tobacco use among adolescents aged 12 through 17 years in a nationally representative sample. *American Journal of Public Health, 86*(11), 1590–1593.

- American Lung Association. (2000, September). *American Lung Association Fact Sheet Smoking*. Retrieved January 22, 2002, from http://www.lungusa.org/tobacco/smoking_factsheet99.html
- Banks, S. M., Salovey, P., Greener, S., Rothman, A. J., Moyer, A., Beauvais, J., et al. (1995). The effects of message framing on mammography utilization. *Health Psychology, 14*(2), 178–184.
- Botvin, G. J., Goldberg, C. J., Botvin, E. M., & Dusenbury, L. (1993). Smoking behavior of adolescents exposed to cigarette advertising. *Public Health Reports: Journal of the United States Public Health Service, 108*(2), 217–224.
- Cohen, L. M., McCarthy, D. M., Brown, S. A., & Myers, M. G. (2002). Negative affect combines with smoking outcome expectancies to predict smoking behavior over time. *Psychology of Addictive Behaviors, 16*(2), 91–97.
- Detweiler, J. B., Bedell, B. T., Salovey, P., Pronin, E., & Rothman, A. J. (1999). Message framing and sunscreen use: Gain-framed messages motivate beach-goers. *Health Psychology, 18*(2), 189–196.
- Gilpin, E. A., Pierce, J. P., & Rosbrook, B. (1997). Are adolescents receptive to current sales promotion practices of the tobacco industry? *Preventive Medicine, 26*, 14–21.
- Hall, S. M., Muñoz, R. F., Reus, V. I., & Sees, K. I. (1993). Nicotine, negative affect, and depression. *Journal of Consulting and Clinical Psychology, 61*, 761–767.
- Krugman, D. M., Fox, R. J., & Fischer, P. M. (1999). Do cigarette warnings warn? Understanding what it will take to develop more effective warnings. *Journal of Health Communication, 4*(2), 95–105.
- National Cancer Institute. (2001, November). *Changing adolescent smoking prevalence* (NIH Publication No. 02-5068). Smoking and Tobacco Control Monograph No. 14. Bethesda, MD: U.S. Department of Health and Human Services, National Institutes of Health, National Cancer Institute.
- Pennebaker, J. W. (1990). *Opening up: The healing power of expressing emotions*. New York: Guilford Press.
- Pennebaker, J. W. (2002). What our words can say about us: Toward a broader language psychology. *Psychological Science Agenda, 15*, 8–9.
- Pennebaker, J. W., Francis, M. E., & Booth, R. J. (2001). *Linguistic inquiry and word count: LIWC2001*. Mahwah, NJ: Erlbaum.
- Rothman, A. J., & Salovey, P. (1997). Shaping perceptions to motivate healthy behavior: The role of message framing. *Psychological Bulletin, 121*(1), 3–19.
- Rothman, A. J., Salovey, P., Antone, C., Keough, K., & Martin, C. D. (1993). The influence of message framing on intentions to perform health behaviors. *Journal of Experimental Social Psychology, 29*, 408–433.
- Rude, S. S., Gortner, E. V., & Pennebaker, J. W. (2002, June). *Language use of depressed and depression-vulnerable college students*. Paper presented at American Psychological Society, New Orleans.
- Schneider, T. R., Salovey, P., Pallonen, U., Mundorf, N., Smith, N. F., & Steward, W. T. (2001). Visual and auditory message framing effects on tobacco smoking. *Journal of Applied Social Psychology, 31*(4), 667–682.
- Smith, S. M., & Petty, R. E. (1996). Message framing and persuasion: A message processing analysis. *Personality and Social Psychology Bulletin, 22*(3), 257–268.
- Substance Abuse and Mental Health Services Administration. (2002). *National Survey on Drug Use & Health*. Retrieved November 7, 2003, from <http://www.samhsa.gov/oas/NHSDA/2k2NSDUH/Results/2k2results.htm#chap4>