Connoted Strength of Signal Words by Elderly and Non-Native English Speakers

N. Clayton Silver
Department of Psychology
University of Nevada, Las Vegas
Las Vegas, NV 89154

Dana S. Gammella
Department of Psychology
University of Central Florida
Orlando, FL 32816

Amy S. Barlow
Department of Psychology
University of Central Florida
Orlando, FL 32816

Michael S. Wogalter
Department of Psychology
North Carolina State University
Raleigh, NC 27695

ABSTRACT

A number of recent studies have examined the connoted strength of signal words used in sign and product label warnings. These words, such as DANGER, WARNING, and CAUTION, are intended to differentiate various levels of hazard (high to low, respectively). Until recently, most studies have only used college students to evaluate signal words. Other populations who are at least equal to or possess greater risk of injury have not been studied. The main purpose of the present research was to determine whether other populations of persons, namely the elderly and non-native English speakers, derive similar meanings (i.e., connoted levels of hazard) from the signal words as have been shown in previous work for college students, as well as, for a sample of grade-school children tested in Silver and Wogalter (1991). A sample of 98 elderly persons and 135 non-native English speakers rated 43 potential signal words on how careful they would be after seeing each term. The results showed that the rank ordering of the words was consistent across both groups and this order corresponded with the ratings from earlier-studied populations. Moreover, there was a significant negative linear relationship between the number of words the non-native English speakers left blank and ratings of understandability by college students in previous research. The forensic implications and practical relevance of these results for hazard communication to diverse populations are discussed.

INTRODUCTION

Current standards and guidelines on warning design recommend using signal words on signs or labels to connote various degrees of hazard. The standards suggest using DANGER, WARNING, and CAUTION to imply highest to lowest levels of hazard, respectively (ANSI, 1991; FMC Corporation, 1985). DANGER is intended to call attention to situations which involve immediate hazards that will result in severe personal injury or death. WARNING is intended to be used for hazards that could result in severe personal injury or death, and CAUTION is intended to signal hazards which could result in minor personal injury or damage (FMC Corporation, 1985).

However, research on whether people actually interpret different levels of hazard for these words has been equivocal. Although some studies have found no differences among the commonly used signal words (Leonard, Matthews, and Karnes, 1986; Ursic, 1984; Wogalter, Godfrey, Fontenelle, Desaulniers, Rothstein, and Laughery, 1987; Wogalter, Jarrard, and Simpson, in press), other work suggests that DANGER connotes a greater level of hazard than CAUTION or WARNING (Bresnahan and Bryk, 1975; Dunlap, Granda, and Kustas, 1986; Wogalter and Silver, 1990). Yet no study to date has found a significant difference in implied hazard between CAUTION and WARNING (e.g., Dunlap et al., 1986; Leonard, Hill, and Karnes, 1989; Leonard, Karnes, and Schneider, 1988; Silver and Wogalter, 1991; Wogalter et al., in press; Wogalter and Silver, 1990).

Nevertheless, most of these studies used college students as the population that evaluated the words. If signal words are to be truly useful in communicating different levels of hazard, then other populations at risk must understand the meanings intended. In an examination of one at-risk group, namely children, Silver and Wogalter (1991) had elementary and middle-school students rate 43 signal words that were taken from a study investigating a larger set of potential signal words (Wogalter and Silver, 1990). Specific analyses examined the connotations of eight terms (the three commonly-used terms plus five other terms that had been studied in previous research). The primary question addressed was whether the grade-school children and college students would give similar ratings on reported carefulness. The results showed that both populations rated the terms DEADLY, DANGER, WARNING, CAUTION, CAREFUL, NOTICE, ATTENTION, and NOTE in a relatively consistent order from greater to lesser hazard.

Besides young children, there are other at-risk populations who might fail to understand the implied hazard levels of signal words. Two of these populations are the elderly and non-native English speakers. One purpose was to determine whether the terms tested by Wogalter and Silver (1990) and Silver and Wogalter (1991) are rated similarly by elderly and non-native English speaking populations as college and grade-school children.

A second purpose was to examine several measures of understandability that might be used, in part, as criteria for selecting suitable signal words for a broad range of populations. These include: (a) low frequency of being left blank (not rated) by respondents (i.e., fewer missing values), (b) high ratings of understandability by college students, (c) high frequency of appearance in written language (as enumerated in analysis of English-language literature), and (d) shortness of word length. A third purpose was to develop a concise list of terms that most individuals are likely to know and understand for possible use as signal words.

METHOD

Participants

A sample of 233 individuals voluntarily participated in this study. Of these, 98 were residents of various retirement homes in Orlando, Florida (*Mean age* = 74.44, *SD* = 9.20). The remaining 135 participants were non-native English speakers enrolled in an "English as a Second Language" class either at the University of Central Florida in Orlando or at the Refugee Resettlement Program in Springfield, MA (*Mean age* = 23.75, *SD* = 5.45). These non-native English

speakers varied both in culture (represented by over 30 nationalities) and fluency with the English language.

Stimuli and Procedure

Forty-three words from the Silver and Wogalter (1991) study were used. These terms were originally selected from a pool of 84 terms (Wogalter and Silver, 1990) based on two criteria: (a) word length of less than 10 letters, and (b) mean understandability ratings above 4.0 (rated "understandable" and above on the scale).

All participants rated the 43 terms based on the question: "How *careful* would you be after seeing this term?" The question was accompanied by a 9-point Likert-type scale with the following numerical and verbal anchors: (0) not at all careful, (2) somewhat careful, (4) careful, (6) very careful, and (8) extremely careful. Although the scale had verbal labels associated with the even-numbered parts of the scale, participants were told that they could use any whole number between (and inclusive of) 0 and 8 in making their ratings. Participants received one of two random orders of the words and were told not to rate the word if they did not understand what it meant.

RESULTS

Several of the analyses described below used a within-subject analysis to compare the terms. Statistical algorithms generally do not allow the incorporation of data from subjects who have one or more data points missing. While there are a number of data substitution schemes described in the statistical literature, all have at least some unsatisfactory drawbacks. Thus, to simplify the data handling, participants with missing data were deleted. Specific analyses of the missing data are provided later in this section.

Analysis of the Three Common Signal Words

A 3 (signal word: DANGER, WARNING, CAUTION) X 2 (participant group: elderly vs. non-native English speakers) analysis of variance (ANOVA) was performed using carefulness ratings as the dependent variable. A significant main effect among words was shown, $F(2, 392) = 25.75$, $p < .0001$. Subsequent Newman-Keuls range tests showed that DANGER ($M = 6.99$) received significantly higher carefulness ratings than WARNING ($M = 6.46$) and CAUTION ($M = 5.94$). Moreover, WARNING had significantly higher carefulness ratings than CAUTION (all $ps < .01$). These words are shown in bold print in Table 1.

Analysis of the Eight Previously Researched Signal Words

An 8 (signal word: DEADLY, DANGER, WARNING, CAUTION, ATTENTION, CAREFUL, NOTICE, NOTE) X 2 (participant group: elderly vs. non-native English speakers) ANOVA was performed on the carefulness ratings. These means are underlined in Table 1. The ANOVA showed a significant main effect of word, $F(7, 1162) = 61.66$, $p < .0001$ with DEADLY, DANGER, WARNING, CAUTION, ATTENTION, CAREFUL, NOTICE, and NOTE rated from greatest to least on intended carefulness. Subsequent Newman-Keuls range tests showed significant differences among all pairwise comparisons ($ps < .05$) except between DEADLY and DANGER and between ATTENTION and CAREFUL.

TABLE 1

Means, Standard Deviations, and Percentages of Missing Carefulness Ratings of Signal Words for Elderly and Non-Native English Speakers.

	Elderly Participants			Non-Native English Speakers		
	mean	SD	% missing	mean	SD	% missing
NOTE	4.30	2.45	2.0	4.76	2.37	12.6
NEEDED	4.80	2.30	3.1	5.04	2.29	7.4
REMINDER	4.80	2.07	2.0	4.65	1.93	39.3
NOTICE	5.00	2.20	2.0	4.99	2.07	17.0
NECESSARY	5.06	2.20	2.0	5.44	2.15	13.3
REQUIRED	5.17	2.07	2.0	5.01	2.13	29.6
CAREFUL	5.24	2.19	1.0	5.97	2.02	15.6
PREVENT	5.33	2.11	0.0	5.35	2.04	28.1
ATTENTION	5.43	2.22	1.0	6.05	1.88	6.7
IMPORTANT	5.59	2.16	0.0	6.22	2.03	3.7
NO	5.82	2.31	5.1	5.84	2.55	4.4
ALERT	5.91	2.02	1.0	5.57	2.09	37.8
CAUTION	5.92	2.03	1.0	5.96	2.20	17.0
DONT	5.94	2.03	0.0	5.71	2.49	5.9
RISKY	5.96	1.90	1.0	5.63	2.25	31.1
INJURIOUS	5.99	2.01	1.0	5.91	1.94	37.0
BEWARE	6.08	2.03	1.0	5.54	1.92	40.0
ALARM	6.09	2.11	0.0	5.86	2.14	12.6
PROHIBIT	6.14	2.04	0.0	5.68	2.26	28.1
NEVER	6.28	2.06	1.0	5.58	2.89	6.7
VITAL	6.31	2.00	1.0	5.14	2.47	35.6
HALT	6.35	2.13	1.0	5.31	2.23	49.6
SERIOUS	6.43	1.70	1.0	5.91	2.02	8.1
CRUCIAL	6.44	1.87	0.0	5.18	2.32	4.0
WARNING	6.49	1.64	0.0	6.44	1.70	17.0
UNSAFE	6.55	1.83	1.0	5.60	2.09	23.0
FORBIDDEN	6.55	2.05	0.0	5.89	2.06	46.7
HOT	6.61	1.96	0.0	5.30	2.51	3.0
HARMFUL	6.68	1.72	2.0	5.69	2.15	37.0
HAZARD	6.72	1.82	0.0	5.51	2.08	44.4
URGENT	6.82	1.73	3.1	6.17	2.00	34.8
SEVERE	6.84	1.55	1.0	5.20	2.03	41.5
CRITICAL	6.89	1.62	2.0	4.96	2.43	22.2
STOP	6.96	1.70	0.0	6.45	2.17	3.0
DANGER	7.00	1.50	0.0	6.99	1.50	10.4
DANGEROUS	7.04	1.38	0.0	7.20	1.30	8.9
TOXIC	7.08	1.79	2.0	6.57	2.09	26.7
HAZARDOUS	7.17	1.51	0.0	5.03	2.26	52.6
FATAL	7.57	1.40	0.0	6.59	2.14	27.4
POISON	7.57	1.32	0.0	7.17	1.56	14.1
LETHAL	7.57	1.38	4.1	6.03	2.40	45.9
EXPLOSIVE	7.62	0.82	1.0	6.75	2.02	38.5
DEADLY	7.69	0.96	0.0	6.51	2.11	20.7

Note. Words are ordered according to the mean carefulness ratings of the elderly participants. Analyses of the bold and underlined words are described in the text.

Participant By Signal Word Interaction

Neither of the aforementioned ANOVAs showed a significant main effect of participant group. However, signal word and participant group interacted in the 2 X 8 ANOVA, $F(7, 1162) = 4.95$, $p < .0001$. Using Satterthwaite's (1946) correction for testing simple effects in mixed designs, the means showed that the non-native English speakers rated the term CAREFUL significantly higher than the elderly did, $F(1, 652) = 5.49$, $p < .02$. However, the elderly rated DEADLY significantly higher than the non-

517

native English speakers did, $F(1, 652) = 10.68$, $p < .002$. The other terms did not significantly differ between the two participant groups ($ps > .05$).

Consistency Between Participant Groups

Although there was a significant interaction between signal word and participant group with regard to mean carefulness ratings, the rank order of the signal words was fairly consistent, T (Kendall's Tau) $= .86$, $p < .0015$. In addition, the rank order of the terms by the elderly and the non-native English speakers was consistent with those of the grade-school and college students in earlier research (Silver and Wogalter, 1991), W (Kendall's coefficient of concordance) $= .93$, $p < .0001$. Moreover, the overall mean ratings for the elderly were highly correlated with mean ratings from the college students at the University of Richmond (Wogalter and Silver, 1990), Appalachian State University, and grade-school students (Silver and Wogalter, 1991). These correlations ranged from .82 to .96. The correlations for the non-native English speakers were slightly lower in relation to these groups, (rs ranged from .57 to .63). Lastly, there was a fairly strong correlation between the mean ratings of the elderly and non-native English speakers, $r = .60$, $p < .0001$.

Understandability Measures and Analysis of Missing Values

Participants were told that if they did not understand some of the terms that they should not rate them (i.e., leave them blank). Thus, the quantity of missing values could be used as a measure of the participants' understanding of the terms. Examination of the data showed that the elderly left blank less than 5% of the ratings, but as Table 1 shows, the non-native English speakers left blank ratings at a much higher rate. Words like HOT and STOP were left blank by less than 3% of the non-native English speakers, whereas, other words like HALT and HAZARDOUS were left blank by nearly 50% of these participants.

In relation to data collected in earlier research, the missing values of the non-native English speakers were negatively (and significantly) correlated with the mean understandability ratings of college students ($r = -.71$, $p < .01$). That is, as the number of missing ratings increase, the rated understandability decreases. Moreover, words most frequently left blank by the non-native English speakers were also terms used less frequently in the English language (Francis and Kucera, 1982), $r = -.46$, $p < .01$. Finally, as the number of letters comprising the words increased the number of missing ratings also increased, $r = .28$, $p < .05$.

Short list of terms. Using the Silver and Wogalter (1991) and the present study's data, a short list of terms was formed. Two criteria were used in the selection process. One was that the words had to be known by at least 95% of fourth and fifth grade students (from Silver and Wogalter, 1991). The other is that words had to be known by 80% or more of the non-native English respondents. The resultant set of terms is shown in Table 2 together with the mean carefulness ratings of the college students, fourth and fifth graders (Silver and Wogalter, 1991), the elderly, and the non-native English speakers.

DISCUSSION

In general, words connoting greater carefulness to the elderly were the same as those of the non-native English speakers. The pattern of results was also similar to those of

TABLE 2

Mean Carefulness Ratings of Signal Words Known by 95% or More by Fourth and Fifth Graders (from Silver and Wogalter, 1991) and by 80% or More of the Non-Native English Speakers. Also Shown are the Ratings of these Terms by the College Students and Elderly Participants.

	Silver and Wogalter (1991)		Current Study	
	College Students	4th & 5th Graders	Elderly	Non-Native English Speakers
NOTICE	4.01	5.39	5.00	3.64
CAREFUL	4.76	5.86	5.23	5.88
ALARM	5.01	6.16	6.09	4.87
IMPORTANT	5.06	5.95	5.59	5.64
CAUTION	5.22	6.64	5.91	4.75
DONT	5.24	6.12	5.93	4.54
NO	5.60	5.63	5.81	4.68
SERIOUS	5.73	6.90	6.43	5.52
NEVER	5.93	6.09	6.27	5.34
WARNING	6.13	6.52	6.49	5.58
HOT	6.21	6.00	6.61	4.40
STOP	6.43	6.11	6.95	6.55
DANGER	6.49	7.12	7.00	7.63
DANGEROUS	6.64	7.18	7.04	7.66
POISON	7.00	7.49	7.57	7.93

Wogalter and Silver (1990) and Silver and Wogalter (1991) who tested college and grade-school children. One exception was that the term ATTENTION had greater carefulness ratings than the term CAREFUL among the elderly and non-native English speakers. This pattern was reversed for all other populations. Moreover, among the non-native English speakers, DANGER had grater carefulness rating than DEADLY. This result was reversed for all other populations.

The higher carefulness rating for DANGER as compared to WARNING and CAUTION corroborates the findings of several earlier studies (Bresnahan and Bryk, 1975; Dunlap, Granda, and Kustas, 1986; Silver and Wogalter, 1991; Wogalter and Silver, 1990). However, the significantly higher carefulness ratings for WARNING as compared to CAUTION does not concur with previous research. Thus, while standards and guidelines (e.g., ANSI, 1991; FMC Corporation, 1985) have advocated and denoted a difference in hazard level between the two words, no empirical study, except for the present one, has found a significant difference. Explanations for this finding are not easily forthcoming, however, several possibilities can be noted. One is that the two groups of participants in the present study may perceive themselves as more vulnerable than either college or grade-school students, and as a consequence, they may make more specific, finer gradations among various levels of hazard. For example, most elderly persons have considerable exposure to warnings on various pharmaceutical containers and medical devices, and as result of this experience, they may have formed knowledge that delineates a difference in connoted hazard level between WARNING and CAUTION. However, this particular explanation does not account for why the non-native English speakers would make this differentiation. Possibly, in their limited exposure to the English language, the non-native English speakers received training on the intended meanings of the terms (perhaps by paying close attention to the gradations of English or to verbiage on products manufac-

tured by English-speaking countries while in their native lands). But, perhaps, a more rational approach to the difference between WARNING and CAUTION shown in this study is to withhold judgment until this pattern is replicated in other research. But even if these two words are found to be statistically different in subsequent research, some practical judgments should be made regarding the importance or even the adequacy of the difference. Given previous findings, a strong argument could be made that other terms might better delineate differences in hazard level.

The results also showed a strong relationship between the understandability ratings of college students (cf. Silver and Wogalter, 1991) and the percentage of missing ratings of non-native English speakers. This result suggests that the same underlying dimension is being captured by both measures. Of the 23 words that were known by 95% or more of the fourth and fifth grade students (cf. Silver and Wogalter, 1991), 15 of those were known by at least 80% of the non-native English speakers. The list shown in Table 2 is not the only list that could be constructed. Using different criteria, some aspects of the list would change. A warning designer interested in choosing a set of terms from this list should consider selecting those that are the most understandable to the target population(s) with significant (statistical as well as important) differences along the hazard dimension.

If individuals do not understand the level of hazard communicated by the signal word, then they might not exhibit the appropriate care and injury may result. If an injury does occur, then legal ramifications may ensue, with the manufacturer potentially liable for not communicating the hazard in a prudent fashion. For example, the California Court of Appeals, Fifth Appellate District ruled that if non-English speaking people are likely to use a product, then the manufacturer may need to provide foreign language warnings (Ramirez v. Plough Inc., 1992). Although it would be difficult to provide warnings that would encompass every language and nationality, perhaps English signal words which are more familiar to non-native English speakers, like the terms STOP and DON'T, would be more appropriate. In addition, another approach would be to examine the material usually used in teaching English to non-native speakers. That is, selection of signal words to be used for this population might be based on the terms already being taught to these individuals in English classes. However, examination of these teaching materials might reveal that very few safety words are currently being taught. A solution would be to develop a list of important safety terms and incorporate them into early English-language training curricula. Moreover, greater use of pictorials or icons might be useful in assisting hazard communication when the verbal information cannot be read.

In this study, the words were shown to participants in the form of a list. Thus, the terms were presented without the context that they would likely appear in real-world settings, e.g., on signs and on product labels. Therefore, it is not known whether the current results would hold in more externally valid settings. Additional research involving real-world contexts (as recently demonstrated by Wogalter et al., 1992, in press) is needed to verify whether the signal words communicate appropriate hazard levels to special populations such as grade-school children, the elderly, and non-native English speakers under different conditions.

Most research on signal words has focused exclusively on hazard level connotation. However, the literature on signal words also describes another purpose of these terms (e.g., Westinghouse, 1981): to capture people's attention to warnings. This additional purpose of signal words has not received much empirical study except for recent research by Young (1991). Young (1991) found that alcoholic-beverage warnings captured attention faster when a signal word was present compared to absent. Future studies using reaction time and eye movement measures (Laughery and Young, 1991) will serve to provide more information on the attention-getting aspects of signal words and provide a better base upon which to determine the proper components of warnings.

REFERENCES

ANSI. (1991). *American national standard for product safety signs and labels: Z535.4.* New York: National Electrical Manufacturers Assn.

Bresnahan, T.F., and Bryk, J. (1975, January). The hazard association values of accident-prevention signs. *Professional Safety,* pp. 17-25.

Dunlap, G.L., Granda, R.E., and Kustas, M.S. (1986). *Observer perceptions of implied hazard: Safety signal words and color words* (Research Report No. TR 00.3428) Poughkeepsie, NY: IBM.

FMC Corporation. (1985). *Product safety sign and label system.* Santa Clara, CA: Author.

Francis, W.N., and Kucera, M. (1982). *Frequency analysis of English usage.* Boston: Houghton Mifflin.

Laughery, K. R., and Young, S. L. (1991). An eye scan analysis of accessing product warning information. In *Proceedings of the Human Factors Society 35th Annual Meeting* (pp.585-589). Santa Monica, CA: Human Factors Society.

Leonard, S.D., Hill, G.W., and Karnes, E.W. (1989). How does the population interpret warning signals? In *Proceedings of the Human Factors Society, 33rd Annual Meeting* (pp. 550-554). Santa Monica, CA: Human Factors Society.

Leonard, S.D., Karnes, E.W., and Schneider, T. (1988). Scale values for warning symbols and words. In F. Aghazadeh (Ed.) *Trends in ergonomics/human factors V* (pp. 669-674).

Ramirez v. Plough Inc., Calif. Ct. App, No. F01950, (5th Dist., 1992).

Satterthwaite, F.E. (1946). An approximate distribution of estimates of variance components. *Biometrics Bulletin,* 2, 110-114.

Silver, N.C., and Wogalter, M.S. (1991). Strength and understanding of signal words by elementary and middle school students. In *Proceedings of the Human Factors Society, 35th Annual Meeting* (pp. 590-594). Santa Monica, CA: Human Factors Society.

Ursic, M. (1984). The impact of safety warnings on perception and memory. *Human Factors,* 26, 677-682.

Westinghouse (1981). *Westinghouse product safety label handbook.* Trafford, PA: Westinghouse Printing Division.

Wogalter, M.S., Godfrey, S.S., Fontenelle, G.A., Desaulniers, D.R., Rothstein, P.R., and Laughery, K.R. (1987). Effectiveness of warnings. *Human Factors,* 29, 599-612.

Wogalter, M. S., Jarrard, S. W., and Simpson, S. N. (1992). Effects of warning signal words on consumer-product hazard perceptions. *Proceedings of the Human Factors Society, 36th Annual Meeting* (pp. 935-939). Santa Monica, CA: Human Factors Society.

Wogalter, M. S., Jarrard, S. W., and Simpson, S. N. (in press). Influence of signal words on perceived hazard levels of consumer products. *Human Factors.*

Wogalter, M.S., and Silver, N.C. (1990). Arousal strength of signal words. *Forensic Reports,* 3, 407-420.

Young, S. L. (1991). Increasing the noticeability of warnings: Effects of pictorial, color, signal icon, and border. In *Proceedings of the Human Factors Society 35th Annual Meeting* (pp. 580-584). Santa Monica, CA: Human Factors Society.