

Text Analysis of Open-Ends

Daniel M. Ennis and William Russ

Background: Open-ended questions in surveys have always posed a challenge for the interpreter. When we design closed-ended questions in a survey, we assume that the respondent understands the meaning of the questions. However, to interpret open-ended questions, we assume the opposite – that a survey analyst understands what a respondent means. In some areas where surveys are used there is heavy reliance placed on open-ended questions. This occurs frequently in “consumer perception” or “consumer takeaway” surveys of ads in false advertising cases. Open-ended responses are given considerable weight in these cases where the occurrence of certain words and their counts relative to a control ad are often at the heart of the key evidence on which the decision-maker depends.

The decision-maker could be a judge in a bench trial, a jury, or an arbitrator. The decision-maker could also be a member of the staff of the National Advertising Division who decides cases as part of the advertising industry’s self-regulatory body. It is clear from decisions made by the NAD that the advertiser, the challenger and the NAD itself may have different interpretations of the meaning(s) intended in responses from open-ended questions and hence lead to different counts that may contribute to deciding a case. For example, Schick Manufacturing Inc. challenged The Gillette Company when the latter made reference to “moisture strips”, “intensive moisture” and “moisturize” in advertising their Venus Divine Shaving System for Women¹. The challenger noted a much larger difference between control and test ads in messages related to the “moisturizing” benefit than the advertiser. The NAD’s count was intermediate. Since the meaning of words depends heavily on the context in which they occur and the possible bias of the interpreter, the task of obtaining reliable counts from open-ended questions in surveys is difficult even for a human, and even more problematic for a machine.

Despite the difficulty for machines, computational tools can provide information gleaned from open-ended questions to help an analyst decide the most likely meaning of a word or phrase and the frequency of occurrence of different possible meanings. Although the tools may not replace a human compiler, they offer useful descriptive statistics on the occurrence of individual words and combinations and possibly include a component of textual context as is attempted in natural language processing². In this technical report, we explore some of these computer-aided tools and apply them to an actual survey involving open-ended questions about the meaning of a product label to a consumer audience.

Scenario³: You work for a company that distributes herbicides and fertilizers for home garden applications. You market two types of herbicides for lawns; a herbicide that kills post (weed) emergence on certain types of weeds on contact and a longer lasting herbicide that works pre-emergence and prevents weed initiation. You also market a fertilizer for lawns. Combining these three benefits with a six-month benefit for weed growth prevention provides the basis for a new product and a label claim in which all

four benefits are included. The proximity of the six-month benefit on the label to the other three benefits causes a competitor to challenge your label, arguing that the six-month benefit may be interpreted by consumers to apply to all three of the herbicide/fertilizer benefits. Neither party disputes that the long lasting benefit only applies to weed emergence prevention.

You contract an internet survey using the original label and a control label with the six-month benefit more clearly associated with the prevention benefit to check the validity of the challenger’s position. This survey is conducted among two matched representative samples of 300 home owners who purchase lawn care products or are involved in lawn care and influence the purchase of lawn care products. As is typical of consumer perception (takeaway) surveys, the survey begins with open-ended questions and progressively proceeds to more targeted questions regarding an opinion concerning what the participant thinks the product is intended to do, based only on the label information. Participants were exposed to each label for 30 seconds because shoppers spend limited time inspecting labels.

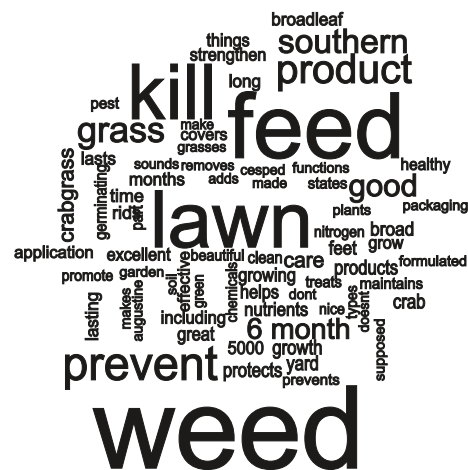


Figure 1a. Word cloud for the original label.



Figure 1b. Word cloud for the modified (control) label.

Word Clouds and Dendrograms⁵: Results from open-ended questions are usually categorized based on the interpretation of a coder. Different coders may produce different results depending on the complexity of the descriptions and the inherent bias of the coders. In the case of Venus Divine, where a shaving product’s “moisture strips” were at issue, it was critical to count the number of times a moisturizing benefit was expressed. Since this benefit could be expressed in different ways, it was not surprising that the advertiser and challenger had different counts. NAD staff conducted their own count and found a number different from both of them.

Although text analysis tools cannot fully resolve this issue without human intervention, at least some text analysis software can provide statistical information on the frequency of occurrence of words or word combinations much more quickly, and for much larger datasets than humans can evaluate, leaving more time for judgment issues to be made by the human interpreter. This can improve the quality and efficiency of open-ended response reports. Examples of these analyses include word clouds and dendrograms. Improvements due to the development of natural language processing in which word context is considered, will improve the automatic analysis of open-ended responses further².

Text Analysis of the Herbicide/Fertilizer Survey: Figures 1a and 1b shows word clouds for the two matched samples of homeowners. The size of the text in the word cloud displays the frequency with which a word or word combination is expressed. What you find remarkable about these word clouds is how seldom the six-month benefit is mentioned and how similar the word clouds are for the two conditions. Numerically, only 8.6% of respondents mentioned the six-month benefit after viewing the original label and only 10.3% mentioned this benefit with the modified label where the six-month benefit was clearly associated with prevention. These results imply that the six-month benefit did not resonate with consumers and that the particular placement of the benefit was not consequential.

Dendrograms of text show the degree of association among the words used and provide clusters of these associations. Figures 2a and 2b show dendrograms for the words used in the open-ends for both the original label and the modified label acting as a control. The six-month benefit falls into a different group for both surveys than two of the three main benefits (weeding and feeding) and for both conditions it is associated with prevention. Once again, the two conditions appear to be quite similar.

These results mean that the message of the six-month benefit was not strong as it was hardly noticed and certainly not associated with all three of the benefits, as the challenger maintained. If anything, the open-ended associations place the six-month benefit with prevention as it falls in a major cluster with that benefit. It is possible that even if the placement of the six-month benefit was not optimal, knowledgeable home owners who purchase lawn care products or influence their purchase, are aware that contact herbicides and most fertilizers do not last for six months.

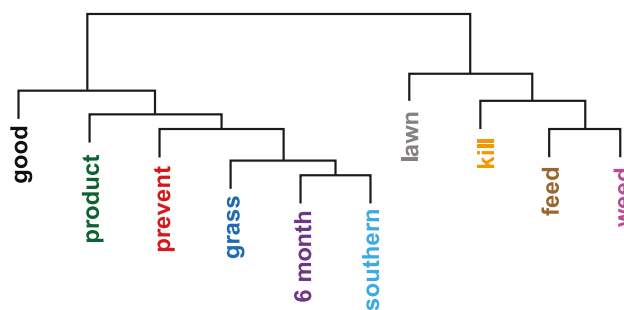


Figure 2a. Dendrogram for the original label.

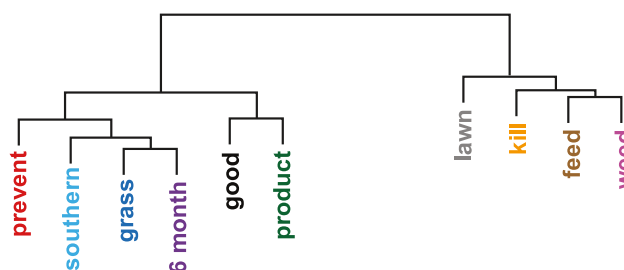


Figure 2b. Dendrogram for the control label.

Conclusion: In these analyses of the open-ended responses, you did not use pre-formed categories into which the open-ended responses were classified. This is how open-ended responses are usually analyzed. Instead, you analyzed word counts and associations and extracted meaning from these statistics using text software tools.

These results put you in somewhat of a dilemma. You agree that the label, as currently constructed, is not literally true as the six-month benefit does not apply to all three of the benefits. You could offer your survey results in defense of the label and probably prevail, but the label could mislead a naïve consumer who knows little about lawn fertilizers and herbicides, or one who scrutinizes labels more carefully and for longer than the average consumer. In the end, you decide to recommend a label change consistent with what the product does. It is not worth the risk of a possible negative NAD outcome, so you recommend conceding to the challenger’s argument, which ends the challenge, and changing the label to communicate what it should have said in the first place.

References and Notes

1. The Gillette Company (Venus Divine® Shaving System for Women) NAD Case #4305 (2005). Challenger: Schick Manufacturing Inc.
2. Pang, B. and Lee, L. (2008). Opinion mining and sentiment analysis. *Foundations and Trends® in Information Retrieval*, 2(1–2), 1-135.
3. This scenario, inspired by an NAD case⁴, is based on an actual survey funded and conducted independently of the advertiser and challenger by *The Institute for Perception and Research Now* for the purpose of method development.
4. Bayer CropScience US (Bayer Advanced 3-in-1 Weed and Feed for Southern Lawns) NAD Case #6033 (2016). Challenger: The Scotts Company, LLC.
5. The word clouds and dendrograms were created by Dr. John Ennis, formerly with *The Institute for Perception* and currently with *Aigora*.