

121. Values and Evaluation: Leximancer as a Tool for Analysing Values in Science Communication Transcripts

Melanie McKenzie

The University of Queensland and Econnect Communication

m.mckenzie@uq.edu.au

Abstract. Best practice science communication research, practice, and evaluation often dictates the use of focus groups and interviews, and yet techniques for analysing data from such collection methods are rarely discussed. This paper reports on the use of Leximancer, an automated text-mining software tool, to examine and compare transcripts from two focus groups. Participants included in this study were part of an upstream engagement process aimed at getting leading Australian farmers together to form a cooperative group to disseminate scientific information about managing climate variability and climate change. Findings from the study show that while cursory automated qualitative analysis of transcripts may not serve as a substitute for in-depth analysis, they may be complementary.

Keywords: Science communication, Values, Focus groups, Engagement, Transcript analysis, Leximancer

Introduction

Qualitative research methods such as focus groups and interviews are commonly employed mechanisms for assessing attitudes and perceptions of participants in science communication engagement activities. They are also frequently recommended as evaluation tools for such activities (1-2). However, while many of these guides describe how to carry out focus groups and interviews, there is little direction offered on what to do with the data once they are collected.

In this study, I conduct an exploratory qualitative investigation of two different focus group conversations using an automated textual analysis tool known as Leximancer. Leximancer is a highly customisable text mining software tool that uses word frequency and co-occurrence to identify families of terms (3). In essence, Leximancer uses lexical relationships and a thesaurus to develop “concepts” and then “themes” from documents (4). It has been used in both academia and industry to examine a variety of texts including interview transcripts, websites, surveys and emails, in fields such as History, Literature, Media Studies, Sociology, Politics, Psychology, Management, Business, Tourism and general Communication; however, its use and profile in science communication has been limited to date.

The observations made in this paper are not intended as a best practice guide, but offer some insight into how Leximancer may (or may not) be useful for analysing focus group transcripts.

This study is part of a larger project examining participant, organiser and facilitator values in emergent engagement. The study’s specific aim is to determine the usefulness of Leximancer as a tool for identifying values that emerge from value-oriented talk in upstream science communication. However, this specific aim does not distract from the goal of this paper in assessing its usefulness for science communication focus groups more generally, since I am simply considering Leximancer’s “themes”, which are made up of “concepts”, as potential values. While at the early stages of this project, where emergent values are not defined as yet, I also make some preliminary observations about the use of Leximancer to elicit “values” in transcribed conversations.

Method

Data were obtained by recording focus group discussions, which were then transcribed and subsequently analysed using Leximancer. During the course of the analysis, and after, I reflected on the subjective usefulness of Leximancer for analysing science communication focus group conversations, and particularly for eliciting values in such conversations.

Participants

In March 2010, twenty-four of Australia’s leading farmers from across the meat, cropping, dairy, sugar, bee keeping, wine and wool industries met in Canberra for a two-day induction into a group to be known as “Climate Champions”. The Climate Champions program was established in recognition of the role of peer interaction in how farmers gain new knowledge and adopt new practices. It aims to put farmers who are knowledgeable about managing and adapting to climate variability and climate change in touch with other farmers. From a science communication

research perspective, the Climate Champions program presents a rare opportunity to observe emergent upstream engagement, as the group was asked to define their own objectives and criteria for success before the trajectory of the program was decided.

The program is supported financially by five Rural Research and Development Corporations: Grains Research and Development Corporation, Meat and Livestock Australia, Dairy Australia, Rural Industries and Development Corporation, and Sugar Research and Development Corporation. It also receives communication support from an environmental and science communication consultancy called Econnect Communication, as part of their contracts with the Managing Climate Variability Program—a research funding body funded by the Australian Commonwealth Government, and the Grains Research and Development Corporation’s Climate Change Communication Campaign.

As part of the induction to the program, participants were separated into small groups (four to six people) for facilitated discussions. This paper examines two of those conversations in detail.

Focus groups

Facilitators were nominated by the organisers of the event and provided with a facilitator’s guide outlining questions for each session. The conversations examined in this paper come from a forty-five minute session designed to elicit participants’ responses to three value oriented questions:

- How will you know if the Climate Champions program has been successful?
- What will be different because of the Climate Champion program?
- What will be the value of Climate Champions?

Facilitators for the two groups used in this study consisted of ‘K’ from one of the sponsor organisations and myself, ‘M’.

Leximancer analysis

One of Leximancer’s most appealing features is that it enables the user to automatically identify significant themes (i.e., words with the greatest number of relationships to other words) and concepts (i.e., words that occur frequently) in text without any prior knowledge of its contents. With this in mind, I wanted to know what themes (i.e., potential values) would emerge from analysis of the conversations—both taken together as part of the upstream engagement as a whole, and considered separately. What were the potential values for both groups together? And what were the potential values for each of the groups?

In addition, I wanted to learn about Leximancer’s ability to account for the influence of context in conversation. Did it matter if turns in speech were represented sequentially? Would there be a difference in results if facilitator and participant speech were extracted from one transcript to produce two and then combined in Leximancer, compared to leaving the transcript intact (i.e., speech combined contextually)?

To explore these questions, transcripts from two of the Climate Champions focus groups were loaded into Leximancer v3.5, and compared in the ways described above. The standard options were selected, as well as “merge word variants” (e.g., communicate and communicating) and “apply dialog tags” (i.e., M: and K: to denote who is speaking) selected. Transcripts were parsed to produce separate transcripts for each of the facilitator’s speech and each group’s speech (i.e., six transcripts in total: M’s transcript intact, K’s transcript intact, M’s speech alone, K’s speech alone, M’s group’s speech alone and M’s group’s speech alone).

Concept maps for each of the analysed conditions were generated (shown in Section 3 below) under default viewing conditions (i.e., none of the sliders controlling detail have been adjusted). Labelled circles denote “themes”—larger circles are considered main themes and colours are assigned randomly. These “themes” are generated by related “concepts” identified in the text, which are denoted by dots within the circles. Lines connecting the dots denote pathways in the text between concepts. For the purposes of this study, I will focus on the comparison of themes, as I am investigating the possible use of themes as potential “values”.

Some may consider the level of analysis performed in this study to be rudimentary. Indeed, Leximancer is equipped with more features and options than what will be discussed here. However, by adhering to Leximancer’s more basic features, I am both minimising researcher influence on the outcomes of the data, and simulating a realistic level of knowledge that other science communicators might be likely to acquire after undertaking a similar degree of introductory training as I had, or less.

Findings

Analysis of two groups combined

Leximancer concept maps may be useful for gaining a sense of the content of text, without actually knowing what it contains. Figure 1 below shows the concept map generated for both group transcripts in this study, analysed together. Results show four main themes encompassing many concepts (i.e., relatively larger circles): “change”, “people”, “research” and “successful”; and seven other themes (i.e., relatively smaller circles): “farm”, “time”, “year”, “saying”, “terms”, “management” and “cause”.



Figure 1. Concept map for all group data (both transcripts) in context

In this study, I was interested in the themes that emerged as potential values in the conversations. In observing the interactions during meeting, facilitating one group and transcribing the recordings, I was familiar with some of the content and was able to make some decisions about acceptance criteria for themes as “values” in the conversation. For example, the themes “cause”, “saying” and “terms” from Figure 1 seemed like inappropriate values for the conversations, and so I decided to investigate this assumption further.

I noticed that two of these suspicious themes—“saying” and “cause”—are located in the periphery of the cluster and not touching or overlapping any of the other themes in Figure 1. These themes could therefore be taken to relate less to the other themes/concepts.

I also looked at the thematic summary (shown in Figure 2), which gives connectivity and relevance ratings for each theme. I noticed that “saying” and “cause” also ranked low in the list in terms of connectivity and relevance.



Figure 2. Thematic summary for all group data (both transcripts) in context

Further investigation of these terms in the transcript showed that “cause” was used mainly as an abbreviation for “because” (e.g., “cause they’re going to...”) and “saying” was used to describe what was happening in the conversation (e.g., “I was saying...” and “So you are saying...”). This further confirmed my assumption that their suitability as “values” was questionable.

While the theme “terms” ranked more highly in the comparison in Figure 2, the transcript revealed that it was used as part of the expression “in terms of” and was therefore also unlikely to be considered an appropriate “value”.

Leximancer facilitates the exclusion of particular terms in its analysis in at least two ways: words can be added to the default stop list of words that are removed from the analysis prior to generating concepts, or concepts can be removed prior to the generation of themes. However, manipulation of the data in this way introduces user subjectivity, detracting from any claims that might be made about neutrality in making use of Leximancer. Furthermore, the removal of concepts may not elicit more refined results. For example, removing old concepts may result in new concepts of questionable use to the user, as illustrated in Figure 3 where the concepts “cause”, “saying”, and “terms” (i.e., inappropriate values) were removed from the emergent concept list.



Figure 3. Concept map for all group data (both transcripts) in context, “cause”, “saying” and “terms” removed as concepts

While the four original main concepts remained (i.e., “change”, “research”, “people” and “successful”), the relative size of “successful” decreased. New themes that might be considered useful potential values were added: “Climate Champions” and “money”; however, new themes that are unlikely to be useful as values also emerged: “bloody” and “stuff”.

Comparison of two groups analysed separately

Leximancer concept maps may be useful for comparing the content of two texts, without actually knowing what they contain. For example, Figures 4 and 5 below show the concept maps generated for each of the group transcripts in this study, analysed separately.

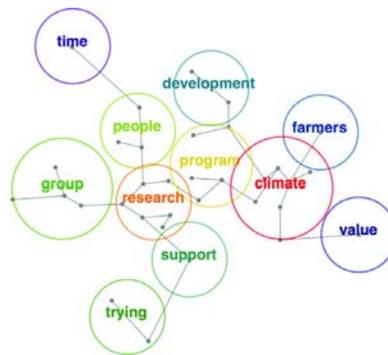


Figure 4. K group in context

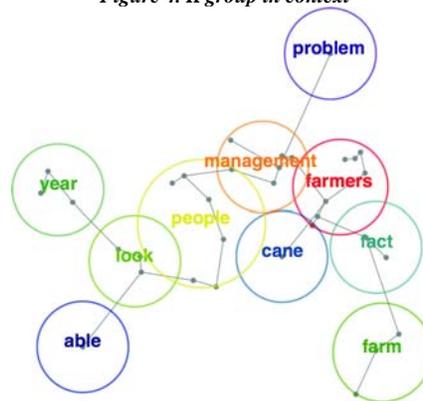


Figure 5. M group in context

At a glance, a comparison between the two groups showed few circles with the same labels, suggesting few similarities. However, a perfunctory explanation of each of the texts could be produced using some of its themes. For example, “something about developing a research program having to do with farmers and the climate” for K group and “something about cane farmers using facts for farm management” for M group. While these clusters of themes provide a sense of the topic of the conversations, without prior knowledge about what occurred in the transcripts, it would be

difficult to know how representative such stories might be.

Concept maps could be used for more direct comparisons of the data. For example, Table 1 below was produced to show a simple comparison between the themes of both groups.

Table 1. Comparing themes between groups

Comparison	Themes
Common themes	“people”, “farmers”
Similar themes	“time” (K)/ “year” (M), “trying” (K)/ “able” (M)
Uncommon themes	(K) - “climate”, “research”, “program”, “group”, “support”, “develop- ment”, “time”
	(M) - “management”, “look”, “farm”, “fact”, “cane”, “able”, “prob- lem”

In constructing the table, I could reason why some themes were common, similar and different. However, without knowing the context that each of these terms was used, it would be difficult to know whether “people” was a common theme for both groups for the same reasons or whether uncommon themes were really uncommon or describing similar things (e.g., “research” and “look”).

What was particularly interesting were the differences between themes in M group and K group taken separately (as in Figures 4 and 5), and then combined (Figure 1). Some of the main concepts stayed the same (e.g., “people” and “research”), while others were lost and/or were replaced by new themes (e.g., “climate” and “program” were lost, while “change” and “successful” were gained). Instead of having all the themes from each of the conversations as a contributing theme in the combined analysis, some kind of mediation had occurred and a more refined concept map resulted. In other words, individual speech contributions did not maintain their prominence when combined with other contributions.

The influence of context

In addition to mediation of contributions by members of different groups, Leximancer may be sensitive to context (e.g., ordering of turns in speech) in conversation. To investigate this sensitivity, I separated speech contributions from the facilitators of each group from the group contributions, and then analysed the four transcripts together to produce the concept map in Figure 6 below. Note that the text itself is the same text as was used to generate Figure 1, but in this case, the contributions were considered out of context (i.e., not in general “facilitator question - participant answer” format).

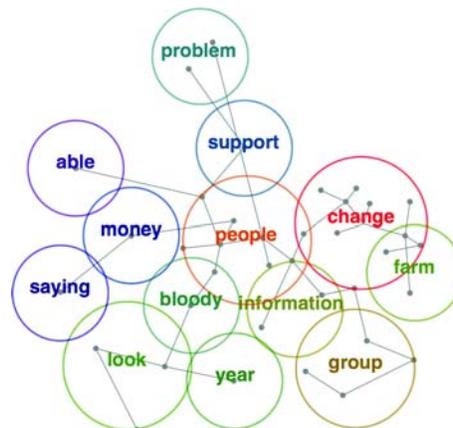


Figure 6. Concept map for all group data (both transcripts) out of context

A comparison between Figures 1 and 6 showed a slight refinement (i.e., fewer number) of themes when the transcripts were taken in context. Once again, the results of the combined data gave a sense of mediation within the conversation. For example, adjectives like “bloody” and directives like “look” may have been important in the conversation rhetorically, but did not contribute to the themes when taken in context.

Discussion

The combined results of these comparisons in the analysis lead to a few observations about the use of Leximancer

to analyse focus group conversations.

First, the extent to which Leximancer may be considered useful as a tool depends on what it is being used for. Leximancer aims to help the analyst avoid fixation on potentially atypical or anecdotal evidence by making them aware of the broader context and significance of concepts (5). Analysers invariably approach their data with expectations and prior assumptions about their meaning. Leximancer provides them with the opportunity to quickly test their assumptions. For example, many of the themes I would have initially expected to be important in the transcripts in this study (e.g., policy making and funding) were missing from the Leximancer analysis, while other themes (e.g., money and year) were unexpected. Leximancer offers a way to test assumptions about data like these, and be challenged by the results. In this way, analysts can widen their own value-lenses and be reflexive in their analysis.

Second, Leximancer's usefulness for analysing focus group conversations is limited by the user's understanding of its operation. The way that I have described its use here—to perform a cursory investigation through its default settings—leads to a very superficial understanding of the data. While it is possible to gather the essence of a conversation using basic features in Leximancer, doing so does not provide a particularly meaningful analysis on its own. It is difficult to make sense of the themes in Leximancer without an understanding of who is saying what, and in what context. For example, in this study, it is not sufficient to have the theme “research” (which is both a noun and a verb) without knowing that the term was generally used to describe the farmers' desire to participate in research and give feedback to researchers. Likewise, themes cannot be compared between groups without an understanding of how those themes emerged (e.g., “terms” as part of the expression “in terms of”).

While Leximancer can be configured to perform a more in-depth investigation than what has been described here, there are tradeoffs for users who engage Leximancer's more advanced features for analysing focus group conversations. Not unlike traditional, non-automated qualitative analyses, users should be aware that their own values influence what they consider to be important in the conversation they are examining. Leximancer offers a variety of ways to explore this influence—from choosing words to be included and excluded, to considering certain combinations of words and numbers of lines. For example, in further iterations of the analysis described in this study, I would exclude words such as “stuff” and “bloody” because they are not themes that I am interested in. However, in doing so, I would be detracting from one of Leximancer's main strengths: the ability to limit the influence of researcher bias to produce themes from data.

Thirdly, from the comparisons in this paper, I believe that Leximancer does have some sensitivity to the influence of context, albeit not in the way that users might like it to have. There was a difference in results when facilitator and group speech were separated and then analysed, but to what degree this analysis is able to account for the flow of conversation (e.g., facilitator question – participant answer) is not clear from this study.

Finally, in examining Leximancer's ability to extract values, I think its strengths lie in its use as a tool for reflexivity. In this analysis, consideration of themes as potential values was useful only insofar as I was able to use my own judgement about what might suitably be considered a value.

In summary, while Leximancer appears to be a useful data-mining tool, it may not serve as a substitute for traditional thematic analysis of conversations when used in the way I have described. However, Leximancer analyses may add value to traditional transcript analysis techniques, even if it might not replace them.

Acknowledgements

I would like to thank Econnect Communication, the Grains Research and Development Corporation and Managing Climate Variability program for their support in organising and permitting data collection during this event. I would also like to acknowledge Richard Fitzgerald and Jenni Metcalfe for their comments on earlier versions of this manuscript.

References

1. Evaluation: Practical Guidelines [monograph on the internet]. Swindon: Royal Councils UK; 2005 [cited 2010 Oct 22]. Available from: <http://www.rcuk.ac.uk/cmsweb/downloads/rcuk/publications/evaluationguide.pdf>
2. So did it work? Evaluating public understanding of science events [monograph on the internet]. London: COPUS; 1996 [cited 2010 Oct 28]. Available from: <http://www.copus.org.uk/downloads/sodiditwork.pdf>
3. Cretchley J, Rooney D, Gallois C. Mapping a 40-year history with Leximancer: Themes and concepts in the Journal of Cross-Cultural Psychology. *J Cross-Cultural Psychol.* 2010; 41(3): 318-328.
4. See Smith & Humphreys 2006 for a fuller discussion of how Leximancer transforms lexical co-occurrence into semantic patterns.
5. Smith AE, Humphreys MS. Evaluation of unsupervised semantic mapping of natural language with Leximancer

concept mapping. *Behav Res Meth.* 2006; 38(2): 262-279.