Using Semantic Technologies to Analyze the Semantic Orientation of Religious Sermons –
A Validation of the Early Work of McLaughlin

Denise A. D. Bedford, Goodyear Professor of Knowledge Management, Kent State University

Author Note

Correspondence concerning this article should be addressed to Denise A. D. Bedford,
Information Architecture and Knowledge Management, School of Library and Information
Sciences, Kent State University, Kent Ohio 44242-0001. E-mail: dbedfor3@kent.edu

Abstract

This research derives from a growing awareness in the knowledge management community of three factors: the value of language to knowledge management, the value of knowledge in all economic sectors and all aspects of human endeavour, and the “knowledge-richness” of belief systems and religion. Three research questions are addressed: What is the nature of language found in sermons? Is the use of semantic analysis technologies a feasible method for increasing our understanding of language patterns and characteristics? And, finally, Are there different approaches to the use of language in sermons across Christian religious communities? The research leverages semantic criteria defined by the early work of Raymond McLaughlin on the use of intentional and extensional language. McLaughlin’s research was necessarily limited in scale and scope and performed manually in 1940. In 2011, this research leverages semantic technologies to apply his well-formed semantic criteria to a larger scale (300 sermons) and broader scope (nine religious denominations). The research results suggest that McLaughlin’s criteria retain their value to language analysis today, that semantic technologies are a practical approach to applying these criteria to the use of language in religious communities, but that there are variations in the conclusion drawn by McLaughlin 70 years ago. The primary result suggests a high degree of balance of intentional and extensional language in modern day sermons.

Language, Communication and Knowledge Management

This research derives from a growing awareness in the knowledge management community of three factors: the value of language to knowledge management, the value of knowledge in all economic sectors and all aspects of human endeavour, and the “knowledge-richness” of belief systems and religion.

Communication and language are essential elements of all stages of a knowledge life cycle. By the knowledge life cycle we mean all stages – from the discovery of a knowledge gap, to idea generation, to knowledge creation, mobilization, propagation, capture, management and preservation. In the knowledge management literature and dialog, communication is generally recognized as a key factor. Language has typically received less attention from the knowledge management community. Language, in the form of semantics, though, is assuming a more prominent role in the knowledge management domain with the popularization of semantic analysis technologies and the increasing use of computational intelligence methods.

Traditionally, the field of knowledge management has focused on knowledge that is created in work-related settings, in high profile organizations, and sectors which have high economic value. As the knowledge economy expands into all aspects of society, there is increased awareness of the value of understanding how people transfer knowledge in everyday contexts, beyond the formal work or business environment. Some important knowledge management lessons are resulting from the study of knowledge practices in the skilled trades, manufacturing and farming. Knowledge management is relevant to all aspects of our lives and our economy in the 21st century. It is important for knowledge management professionals to extend their understanding beyond the Fortune 500 company or the high profile public sector organization.

Religious communities are “knowledge rich” in both foundational knowledge and beliefs and knowledge practices. A large portion of the population in any society participates in a religious community. Most religious or belief systems today have an explicit foundational knowledge and belief system, and a set of established practices for mobilizing and propagating (i.e., transferring or handing down) that knowledge across communities and over time. To date, though, the knowledge management literature has not considered religion as an area of study or research. What research exists on knowledge management in religion derives from the work in the theological community. This prevents both a gap and an opportunity. The goal of this

research is to take a small step forward in bridging that gap, to establish deeper research collaboration between the theological and knowledge management communities, and to provide an opportunity for knowledge management professionals to gain a deeper understanding of the role of language in knowledge

Research Context and Questions

There are many belief systems across the globe. While there are common elements to all belief systems, we believe that common knowledge practices are more likely to be found within a single system. The Christian religion provides us with a well defined knowledge foundation and set of beliefs. It also provides us with a common knowledge practice – use of weekly sermons – as a method of knowledge mobilization and propagation from religious leaders to religious community leaders. A sermon is oration, a form of discourse generally delivered as a monologue by clergy or other recognized religious experts (Ortberg 2007). Sermons are preached – using language which is chosen to effectively instruct the religious community, and to convey theological, religious or moral knowledge. The purpose of a sermon may address a wide range of issues, including general religious issues and topics, conversion efforts, narratives which tell stories and convey morals, and interpretations of current events. Some of the most effective sermons are drafted as stories, engaging the listener in the narrative.

In the early 21st century, sermons are available not only interactively in a place of worship such as a church, but also via broadcast communications, as transcripts and podcasts downloadable from the World Wide Web. The widespread available provides an opportunity to expand our understanding of the use of language in knowledge mobilization and propagation within religious communities.

Until recently any analysis of the language patterns has been a laborious and manual process. Human-based, manual analysis also carries a risk of high subjectivity and variability across individual researchers. The labor-intensive nature of the work may tend to limit the scale and scope of the analysis. Subjectivity, limited scale and scope are all factors which limit our ability to generalize. The availability of user-definable semantic analysis technologies addresses some of these risks. Tools available today allow us to encode analytical parameters and apply them consistently to large bodies of text. This provides data which can be objectively reviewed and interpreted by several researchers.

Understanding the use of language in sermons may increase our understanding of effective use of language to support knowledge management in other domains. If semantic technologies can be used to analyze language automatically, we can increase both the scale and scope of our analysis – to produce generalizable and reliable results. Given the fact that all belief systems have variations in practice – and this practice may be reflected in the use of language – we explore whether variations in practice are reflected in language.

This research strives to answer three essential questions:

- Question 1: What is the nature of language found in sermons?
- Question 2: Is the use of semantic analysis technologies a feasible method for increasing our understanding of language patterns and characteristics?
- Question 3: Are there different approaches to the use of language in sermons across Christian religious denominations?

Research Methodology

The research methodology for each of the three questions is described below.

Question 1: What is the Nature of Language Found in Sermons?

There are multitudes of ways to describe the nature of language. This research focuses on the use of *intentional* and *extensional language* in sermons as an indication of different structures and approaches to conveying knowledge and ideas (Korzybski 1938). Generally speaking, semanticists characterize *intentional language* as more abstract and conceptual, whereas *extensional language* is more concrete and enumerative of real examples and properties. In the context of religion, *intentional language* may tend to be more dogmatic but also have a closer resemblance to narrative and storytelling. Conversely, extensional language may tend to be more balanced and qualified and more characteristic of logical arguments and rationality. Two commonly held assumptions are important to this research question, specifically:

- From the field of theology – that sermons from fundamentalist preachers will tend to use more *intentional language*, and sermons from liberal preachers will tend to use more *extensional language*;
- From the field of knowledge management -- that narrative language style is more effective for engaging listeners and transferring knowledge than are argument- and rational-based language styles.

An example of early work in semantics came in 1940 when Raymond McLaughlin (McLaughlin 1940) translated general ideas of intentional and extensional into concrete, observable linguistic criteria. In an era many years before the widespread availability of computers, McLaughlin applied his characteristics manually to sermons from two preachers: Harry Emerson Fosdick, pastor of the Riverside Church of New York City, a liberal pastor of what is even today considered to be a liberal religious denomination, and the well-known fundamentalist mass evangelist preacher, Oral Roberts. McLaughlin found 15 sermons from each pastor and analyzed their language for frequency of occurrence of intentional and extensional characteristics. McLaughlin's criteria are described below.

Table 1.
McLaughlin's Semantic Criteria for Intentional and Extensional Language

Semantic Type	Semantic Criteria
Intentional Criteria	Allness terms as represented by such terms as all, every, entire, whole, none....
	Superlative terms such as best , worst, most, least, only, matchless,
	Two-valued terms such as either:or, if: when, if: not,.....
	"Is" of statements which represented identification of objects
	"Is" of as predication terms such as "roses are beautiful"
Extensional Criteria	Comparative terms such as higher, lower, more, less,
	Quantifying terms or precise numerical designations such as 60, sixty, second,
	Conditional terms such as if, but, except, perhaps, unless,
	Consciousness of projection terms such as seems, appears, in my opinion,
	Pseudo-quantifying terms or terms that loosely represent the idea of amount, size such as many, much, few, lots,

McLaughlin's characterizations were well formulated, but their generalizability and reliability was limited due to the limited number of cases to which they were applied (e.g., 30 sermons total), and also due to the scope of the language practices to which they were applied (e.g., two preacher representing only the extremes of the religious spectrum). To the best of our knowledge, McLaughlin's good work has not been carried forward, despite the fact that he

recorded important suggestions for further research. Seventy years later, we take a small step to extend his research to a broader context. McLaughlin's research into language patterns, if generalizable in a larger collection, could inform research in knowledge management.

Expected Result for Question 1: We expect the research to demonstrate the use of both intentional and extensional language in sermons, to indicate a greater propensity of intentional or extensional language depending on the nature of religious practices and beliefs (i.e., greater use depending on liberal, middle-ground, or conservative religious practices and beliefs).

Formulating a research methodology and criteria to test this expectation was not a trivial task. The research methodology and criteria are therefore grounded on the early work of semanticists and linguists. It is important to remember that semantics is a subject domain with a long history, one that predates the availability of semantic analysis technologies. And that the most effective use of semantic analysis technologies depends on leveraging human knowledge.

To extend McLaughlin's research, a test collection of 300 sermons was created. The three hundred sermons were representative of a broader spectrum of nine conservative/fundamentalist to liberal Christian religious denominations, including:

- Evangelical Baptist
- Evangelical Lutheran
- Seventh Day Adventist
- Methodist
- Catholic
- Episcopalian
- Presbyterian
- Congregational
- Unitarian

These communities were selected because they tended to provide a representative sample of conservative, liberal and middle-ground religious knowledge and practices within the larger Christian community. Sermons in each community represent multiple pastors and were delivered over at least a two year time period. All sermons were drawn from publicly available websites by the researcher. No other pre-selection criteria were applied. The test collection details are provided in Table 2 below.

Table 2.
Summary Description of Research Collection

Source	Number of Sermons
Evangelical Lutheran Sermons	30
Evangelical Baptist Sermons	60
Seventh Day Adventist Sermons	30
Catholic Sermons	30

Episcopalian Sermons	30
Methodist Sermons	30
Presbyterian Sermons	30
Congregational Church Sermons	30
Unitarian Sermons	30

We would like to note that the collection comprised a larger number of Evangelical Baptist Sermons because the publicly available, free sermons tended to be shorter than other sermons found on the web. The greater number was used to ensure that there was equivalent language content for the analysis for this particular religious denomination.

Question 2: Is the use of semantic analysis technologies a feasible method for increasing our understanding of language patterns and characteristics?

Text analysis and semantic analysis technologies are widely available on today's commercial market. There is a distinction between text analysis and semantic analysis technologies which is important to note when approaching a research task which involves such fine-grained language characterization. Text analytics software typically leverages statistical methods which focus on the patterns of language, and are designed to discover common co-occurrences and alignments of words. These methods typically have built-in algorithms for discovery and presentation. Semantic analysis methods on the other hand leverage a deep understanding of natural language including language-specific dictionaries, grammatical rules and part of speech tagging, and morphological rules. Semantic analysis methods also typically enable end-user or determination of important semantic elements and patterns within language. For this reason, semantic analysis technologies are better suited to this research than are text analytical tools.

When properly configured and engineered with embedded human knowledge, semantic analysis technologies have been observed to produce as, or more, objective results than those produced manually through human analysis. The rationale for this conclusion is the use of explicitly codified and traceable human knowledge and conceptual rules in constructing the profiles. Iterative review and fine-tuning of semantic analysis profiles leads to observable and traceable improvements in results.

Expected Result for Question 2: Given the capability to define semantic profiles at a fine-grained natural language processing, the expectation is that McLaughlin's criteria can be automated and applied to a broader scope of sermons and a larger collection.

The research task for Question 2 involved the construction of a single semantic profile representing both intentional and extensional language features. This task was comprised of three steps: (1) development of an understanding of McLaughlin's definition of each criteria; (2) interpreting that understanding as a set of concepts and conceptual rules; and (3) the construction of a semantic profile for SAS/Categorization Suite to use in processing the 300 sermons in the test collection. Each criteria was researched in the linguistic and semantic literature to develop an understanding. Next, that understanding was translated into concept lists and conceptual rules by consulting dictionaries, glossaries, thesauri and word lists design to advise writers found in the English grammar literature. Some criteria were easier to interpret and represent than were others. The profile represents a first attempt at a robust interpretation, but we believe there is room for improvement of this first attempt. Table 3 describes how the researcher interpreted each of McLaughlin's criteria.

Table 3
McLaughlin's Semantic Criteria Interpreted for Semantic Analysis

Semantic Type	Semantic Criteria	Semantic Rule Interpretation
Intentional Criteria	Allness terms as represented by such terms as all, every, entire, whole, none....	200 concepts comprised largely of single word adjectives and adverbs derived from common word lists, dictionaries, glossaries and thesauri
	Superlative terms such as best, worst, most, least, only, matchless,	2,061 concepts derived from single word adjectives and compound adjectival phrase with superlative connotations derived from common word lists, dictionaries, glossaries and thesauri
	Two-valued terms such as either:or, if: when, if: not,.....	32 concepts representing individual and combined concepts representing multivalve contexts
	"Is" of statements representing both identification and predication were combined	39 concepts which represented both identification and predication semantics as these are typically difficult to distinguish by humans and are more challenging to interpret at the machine level.
Extensional Criteria	Comparative terms such as higher, lower, more, less,	241 concepts representing single and multi-word concepts that represent comparators
	Quantifying terms or precise numerical designations such as 60, sixty, second,	97 concepts of both alphabetical and numerical characterizations of quantity
	Conditional terms such as if, but, except, perhaps, unless,	114 concepts representing single and multi-word conditional terms derived largely from thesauri, word lists, glossaries
	Consciousness of projection terms such as seems, appears, in my opinion,	53 concepts representing single and multi-word terms derived largely from thesauri and word lists
	Pseudo-quantifying terms or terms that loosely represent the idea of amount, size	73 concepts representing single and multi-word terms derived largely from thesauri and

	such as many, much, few, lots,	word lists
--	--------------------------------------	------------

Thirdly, each set of concepts and rules were encoded into a categorization profile in the SAS/Categorization Suite client. Figures 1 and 2 provide screen shots of the encoded concepts and rules for two intentional criteria, “Allness” and “Consciousness of Projection”. Figures 3 and 4 provide screen shots of the encoded concepts and rules for two extensional criteria, “Comparators” and “Conditional”.

Figure 1.
Screen Short of “Allness” Concepts and Rules
Intentional-Extensional Categorization Profile (SAS/Categorization Suite)

Figure 2.
Screen Short of “Consciousness of Projection” Concepts and Rules
Intentional-Extensional Categorization Profile (SAS/Categorization Suite)

Figure 3.
Screen Short of “Comparator” Concepts and Rules
Intentional-Extensional Categorization Profile (SAS/Categorization Suite)

Figure 4.
Screen Short of “Conditional” Concepts and Rules
Intentional-Extensional Categorizational Profile (SAS/Categorization Suite)

Description of Semantic Processing of Sermons

The SAS/Categorization Suite technologies analysis content through a three-step process. The *first step* involves generated a semantic index of the document. This involves part of speech (POS) tagging of each word in the document. This semantic index is then used as a baseline for further semantic analysis. In the case of intentional-extensional analysis, the *second step* involves applying the profile to each and every sermon in the test collection to determine whether the item meets a minimum threshold in terms of the “goodness of semantic fit” to the semantic criteria. This threshold is established by the researcher to represent the minimal level of match of the content to the concepts and rules. The “goodness of semantic fit” is represented by an indicator of PASS or FAIL – did the document meet the threshold that the research set for achieving for example a high degree of use of Comparative language? It PASSED if it met the threshold, which was set in this case at 50%. A 50% threshold meant that at least 50% of the semantic criteria specified in the profile had to be evident in the document. If it did not reach that threshold, it was rated as having FAILED. The 50% threshold of occurrences was defined consistently with what a person might expect as a strong match to the specifications for each

criterion. This is a fairly high threshold to achieve but the researcher wanted to set a high bar for the content. Finally, the *third step* is to calculate the individual threshold for each document that passed the 50% threshold.

Step 1: Semantic Index Generation

For the purpose of this research the semantic index was well formed. Given the volume of data and the size of the resulting semantic index, this is not generally presented either to the end user or in research results. The semantic index, though, was well formed through the use of sophisticated natural language processing methods and no-proprietary POS tag sets.

Step 2: “Goodness of Fit” PASS and FAIL

Figures 5 and 6 provide examples of the outputs of the semantic processing. Figure 5 illustrates the results for the use of “Comparator” terms in Congregational sermons. As we can see, each of the Congregational Sermons passed the threshold for these criteria.

Figure 5
Semantic Analysis Results for “Comparator” Terms in Congregational Sermons

Test File	Result	Relevancy
sermon-honeyandhoneydew.txt	PASS	0.96
sermon-honeyandhoneydew.txt	PASS	0.9391
sermon-honeyandhoneydew.txt	PASS	0.9379
sermon-honeyandhoneydew.txt	PASS	0.9126
sermon-honeyandhoneydew.txt	PASS	0.9098
sermon-honeyandhoneydew.txt	PASS	0.9056
sermon-honeyandhoneydew.txt	PASS	0.9028
sermon-honeyandhoneydew.txt	PASS	0.8252
sermon-honeyandhoneydew.txt	PASS	0.814
sermon-honeyandhoneydew.txt	PASS	0.8029
sermon-honeyandhoneydew.txt	PASS	0.8029
sermon-honeyandhoneydew.txt	PASS	0.8001
sermon-honeyandhoneydew.txt	PASS	0.7867
sermon-honeyandhoneydew.txt	PASS	0.7867
sermon-honeyandhoneydew.txt	PASS	0.7973
sermon-honeyandhoneydew.txt	PASS	0.7973
sermon-honeyandhoneydew.txt	PASS	0.7959
sermon-honeyandhoneydew.txt	PASS	0.7959
sermon-honeyandhoneydew.txt	PASS	0.7945
sermon-honeyandhoneydew.txt	PASS	0.7917
sermon-honeyandhoneydew.txt	PASS	0.7917
sermon-honeyandhoneydew.txt	PASS	0.7903
sermon-honeyandhoneydew.txt	PASS	0.7875
sermon-honeyandhoneydew.txt	PASS	0.7861
sermon-honeyandhoneydew.txt	PASS	0.6982
sermon-honeyandhoneydew.txt	PASS	0.6834
sermon-honeyandhoneydew.txt	PASS	0.682
sermon-honeyandhoneydew.txt	PASS	0.682
PastoralPaper082210.txt	PASS	0.6764

Figure 6 illustrates the results for the use of “Superlative” terms for Episcopalian sermons. As we can see, nine of the sermons failed to reach the threshold of 50%.

Figure 6
Semantic Analysis Results for “Superlative” Terms in Episcopalian Sermons

SEMANTIC ANALYSIS OF SERMONS

The screenshot shows the SAS Content Categorization Studio interface. On the left, a tree view displays the 'Sermon Language Profile' with various linguistic categories. The main window shows a list of test files and their results. The results are organized into three columns: 'Test File', 'Result', and 'Relevancy'.

Test File	Result	Relevancy
Pentecost Sunday 2010.txt	PASS	0.6726
Christmas Eve Sermon.txt	PASS	0.6704
Grace Episcopal Church: Where You There.txt	PASS	0.6701
Advent 3, 2003.txt	PASS	0.6687
Grace Episcopal Church: God's Love Made Manifest Another Way.txt	PASS	0.6686
Rev. Laila Epphary.txt	PASS	0.6683
Sermon Christmas 2006.txt	PASS	0.6683
Palm Sunday 2011.txt	PASS	0.6688
MLK Sermon 2011.txt	PASS	0.6688
Grace Episcopal Church: The Thin Places.txt	PASS	0.6683
Grace Episcopal Church: Community of God's People.txt	PASS	0.6683
Grace Episcopal Church: Who Do You Say That I AM.txt	PASS	0.6681
Grace Episcopal Church: Christ: The King Sunday.txt	PASS	0.6681
Grace Episcopal Church: Nurturing the Seed.txt	PASS	0.6680
Sermon: To Prepare in to Say Yes.txt	PASS	0.6678
Sermon: The God of Challenge and Compassion Comes in the Messia	PASS	0.6678
Sermon: The Cry of the Prophet.txt	PASS	0.6678
Sermon: Preparing for the Second Coming.txt	PASS	0.6678
Sermon: To Share the Peace of God.txt	PASS	0.6676
Indifference.txt	PASS	0.6676
Grace Episcopal Church: The End Gives Way to the Beginning.txt	PASS	0.6673
Sermon: Baptism Has Meaning Throughout the Ages.txt	FAIL	
Grace Episcopal Church: Travels in Faith.txt	FAIL	
Grace Episcopal Church: Time for a Change in our Routines.txt	FAIL	
Grace Episcopal Church: The Inevitable Human Struggle.txt	FAIL	
Grace Episcopal Church: Stewardship: God's Command to Manage.txt	FAIL	
Grace Episcopal Church: Self-Reflection.txt	FAIL	
Grace Episcopal Church: God's Most Precious Gift.txt	FAIL	
Grace Episcopal Church: Faith: A Way of Living.txt	FAIL	
Grace Episcopal Church: Emmanuel: God with Us.txt	FAIL	

The number of sermons from each community which met the 50% threshold is presented in Tables 4a and 4b. There was a high rate of success in meeting the threshold, with the exception of two indicators: Two-Valued Term language, and Quantifying Language. The results of this step provided sufficient data to use in moving to the next step – which involved analyzing the individual “Goodness of Fit” measures.

Table 4a
Number of Sermons Achieving Minimum 50% Threshold

	Baptist Sermons		Catholic Sermons		Congregational Sermons		Episcopalian Sermons		Lutheran Sermons	
	N=60	%	N=30	%	N=30	%	N=30	%	N=30s	%
Intentional Language Content										
Superlative Rating	27	45.00%	26	86.67%	26	86.67%	21	70.00%	17	56.67%
Identification Predication Language Ratings	56	93.33%	29	96.67%	30	100.00%	29	96.67%	28	93.33%
Allness Rating	56	93.33%	28	93.33%	30	100.00%	29	96.67%	29	96.67%
Two Valued Term Rating	0	0.00%	3	10.00%	1	3.33%	0	0.00%	2	6.67%
Extensional Language Ratings										
Consciousness of Projection Rating	43	71.67%	28	93.33%	29	96.67%	28	93.33%	29	96.67%
Comparative Language Ratings	44	73.33%	28	93.33%	30	100.00%	27	90.00%	28	93.33%
Quantifying Language Rating	7	11.67%	9	30.00%	9	30.00%	5	16.67%	7	23.33%
Conditional Language Ratings	59	98.33%	29	96.67%	30	100.00%	29	96.67%	29	96.67%
Pseudo Quantifying Terms	53	88.33%	28	93.33%	29	96.67%	29	96.67%	27	90.00%

Table 4b
Number of Sermons Achieving Minimum 50% Threshold

	Methodist Sermons		Seventh Day Adventist Sermons		Presbyterian Sermons		Unitarian Sermons	
	N=30	%	N=30	%	N=30	%	N=30	%
Intentional Language Content								
Superlative Rating	18	60.00%	24	80.00%	29	96.67%	27	90.00%
Identification Predication Language Ratings	29	96.67%	29	96.67%	30	100.00%	30	100.00%
Allness Rating	30	100.00%	28	93.33%	30	100.00%	30	100.00%
Two Valued Term Rating	1	3.33%	1	3.33%	1	3.33%	2	6.67%
Extensional Language Ratings								
Consciousness of Projection Rating	30	100.00%	29	96.67%	30	100.00%	30	100.00%
Comparative Language Ratings	26	86.67%	28	93.33%	30	100.00%	30	100.00%
Quantifying Language Rating	6	20.00%	17	56.67%	1	3.33%	11	36.67%
Conditional Language Ratings	30	100.00%	29	96.67%	30	100.00%	30	100.00%
Pseudo Quantifying Terms	30	100.00%	30	100.00%	30	100.00%	30	100.00%

Step 3: Analysis of Individual “Goodness of Fit” Measures

Each sermon was processed against each of the eight semantic criteria. The “Goodness of Fit” data for each sermon was collected and used as the research data set for analysis in Question 3. Figures 5 and 6 above illustrate the individual “Goodness of Fit” metric for each sermon for a given criteria. “Goodness of Fit” was only calculated for those sermons that achieved the 50% threshold. For example, in Figure 5, the “Goodness of Fit” indicator for Comparator language for the Congregational sermons entitled “Jesus and the Liberal Church” was 85.47. This meant there was a high rate of use of comparator concepts in this sermon. This could be compared to the 90.98 metric for comparator terms for the Lutheran sermon on Epiphany.

Because the amount of data produced for all 300 sermons is significant, we have presented a summary of Averages, High and Low “Goodness of Fit” for each criterion by community. These data are available in Tables 5a and 5b.

Discussion of the Effectiveness of the Semantic Analysis Approach

The use of semantic technologies was an exploratory research approach. In general, we would suggest that characterizing the use and types of language patterns is challenging when performed manually and when performed using technology. The advantage of using the technology approach is that it forces the human researcher to make their criteria explicit and available for review and critique by others. In other words, it potentially adds a degree of rigor and objectivity to the analysis. This allows others to review and improve upon the characterization.

It also produces predictable and auditable analytical results. Results which do not seem to make sense can be traced back to the explicit rules in the profile. They can be refined and adjusted until they align with human expectations. It also allows the criteria to be applied consistently to large collections. This is valuable for testing the scope and scale of our conclusions. In this case, being able to expand both scale and scope proved to be quite valuable in the final analysis.

Semantic profiling generally involves a parallel manual review as validation. Manual validation allows us to refine the profile and improve performance. Manual review was performed on two sermons from each denomination to ensure that the semantically generated results paralleled what a person would have concluded.

A caution is also included here regarding the nature of semantic technologies which may or may not be suited to this type of analysis. The software used for this research, the SAS/Content Categorization Suite, can be configured to generate individual results for each sermon processed. This supports review and validation. Semantic analysis technologies which simply produce aggregated results and only construct an index which is internal to the software application are not well suited to validation and verification of results.

Table 5a.
Detailed “Goodness of Fit” Rates for Sermons by Community
(Numbers Represent Averages, Highs and Lows)

	Baptist Sermons		Catholic Sermons		Congregational Sermons		Episcopalian Sermons		Lutheran Sermons	
	Ave.	High Low	Ave.	High Low	Ave.	High Low	Ave.	High Low	Ave.	High Low
Intentional Language Content										
Superlative Rating	100	100	64.47	67.09 66.86	66.97	67.54 66.72	66.87	67.26 66.73	66.98	67.94 66.67
Identification Predication Language Ratings	100	100	100	100	100	100	100	100	100	100
Allness Rating	76.09	85.51 66.67	79.07	86.35 67.17	82.37	88.86 67.5	79.88	85.35 67	79.78	90.2 66.83
Two Valued Term Rating	All Fail		61.71	61.71	61.11	--	All Fail	--	64.98	65.87 64.09
Extensional Language Ratings										
Consciousness of Projection Rating	100	100	100	100	100	100	100	100	100	100
Comparative Language Ratings	74.36	78.06 67.09	81.24	92.1 67.5	78.11	67.64 80.35	76.79	91.12 67.50	80.06	92.79 68.20
Quantifying Language Rating	67.37	71.93 68.42	73.68	75.79 71.58	73.60	80.35 70.18	72.1	75.79 69.62	75.43	85.61 69.82
Conditional Language Ratings	74.36	81.2 67.52	82.82	92.88 80.63	78.96	100 68.38	81.18	82.91 78.63	83.07	100 68.95
Pseudo Quantifying Terms	52.92	55.84 50	50.80	56.49 50.65	53.51	58.44 50.65	51.88	54.55 50	52.75	64.29 50.00

Table 5b
Detailed “Goodness of Fit” Rates for Sermons by Community
(Numbers Represent Averages, Highs and Lows)

Intentional Language Content	Methodist Sermons		Seventh Day Adventist Sermons		Presbyterian Sermons		Unitarian Sermons	
	Ave.	High Low	Ave.	High Low	Ave.	High Low	Ave.	High Low
Superlative Rating	66.81	66.94 66.73	64.04	67.04 66.86	64.58	62.02 66.78	63.71	67.31 66.83
Identification Predication Language Ratings	100	100 100	100	100 100	100	100 100	100.00	100.00 100.00
Allness Rating	77.67	85.18 66.83	85.51	86.01 67.67	78.37	85.51 67.67	83.15	86.68 68.17
Two Valued Term Rating	61.11	--	62.9	--	61.11	--	62.50	61.71 62.90
Extensional Language Ratings								
Consciousness of Projection Rating	100	100 100	100	100 100	100	100 100	100.00	100.00 100.00
Comparative Language Ratings	79.40	91.40 67.36	78.85	92.38 67.36	78.62	90.98 67.50	80.73	92.52 68.76
Quantifying Language Rating	72.74	79.33 72.28	73.70	76.84 70.88	74.74	--	66.97	78.95 71.23
Conditional Language Ratings	79.09	92.31 79.77	82.85	93.73 67.81	79.20	92.59 79.49	83.34	94.87 70.09
Pseudo Quantifying Terms	52.42	55.19 50.00	53.38	57.79 50	53.53	58.44 50.65	52.99	57.79 51.95

Research Results

Questions 1 and 2 largely address methodology questions. Question 3 is the primary focus of the research: Are there different approaches to the use of language in sermons across Christian religious denominations?

The results from the current research suggest that when tested against a broader set of sermons drawn from nine religious denominations, some variations were observed. In general, though, there was general consistency across both *intentional* and *external language criteria* and communities.

Use of Intentional Language

The research results suggest that there are differences in the use of *intentional language*, specifically in one criterion: Superlatives. The variation in the use of Superlatives did not align obviously with liberal or conservative perspectives. The outliers at the high end were found in Evangelical Baptist, Presbyterian and Unitarian. The outliers at the low end of the scale were found in Lutheran and Methodist sermons. No other variations in use of *intentional language* were observed.

This varies from McLaughlin's observations in 1940. In his research, McLaughlin observed that sermons from liberal preachers exhibited more *Intentional language* than did sermons from conservative preachers. Specifically, he suggested that liberal preachers generally used more "Allness" and "Superlatives", whereas conservative preachers used more "Is" of identification terms and "Quantified" language. We suggest that the difference derives from the small sample size and the fact that McLaughlin's comparison focused on the extreme perspectives and practices of the religious denomination.

Use of Extensional Language

Our research results suggest that there are some small differences in the use of *extensional language*, specifically in the use of "Consciousness of Projection", "Comparator," and "Quantifying" language. In the use of "Consciousness of Projection" language, there was only one outlier – Baptist sermons were at the very low end of the scale for use of "Consciousness of Projection" language. For "Comparator" language, the outlier was found in Baptist sermons which exhibited a low rate of use. For "Quantifying" language, low rate of use was observed for Baptist, Seventh Day Adventist and Presbyterian sermons.

There is some consistency between this and McLaughlin's research in the patterns of use of extensional language. In his research, McLaughlin observed that sermons from liberal preachers exhibited more *Extensional language* than did sermons from conservative preachers. While we did not find a higher rate of use of extensional language for liberal or center religious denominations, we did find a lower rate of use of extensional language in conservative religious sermons. Specifically, McLaughlin and our research observed a low rate of the use of Consciousness of Projection language. In addition, we observed – as did McLaughlin – a high rate of use of Conditional language across all sermons from all communities.

General Conclusions

In general, though, the research results would tend to suggest that sermons from all religious denominations take a balanced approach to the use of *intentional* and *extensional language*. This observation is derived not only from reviewing language practices in sermons across religious denominations, but within individual sermons. The existence of “balance” is the key finding of this research.

We believe that McLaughlin's characterization of the language criteria remain valuable in 2011. We agree with McLaughlin's suggestion from 70 years ago that more work needs to be done to understand and characterize the “Is” identification and predication criteria. We also believe, though, that other criteria may be added to the analysis – including indications of positive and negative sentiment.

Semantic analysis technologies are a feasible approach to expanding the use of McLaughlin's criteria to investigate the use of language in other domains. As an initial test outside of the domain of religion, we hope to expand our exploratory research into the use of intentional and extensional language to news stories in the popular media, on political speech, and on judicial decisions.

Extending the research further within the domain of religious knowledge, the next steps for this research team would be to attempt to characterize and categorize the nature of religious knowledge itself. Semantic analysis technologies, leveraging foundational knowledge from religious experts, may provide a practical tool for understanding and organizing religious knowledge.

Biographical Statement

Denise Bedford is currently the Goodyear Professor of Knowledge Management at the College of Communication and Information, Kent State University. She teaches courses in knowledge management, communities of practice, economics of information, semantic analysis, enterprise architecture, business intelligence, organizational network analysis, and information environments. Her educational background includes a B.A. in History, in Russian Language/Literature, and in German Language/Literature from the University of Michigan; an M.A. in Russian History also from University of Michigan; an M.S. in Librarianship from Western Michigan University, and a Ph.D. in Information Science from University of California, Berkeley.

SEMANTIC ANALYSIS OF SERMONS

References

- Ortberg, J. (2007). Birthing a Sermon: A step-by-step guide to bringing the text alive. *Leadership Journal* <http://www.christianitytoday.com/le/2007/summer/12.38.html>.
- Korzybski, A. (1938). Outline of General Semantics. *General Semantics Papers from the First American Congress for General Semantics*. Collected and arranged by Hansell Baugh. Arrow Editions, New York, 1938.
- McLaughlin, R. W. (1940). Intentional-Extensional Language as a Measure of Semantic Orientation. *Bulletin of the Evangelical Theological Society*, 143-151. McLaughlin was associated with the Conservative Baptist Seminary, Boulder Colorado, in 1940.