Pennebaker, J.W., & Chung, C.K. (in press).
Computerized text analysis of Al-Qaeda transcripts.
In K. Krippendorff & M. Bock (Eds.), *A content analysis reader*. Thousand Oaks, CA: Sage.

Computerized Text Analysis of Al-Qaeda Transcripts

James W. Pennebaker and Cindy K. Chung
The University of Texas at Austin
January 2007
* * * * * * * * * * * * * * *

Executive Summary

Two computerized approaches to analyzing 58 Al-Qaeda transcripts were employed. The first examined the linguistic style of Bin Ladin and al-Zawahiri; the second explored the content-related themes of their statements and interviews. Findings include:

- Compared to other extremist groups, the texts from Al-Qaeda are more emotional, angry, and concerned with other groups and governments
- From a stylistic perspective, Bin Ladin is evidencing a striking increase in the rate of positive emotion words as well as negative emotion words – especially anger words. He is also showing higher rates of exclusive words which often marks greater cognitive complexity in thinking.
- Zawahiri is evidencing a surprising shift in his use of 1st person singular pronouns. This dramatic increase suggests greater insecurity, feelings of threat, and, perhaps, a shift in his relationship with Bin Ladin. Overall, Zawahiri tends to be slightly more positive and significantly less negative and less cognitively complex than Bin Ladin in his statements.
- Using a new meaning extraction methodology, both Bin Ladin and Zawahiri have been devoting an increasing amount of their statements in talking about Bush, Iraq, and Afghanistan. At the same time, attention to Saudi Arabia and the Holy Lands as well as other Islamic hotspots has been decreasing dramatically.

Implications of using a new generation of text analysis procedures for political and other documents are discussed

* * * * * * * * * * * * * * *

Traditionally, the analysis of political documents has been done by experts in policy and international relations using a mix of discourse or content analyses (e.g. Hart, 1984; North, Hoslti, Zaninovich, & Zinnes, 1963; Simonton, 1988; Walker, 2000; Winter, 2005). The goal of these analyses has typically been to understand the meaning of the documents themselves or, perhaps, the motivation or intention of the authors. Although computerized approaches to text analysis have been around since the 1960s, they have not been widely used. With recent advancements in technology, computational linguistics, and the psychology of language, computerized text analyses are increasingly efficient and reliable.

Queries can be sent to the authors at Department of Psychology A8000, 1 University Station, University of Texas at Austin, Austin, TX 78712 (Pennebaker@mail.utexas.edu or CindyK.Chung@mail.utexas.edu). Thanks are extended to Conor Seyle for his analyses and expertise on extremism, and Andrew Lupo and Janna Miller for their assistance with the text preparation.

For the last 10 years, we have been exploring some novel ways to think about people's use of language. Rather than focusing on the meaning of a document per se, we have been studying how the language people use in writing a document (or even naturally speaking) reflects who the authors are. That is, the ways people use language reflect their basic social and psychological state. Others have discovered this as well. Weintraub (1981) hand counted thousands of documents and speech samples and provided compelling evidence that people's use of certain parts of speech were diagnostic of their emotional wellbeing. Martindale (1990), Mergenthaler (1996), Stone, Dunphy, Smith, and Ogilvie (1966), Gottschalk & Glaser (1969), and others have also contributed to the measurement of natural language use in psychology. Today, there are an increasing number of applications of word analyses in clinical (e.g. Gottschalk, 2000), criminal (e.g. Adams, 1996; 2004), cultural (e.g. Boroditsky, 2001; Tsai, Simenova, & Watanabe, 2004) and personality assessments (e.g. Oberlander & Gill, 2004; Pennebaker & King, 1999)

We have been approaching language use in two relatively independent ways. The first examines peoples' social and psychological states by analyzing their use of function and emotion words. Function words include pronouns, prepositions, articles, conjunctions, and auxiliary verbs. We find that the use of these word categories are diagnostic of emotional states (e.g., depression, self-esteem, suicide proneness), biological states (e.g., testosterone levels, heart disease proneness), personality (neuroticism, immediacy), cognitive styles (complexity of thinking, psychological distancing), and social relationships (honesty, dominance). Indeed, function words are powerful correlates of demographic variables such as age, sex, and social class (for a review, see Pennebaker, Mehl, & Niederhoffer, 2003).

Whereas the function word approach explores people's writing or speaking styles, the use of nouns, regular verbs, and many adjectives and adverbs reveals the content of their thinking. This style versus content distinction is quite important. Indeed, our second strategy to studying language focuses on groups of content words and how they cluster together in natural speech or writing. Much like recent developments in artificial intelligence (e.g., Latent Semantic Analysis, clustering methods, and other data mining strategies), we are developing approaches to automatically determining the underlying meaning structures in text.

The purpose of this project is to apply function word analysis and a meaning extraction method to the Al-Qaeda text samples. The function and emotion word analyses provide a sense of the social and psychological dimensions that we see in Bin Ladin and al-Zawahiri. The meaning extraction strategy allows a way of determining the themes that they are emphasizing in their public statements.

The text samples provided by the FBI were cleaned prior to analyses. Only the actual words spoken by Bin Ladin or Zawahiri were retained. Translator interpretations, extended religious or other quotations, or interviewer questions were removed. Spelling was corrected and normalized across texts. The final sample consisted of 58 texts of which 36 were authored by Bin Ladin, 17 by Zawahiri, 3 by both, and 2 unknown. As a comparison sample, we used a corpus of 17 files previously identified by Allison Smith (2004) as representing a sample of terrorist groups. This sample included statements from five different terrorist groups

representing a wide range of time periods, ideologies, and geographic locations. Groups represented in this sample include the Sicarii group of ancient Palestine (2 texts), the Front du Liberation du Québec (1 text), the Shining Path (4 texts), Hamas (1 text), and the Army of God (9 texts).

Social and Psychological Profiles: Analysis of Function and Emotion Words

For the last several years, we have been developing a computerized method by which to assess people's use of function and emotion words. The program, Linguistic Inquiry and Word Count (LIWC, Pennebaker, Francis, & Booth, 2001; Pennebaker, Booth, & Francis, 2006), is a relatively simple word counting program that calculates the percentage of words within several dozen categories that are used within any given text. The underlying logic of the program is that it searches for groups of words that have been predefined as matching the various categories of interest. For example, the program searches for and counts words that are related to the construct of anger. Groups of judges were used who agreed that words such as hate, kill, angry, outrage, etc were all anger-relevant words. The LIWC program looks at each text file separately and simply calculates the percentage of words in the entire text that match the words in the predefined anger dictionary as well as over 70 other language categories.

The basic LIWC program was used to analyze each of the Al-Qaeda files. Rather than focus on all 70+ LIWC dimensions, the current project examined the 16 or so language dimensions that have been found to be most correlated with social and psychological variables. For example, use of pronouns are closely related to depression, social status, individual and group identity, and insecurity. Certain classes of prepositions are associated with cognitive complexity. Both positive and negative emotion words are linked to the emotional state of the author. We first provide a comparison between the Alison Smith corpus with the two Al-Qaeda authors. A comparison between Bin Ladin and Zawahiri follows.

Al-Qaeda versus other extremist group differences. As can be seen in Table 1, simple LIWC analyses paint a striking difference between other extremist groups with the two Al-Qaeda authors. Compared to authors of the Smith corpus, Bin Ladin and Zawahiri focus more on other individuals (as seen in 3rd person plural pronouns) and are more emotional in their statements. They also pay less attention to past events than other groups.

The use of 3rd person plural pronouns (e.g., they, them) is highly significant. In our analyses of online extremist groups such as American Nazis, animal rights groups, etc, we find that 3rd person plural pronouns are the best single predictor of extremism as rated by independent judges (Seyle & Pennebaker, 2004). It suggests that the group is defining itself to a large degree by the existence of an oppositional group. A high usage of words like 'they' and 'them' indicates that the speakers are addressing people who they believe share the same world view and are attempting to bring the audience closer to their world view.

It is also of interest that the Al-Qaeda speakers are far more emotional in their use of both positive and negative emotion words than authors of the Smith corpus. In natural conversation, most people use almost twice as many positive emotion words than negative emotion words. It is interesting to note the high degree of negative emotions among the Al-Qaeda authors. These effects are due almost exclusively to the remarkably high rate of anger or hostility words (relative to anxiety or sadness words).

Bin Ladin versus Zawahiri. The analysis of the Zawahiri and Bin Ladin files suggest somewhat different speaking and, by extension, thinking styles. Overall, Zawahiri uses bigger words, tends to be more positive in his outlook, and is less time-bound, less immediate, and ultimately less analytical than Bin Ladin. Although there have been some revealing changes in Zawahiri's language of late (see below), most of his earlier communications suggest someone more emotionally detached from his topic and audience than Bin Ladin.

Table 1. Comparison of Public Statements by Bin Ladin, Zawahiri, and other Terrorist Groups

	Bin Ladin (1988 to 2006) N = 28	Zawahiri (2003 to 2006) N = 15	Controls N = 17	p (two-tailed)
Word Count	2511.5	1996.4	4767.5	
Big words (greater than 6 letters)	21.2a	23.6b	21.1a	.05
Pronouns	9.15ab	9.83b	8.16a	.09
I (e.g. I, me, my)	0.61	0.90	0.83	
We (e.g. we, our, us)	1.94	1.79	1.95	
You (e.g. you, your, yours)	1.73	1.69	0.87	
He/she (e.g. he, hers, they)	1.42	1.42	1.37	
They (e.g., they, them)	2.17a	2.29a	1.43b	.03
Prepositions	14.8	14.7	15.0	
Articles (e.g. a, an, the)	9.07	8.53	9.19	
Exclusive Words (but, exclude)	2.72	2.62	3.17	
Affect	5.13a	5.12a	3.91b	.01
Positive emotion (happy, joy, love)	2.57a	2.83a	2.03b	.01
Negative emotion (awful, cry, hate)	2.52a	2.28ab	1.87b	.03
Anger words (hate, kill)	1.49a	1.32a	0.89b	.01
Cognitive Mechanisms	4.43	4.56	4.86	
Time (clock, hour)	2.40b	1.89a	2.69b	.01
Past tense verbs	2.21a	1.63a	2.94b	.01
Social Processes	11.4a	10.7ab	9.29b	.04
Humans (e.g. child, people, selves)	0.95ab	0.52a	1.12b	.05
Family (mother, father)	0.46ab	0.52a	0.25b	.08
Content				
Death (e.g. dead, killing, murder)	0.55	0.47	0.64	
Achievement	0.94	0.89	0.81	
Money (e.g. buy, economy, wealth)	0.34	0.38	0.58	
Religion (e.g. faith, Jew, sacred)	2.41	1.84	1.89	

Note. Numbers are mean percentages of total words per text file. Statistical tests are between Bin Ladin, Zawahiri, and Controls. Documents whose source indicates “Both” (n=3) or “Unknown” (n=2) were excluded due to their small sample sizes.

We are hesitant to over-analyze the historical distinctions between the two Al-Qaeda authors at this time. The reason is that Zawahiri has been changing in his speaking style rather dramatically and progressively over the last year and a half. Most striking has been his changes in pronouns. As can be seen in Figures 1a and 1b, Zawahiri has more than tripled in his use of 1st person singular pronouns – primarily the use of the word "I". At the same time, his use of 1st person plural has remained flat, even dropping slightly in comparison with Bin Ladin. Normally, higher rates of "I" words corresponds with feelings of insecurity, threat, and defensiveness. Closer inspection of his "I" use in context tends to confirm this. Indeed, to the degree that Bin Ladin and Zawahiri are in contact, the relative differences in "I" use would suggest a significant change in their relationship.

The analysis of emotion words in Table 1 suggests that both Bin Ladin and Zawahiri share similar levels of both positive and negative affect. However, closer inspection of Figures 1c and 1d reveal some interesting trends over time. Since 2002, Bin Ladin's use of positive emotion words has risen significantly. Zawahiri has used similar impressively high rates of positive emotion since his first works in 2003. Interestingly, Bin Ladin's use of negative emotion words has been steadily increasing since 1988. Indeed, this effect is due almost exclusively to an increase in anger-related words. When this time factor is considered, it is clear that Zawahiri is significantly less negative and hostile than Bin Ladin.

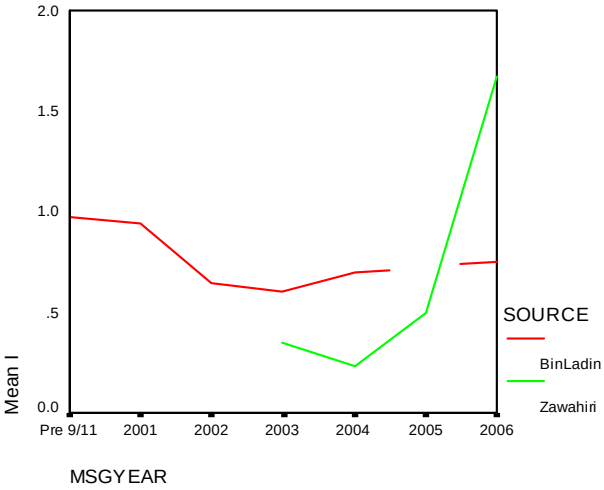
One other category that we pay close attention to is exclusive words. Exclusive words such as *except*, *but*, *exclude*, and *without* signal a person's attempting to make distinctions between what is in a category and what is not in it. We have found that exclusive word use is associated with greater cognitive complexity, telling the truth, and better grades in classes (see Chung and Pennebaker, in press, for review). Although Bin Ladin and Zawahiri do not differ in overall use of exclusive word use, Figure 1e indicates that Bin Ladin's exclusive words have been increasing significantly since 1988. On the other hand, Zawahiri continues to think in less complex ways.

Message and target analyses. Table 2 includes a basic comparison in word use as a function of the types of messages that Bin Ladin and Zawahiri have used. The table excludes the 3 letters and 3 epistles by the authors and only includes the 10 interviews and 42 statements. As is apparent, the language use is somewhat different among several dimensions. Not surprisingly, interviews result in more personal (i.e., 1st person singular) pronouns. Use of 2nd person (you, your) and third person plural (they, them) are more common in statements. The use of 2nd person is generally considered an aggressive form of speaking in one-on-one communications whereas it is quite common in general statements where the "you" is not specific and personal. The 3rd person plural effects are less clear and may reflect the nature of the interviewers' questions and perspective.

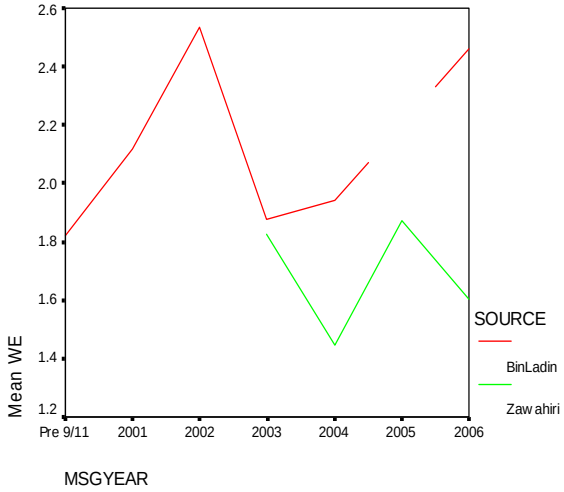
Target analyses comparing likely audiences of the messages were categorized into predominantly Muslim (N = 37), both Muslim and Western (N = 7), and predominantly Western (N = 8). As can be seen in Table 3, the differences in word use as a function of audience were not particularly striking. As might be expected, the authors were significantly more likely to use we (in reference to the speaker and their world), you, and less likely to use 3rd person plural when speaking to Western audiences. When addressing fellow Muslims, Bin Ladin and Zawahiri made reference to "them" in reference to Westerners.

Figure 1. Use of Pronouns and Emotion Words by Bin Ladin and Zawahiri Over Time

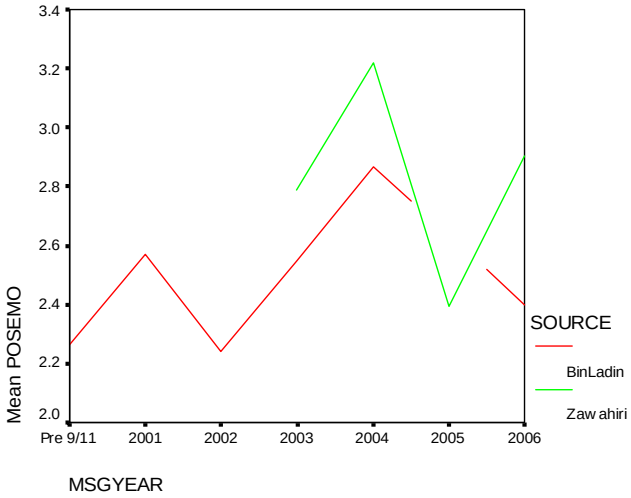
A. First person singular (I, me, my)



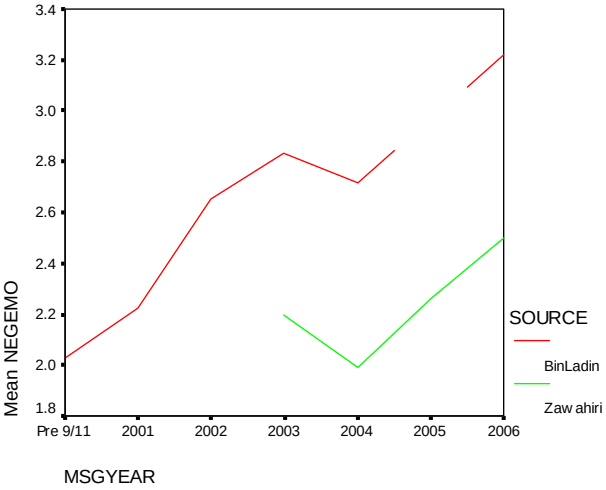
B. First person plural (we, us, our)



C. Positive emotion (happy, love)



D. Negative emotion (hate, sad)



E. Exclusive words (except, but, without)

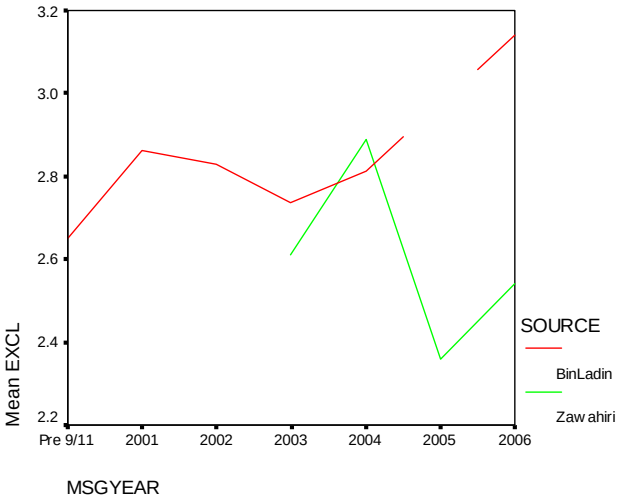


Table 2. LIWC Analyses By Message Type

	Interviews (1994 to 2005) N=10	Statements (1988 to 2006) N=42	p (two-tailed)
Word Count	3329.2	2495.4	
Pronouns	8.26	9.18	
I (e.g. I, me, my)	1.21	0.58	.01
We (e.g. we, our, us)	2.22	1.90	
You (e.g. you, your, yours)	0.32	1.54	.01
Other (e.g. he, hers, they)	2.75	3.65	.01
Prepositions	14.7	14.7	
Articles (e.g. a, an, the)	9.35	9.35	
Exclusive Words (e.g. but, exclude, without)	2.89	2.89	
Affect	4.55	4.55	
Positive emotion words (e.g. happy, joy, love)	2.22	2.22	
Negative emotion words (e.g. awful, cry, hate)	2.32	2.32	
Cognitive Mechanisms	5.19	5.19	.09
Content			
Death (e.g. dead, killing, murder)	0.48	0.48	
Humans (e.g. mother, people, selves)	0.92	0.92	
Money (e.g. buy, economy, wealth)	0.50	0.50	.06
Religion (e.g. faith, Jew, sacred)	1.79	1.79	
Factor 1: Compassion	.16	-.50	.01
Factor 2: Bush in Iraq and Afghanistan	.20	-.21	
Factor 3: Islam-Israeli Conflict	-.03	-.04	
Factor 4: Saudi, Egypt, US, Holy Land	-.08	.40	.06
Factor 9: Worldwide Islamic conflicts	-.08	.27	.06

Note. Numbers are mean percentages of total words per text file. Mean differences between Statements (1988 to 2006) and Interviews (1994 to 2005) are based on two-tailed independent samples t-tests, df=50. Documents whose message type is “Epistle” (n=3) or “Letter” (n=3) were excluded due to their small sample sizes.

Table 3. Target information

	Muslim (1988 to 2006) N = 37	Muslim/West (1996 to 2006) N = 7	West (2002 to 2006) N = 8	p (two-tailed)
Word Count	2742.5	2592.4	1343.6	
Big words (greater than 6 letters)	22.2	22.1	21.4	
Pronouns	8.52	9.49	11.13	.01
I (e.g. I, me, my)	0.61	0.68	1.17	.09
We (e.g. we, our, us)	1.86	1.70	2.77	.03
You (e.g. you, your, yours)	0.94	1.93	3.27	.01
He/she (e.g. he, hers, they)	1.25	1.33	1.16	
They (e.g., they, them)	2.32	2.28	1.42	.06
Prepositions	14.7	14.9	16.0	.05
Articles (e.g. a, an, the)	9.10	8.83	8.25	
Exclusive Words (but, exclude)	2.64	2.64	3.15	
Affect	4.83	5.24	5.46	
Positive emotion (happy, joy, love)	2.49	2.71	2.47	
Negative emotion (awful, cry, hate)	2.32	2.52	2.95	.06
Anger words (hate, kill)	1.37	1.47	1.81	
Cognitive Mechanisms	4.60	4.20	5.00	
Time (clock, hour)	2.21	2.03	2.46	
Past tense verbs	2.14	1.77	2.00	
Social Processes	10.30	11.06	12.60	.07
Humans (e.g. child, people, selves)	0.77	0.85	0.97	
Family (mother, father)	0.50	0.41	0.29	
Content				
Death (e.g. dead, killing, murder)	0.42	0.62	0.80	.06
Achievement	0.91	0.88	1.00	
Money (e.g. buy, economy, wealth)	0.36	0.40	0.47	
Religion (e.g. faith, Jew, sacred)	2.22	1.96	1.39	
Factor 1: Compassion	-.02	.15	-.17	
Factor 2: Bush in Iraq and Afghanistan	.08	.09	-.76	.01
Factor 3: Islam-Israeli Conflict	.10	-.19	-.34	
Factor 4: Saudi, Egypt, US, Holy Land	-.15	.72	.47	.01
Factor 9: Worldwide Islamic conflicts	-.03	-.22	-.39	

Note. Numbers are mean percentages of total words per text file. Mean differences between Muslim, Muslim/West, and West are based on one-way ANOVAs, F(2, 49). Documents whose target audience is "General" (n = 2) or "Extremists" (n = 4) were excluded due to their small sample sizes.

Meaning Extraction Method: Pattern Analysis of High Frequency Content Words

Much like recent breakthroughs in computational linguistics (e.g., Landauer & Dumais, 1997), we have developed a method to determine the major themes that occur in a text sample based on the co-occurrence of high frequency content words. Content words (or open-class words), such as adjectives, adverbs, nouns, and regular verbs, are more telling of conversational or writing topic than of linguistic style. Our meaning extraction method begins by using a word counting program that ranks all the content words in a corpus by frequency of use. The most frequently occurring content words across all texts are compiled into a LIWC dictionary, and their patterns of co-occurrence are assessed using a factor analytic approach. Each resulting factor is made up a coherent group of words that co-occur in the texts. The factors can be used for descriptive purposes or for further analyses both within and across texts. The primary advantage of the meaning extraction method is that it provides intuitively comprehensible themes based on a purely inductive approach.

Instead of imposing a predetermined coding structure on the Al-Qaeda files, the meaning extraction method was used to uncover the major themes in the 57 texts from 1994 to 2006. First, the most frequently occurring content words appearing in at least 25% of text files were made into a LIWC dictionary. Each text file was broken down into text segments of 308 words for analyses. This text segmentation was chosen since the smallest text file had a total of 308 words after the text files had been cleaned. After dividing by 308 words, remaining text segments in text files were included for analyses if they included at least 40 words. This resulted in 519 text segments. LIWC assessed each of these 519 text segments for the occurrence (coded as 1) or absence (coded as 0) of the top 257 occurring content words (including any alternate forms that could be made from their root word). The final data summary, then, can be thought of as a 257 (content word) by 519 (text segment) matrix, with each entry referring to the presence or absence of each content word within each text segment.

A factor analysis on this matrix (principal components analysis with varimax rotation) produced 18 factors with Eigenvalues of at least 2.50, accounting for 22% of the variance (see Table 4). Regression-based factor scores were computed for each of the text segments. The means of the regression-based factor scores for each text file are the basis of further analyses over time and by author.

Word factors over time and by author. The meaning extraction approach to content analysis opens new ways of thinking about language. Each factor exists along a continuum where a high positive number reflects the use of words within the factor. Because most factors include words that are only positively loaded, a negative loading on a factor indicates that the text does not deal with this topic. Table 4 lists the actual words that load most highly on each of the 18 factors. Words that are negatively loaded on the factors are in brackets. Note that each word factor suggests a different theme or meaning unit. The factors are centered around themes of religiosity (factors 1, 7, 17), war and Jihad (factors 2, 3, 6), emotions (factor 14), economics (factors 8 and 16), or Middle Eastern and Western politics and geography (factors 4, 5, 8, 9, 10, 11, 12, 13, 15, and 18). Analyses for Factors 1, 2, 3, 4, and 9 are highlighted for further discussion.

Table 4. Meaning Extraction Results: 18 Factor Solution

1	2	3	4	5	6	7	8	9
Guide	Bush	Crusade	Saudi	Palestine	Child	God	Important	Afghan
Companion	Operate	Islam	Region	Occupy	Kill	Pray	Resource	Iraq
Forgive	War	Muslim	Holy	Century	Woman	Almighty	Aware	Sudan
Praise	Iraq	Campaign	Occupy	Liberate	Innocent	Bless	Role	Victor
Message	Qaeda	Zion	Control	[Taliban]	Bomb	Faith	Grace	Chechnya
Family	Evident	Support	America	[Pakistan]	Death	Serve	Major	Palestine
Lord	America	Israel	Land	[Afghan]	Terror	Accept	Victor	Kashmir
Allah	Afghan	Pakistan	Peninsula	[Omar]		Jihad	UN	Somali
Muhammad	Add	Jew	Force	[Mujahid]		[Sake]	Crime	[Situation]
Patient	[Man]	Jihad	Month			[Allah]		
Mercy	[Allah]	Duty	Regime					
Bless	[Life]		Year					
[Country]	[Lord]		Egypt					
			Jew					
			State					
10	**11**	**12**	**13**	**14**	**15**	**16**	**17**	**18**
Issue	Banner	Human	NewYork	Sword	Plan	Economy	Ladin	Media
Clear	Lead	Free	Washington	Age	Achieve	Dignity	Usama	Huge
Event	[Verily]	Britain	Change	Sacrifice	Resist	Long	Shaykh	Deal
Point	[Friend]	Real	Expose	Head	Return	Oil		State
Sudan	[Unjust]	Begin	Respond	Turn	Egypt	Blood		[Remain]
Act	[Christian]	Destruct	[Abu]	Pride	Small	[Knowledge]		
Conflict	[Belief]	Face	[General]	Humiliate	Policy	[Arab]		
Nature	[Protect]	Day		Heart	International	[Affair]		
[Occupy]	[Campaign]	[Infidel]				[Order]		
		[Money]				[Abdallah]		

All words within a factor are positively loaded except for the words in [brackets]. Only words that have loadings of .25 or higher are included. Words are listed from the highest to the lowest loadings.

Factor 1, for example, can be thought of as a Compassion factor, made up of words such as guide, companion, forgive, praise, message, family, Lord, Allah, Muhammad, patient, mercy, bless. As can be seen in Figure 2a, Bin Ladin's use of compassion-relevant words has increased over the course of his career, peaking in 2002. Zawahiri's loadings are not significantly different from Bin Ladin's and have not fluctuated much in the last four years.

The remaining graphs in Figure 2 highlight meaningful trends in the topics that the authors are emphasizing. Bin Ladin, for example, has been dramatically increasing his references to Bush's involvement in Iraq and Afghanistan (Figure 2B) whereas he has been dropping his references to other Islamic hotspots such as Saudi Arabia, Egypt and the Holy Lands (Figure 2C) and Chechnya, Kashmir, Somalia, and Sudan (Figure 2D). In other words, one can easily see how Bin Ladin's early goals for Al Qaeda have been co-opted by the American involvement in Iraq and Afghanistan.

Whereas the content dimensions are generally quite similar for Bin Ladin and Zawahiri, the two authors are strikingly different about one issue. As can be seen in Figure 2C, Zawahiri is much more consumed by the Islamic-Israel conflict than is Bin Ladin. It is interesting that this topic is a feature of virtually all of his communications.

Message and target analyses. As can be seen in Table 3, Factor 2 (Bush's involvement in Iraq and Afghanistan) is a topic commonly discussed when addressing Muslim audiences but not those in the West. The topics relevant to Factor 4 (Saudi, Egypt, US, and Holy Lands) are used much more when the target audience includes Westerners.

Table 4 lists the use of the factors according to message type. The differences in use of factors may not be so surprising, considering that interview audiences (e.g. Al-Jazeera, Pakistani newspapers) have included a large Muslim population, whereas statements are intended for a wider audience. For example, Factor 1 (Compassion) was much higher in interviews, relative to statements. On the other hand, talk of broader issues, such as control and occupation of various lands (Factor 4) and international Islamic conflicts (Factor 9) was much more common in statements than in interviews.

Discussion and Summary

The ways in which public figures use words give us glimpses into who they are—much in the way that facial expressions, haircuts, and non-verbal gestures do. Because it is difficult for people to control their linguistic styles, the analyses of subtle word use is helpful in gaining insights into the ways people and groups think and relate to their conversational topics, their audiences, and perhaps to themselves. The Al-Qaeda findings hint at some of the social and psychological dynamics behind two of their leaders.

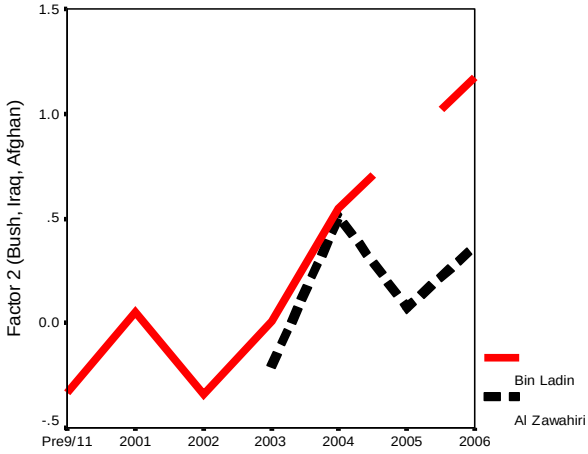
The LIWC analyses suggest that Bin Ladin has been increasing in his cognitively complexity and emotionality since 9/11, as reflected by his increased use of exclusive, positive emotion, and negative emotion word use. Also, since the 2003 invasion of Iraq, both Bin Ladin and Zawahiri have accelerated in their use negative emotion words – especially anger words. In fact, their use of anger and hostility words is much higher than that of statements from other known extremist groups. Also in comparison to other extremist groups, Al-Qaeda's sense of identity is more strongly defined through an oppositional group or government, as indicated by their higher use of third person pronouns.

The meaning extraction strategy results indicate that Al-Qaeda is increasingly focused on disseminating their interpretations of American involvement with Iraq and Afghanistan,

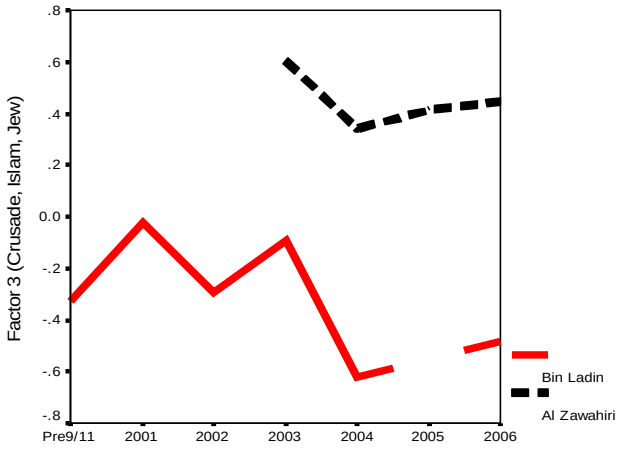
Figure 2: Changes in content over time and by author

A. Factor 1: Compassion

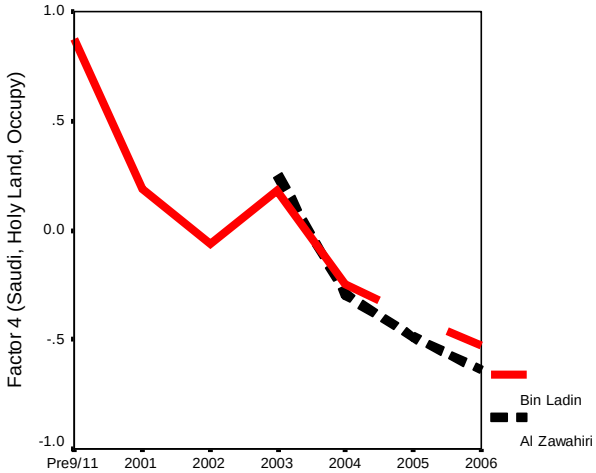
B. Factor 2: Bush in Iraq and Afghanistan



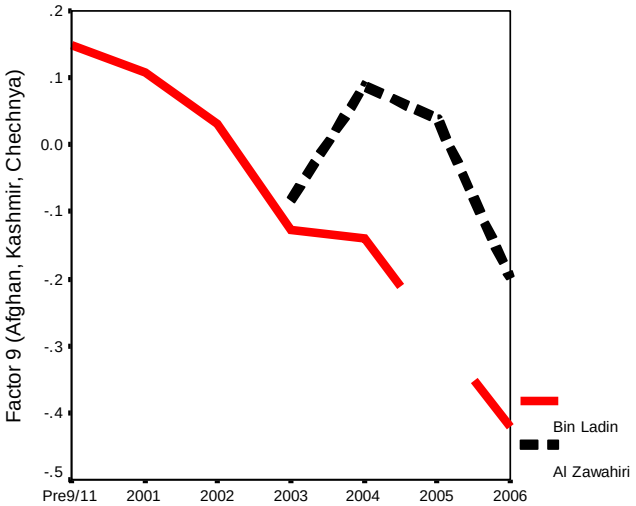
C. Factor 3: Islam-Israeli conflict



D. Factor 4: Saudi, Egypt, US, Holy Land



E. Factor 9: Worldwide Islamic conflicts



especially to Muslim communities. This increased attention to the West has replaced talk of Middle Eastern holy lands and other Islamic conflicts, which was the one of the group's main concerns in statements before 9/11.

Perhaps more important than the actual findings themselves is the introduction of new computer-based text analysis methods. The text preparation time involved spell checking and the removal of interviewers' and transcribers' comments along with extended religious quotations – a process that took approximately 4 hours. The LIWC analyses of linguistic styles of all 57 texts took approximately 12 seconds. Subsequent data analyses reported in this paper required perhaps 3 additional hours. In short, analyses of linguistic style of a large number of texts involved 7 hours and 12 seconds of work.

The meaning extraction method is much more experimental, a bit more time consuming, and involves greater interpretation once the process is concluded. The results reported in this paper involved perhaps 5-10 hours of text analysis work. The underlying logic of meaning extraction is to mathematically discover which groups of words tend to co-occur. These word clusters tend to reflect underlying themes. What makes the meaning extraction method appealing is that there is no predetermined categorization made by linguists, operatives, or even translators. Indeed, this method is not language-determined. We could do the exact same methods on Farsi, Arabic, or Korean language sets without being able to read a single word or character. The only time that translators and/or interpreters would be required is at the <u>end</u> of the analytic procedure.

We realize that our interpretations of the meaning extraction results are superficial. This is where the expertise of the intelligence and diplomatic communities is needed. As computer language analysts, we can say which words hang together. Unfortunately, without deep knowledge of the authors and context, we are restricted in knowing what the themes may reflect.

We are at the dawn of a new era in computerized text analysis. Through continued analyses of linguistic style and automated theme-based analyses, it will be possible to follow the individual and group dynamics of Al-Qaeda and other groups over time. Similar analyses can be conducted with almost any text-based documents surrounding Al-Qaeda or other groups. This could include blogs, posters, emails, overheard conversations, or public statements. With continued refinements in computational linguistics and cross-language research, it will soon be possible to bypass many problems in translation and examine statements in the language in which they were originally spoken or written.

References

Adams, S. (1996). Statement analysis: what do suspects' words really reveal? *FBI Law Enforcement Bulletin, October,* 12-20.

Adams, S. (2004). Statement analysis: beyond the words. *FBI Law Enforcement Bulletin. April,* 22-23.

Boroditsky, L. (2001). Does language shape thought? English and Mandarin speakers' conceptions of time. *Cognitive Psychology, 43,* 1-22.

Chung, C. K., & Pennebaker, J. W. (in press). The psychological function of function words. In K. Fiedler (Ed.), *Frontiers of Social Psychology.* New York, NY: Psychology Press.

Gottschalk, L. A. (2000). The application of computerized content analysis of natural language in psychotherapy research now and in the future. *American Journal of Psychotherapy, 54,* 305-311.

Gottschalk, L. A. & Gleser, G. C. (1969). *The Measurement of Psychological States Through the Content Analysis of Verbal Behavior.* Oxford, England: U. California Press.

Hart, R. P. (1984). *Verbal Style and the Presidency: A Computer-Based Analysis.* New York: Academic Press.

Landauer, T. K., & Dumais, S. T. (1997). A solution to Plato's problem: The latent semantic analysis theory of the acquisition, induction, and representation of knowledge. *Psychological Review, 104,* 211 – 240.

Martindale, C. (1990). *A Clockwork Muse: The Predictability of Artistic Change.* New York, NY: Basic.

Mergenthaler, E. (1996). Emotion-abstraction patterns in verbatim protocols: a new way of describing psychotherapeutic processes. *Journal of Consulting and Clinical Psychology, 64,* 1306-1315.

North, R. C., Holsti, O. R., Zaninovich, G. M., & Zinnes, D. A. (1963). *Content Analysis: A Handbook with Applications for the Study of International Crisis.* Oxford, England: Northwestern U. Press.

Oberlander, J., & Gill, A. (2004). Individual differences and implicit language: personality, parts-of-speech, and pervasiveness. In *Proceedings of the 26th Annual conference of the Cognitive Science Society,* 1035-1050. Chicago, IL, August 2004.

Pennebaker, J.W., Booth, R.J., & Francis, M.E. (2007). *Linguistic Inquiry and Word Count: LIWC 2007*. Austin, TX: LIWC (www.liwc.net).

Pennebaker, J.W., Francis, M.E., & Booth, R.J. (2001). *Linguistic Inquiry and Word Count: LIWC 2001*. Mahwah, NJ: Erlbaum Publishers (www.erlbaum.com).

Pennebaker, J.W. & King, L.A. (1999). Linguistic styles: Language use as an individual difference. *Journal of Personality and Social Psychology, 77,* 1296-1312.

Pennebaker, J. W., Mehl, M. R., & Niederhoffer, K. (2003). Psychological aspects of natural language use: Our words, our selves. *Annual Review of Psychology, 54,* 547-577.

Seyle, D. C. & Pennebaker, J.W. (2004). *"We're right, they aren't" A linguistic picture of extremism*. Paper presented at Society for Personality and Social Psychology, Austin, TX.

Simonton, D. K. (1988). Presidential style: Personality, biography, and performance. *Journal of Personality and Social psychology,55,* 928-936.

Smith, A. (2004). From words to action: Exploring the relationship between a group's value references and its likelihood of engaging in terrorism. *Studies in Conflict and Terrorism, 27,* 409-437.

Stone, P. J., Dunphy, D. C., Smith, M. S., & Ogilvie, D. M. (1966). *The General Inquirer: A Computer Approach to Content Analysis.* Cambridge, MA: MIT press.

Tsai, J. L., Simenova, D., & Watanabe, J. (2004). Somatic and social: Chinese Americans talk about emotion. *Personality and Social Psychology Bulletin, 30,* 1226-1238.

Walker, S. G. (2000). Assessing psychological characteristics at a distance: Symposium lessons and future research directions. *Political Psychology, 21,* 597-602.

Weintraub, W. (1981). *Verbal Behavior: Adaptation and Psychopathology.* New York, NY: Springer.

Winter, D. G. (2005). Things I've learned about personality from studying political leaders at a distance. *Journal of Personality, 73,* 557-584.