Behavioral Sciences of Terrorism and Political Aggression
Vol. 2, No. 2, May 2010, 108–132

Routledge
Taylor & Francis Group

Social language processing: A framework for analyzing the communication of terrorists and authoritarian regimes

Jeffrey T. Hancock,[a]* David I. Beaver,[b] Cindy K. Chung,[b] Joey Frazee,[b] James W. Pennebaker[b], Art Graesser[c] and Zhiqiang Cai[d]

[a]*Cornell University, 320 Kennedy Hall, Ithaca, NY 14853, USA;* [b]*University of Texas, Austin, Texas, USA;* [c]*University of Memphis, Memphis, Tennessee, USA;* [d]*Purdue University, West Lafayette, Indiana, USA*

(*Received 15 June 2009; final version received 3 November 2009*)

Social Language Processing (SLP) is introduced as an interdisciplinary approach to assess social features in communications by terrorist organizations and authoritarian regimes. The SLP paradigm represents a rapprochement of theories, tools and techniques from cognitive science, communications, computational linguistics, discourse processing, language studies and social psychology. The SLP paradigm consists of three broad stages: (1) linguistic feature identification; (2) linguistic feature extraction; and (3) classifier development. In this paper, we detail the SLP paradigm and review several linguistic features that are especially amenable to uncovering the social dynamics of groups that are difficult to assess directly (i.e. through questionnaires, interviews or direct observation). We demonstrate the application of SLP to identify status, cohesion and deception in the case of Saddam Hussein's regime. Specifically, we analyzed the memoranda, letters and public communiqués distributed within and from Saddam Hussein's administration in a recently recovered corpus called the Iraqi Perspectives Project, along with several related projects. We conclude with a discussion of the challenges that SLP faces for assessing social features across cultures in public and captured communications of terrorists and political regimes, along with responses to these organizations.

AQ1 **Keywords:**

Like any other organization, terrorist groups and aggressive political entities need to communicate with each other to coordinate their actions and beliefs and to communicate with the public the narrative that defines their cause. The content and style of such communication can reveal insights about the psychological states of the individual actors in the organization, including personality traits and emotional states. The content and style of communication can also provide clues about the social dynamics and functioning of the group, such as social status and the overall cohesion of the group.

In the present paper we introduce a research framework called Social Language Processing (SLP) that marries social and psychological theory with computational techniques for modeling the relationships between discourse and social dynamics. Specifically, SLP is the application of computational methods to automatically

*Corresponding author. Email: jeff.hancock@cornell.edu

ISSN 1943-4472 print/ISSN 1943-4480 online

DOI: 10.1080/19434471003597415
http://www.informaworld.com

analyze and classify psychological features of individuals and groups based on how language is used by members of the groups. For example, an SLP approach can help to inform questions such as: what are the key words that identify the leader in a group? What are the language patterns of group members that predict the stability of the group and likelihood of defection? What are the linguistic cues that mark when a leader or an organization is deceiving or concealing intent? SLP approaches seek to determine the degree to which people and groups are aligned and share a common perspective from the automatic analysis of their discourse.

The SLP approach is especially useful in the context of narratives by terrorist organizations and authoritarian regimes because of its indirect nature. SLP can be applied to captured internal communication, such as member-to-member conversations or military memoranda, or public communication produced by group leaders, such as Osama bin Laden's speeches. Here we describe how the SLP approach can be applied to narratives from terrorist organizations or authoritarian regimes to learn about three general social dynamics. Specifically, we show how SLP can be used to examine the status, cohesion and deception among members of a group.

We organize the paper into four sections. The first provides an overview of the general SLP paradigm. The second section provides a brief primer on the kinds of discourse features SLP might use. The third section describes various stages of SLP for the study of select social dynamics within terrorist organizations and authoritarian regimes. An SLP approach for understanding status, cohesion and deception in the case of Saddam Hussein's regime is showcased, primarily because of a recently available database, called the Iraqi Perspectives Project (Woods, Pease, Stout, Murray, & Lacey, 2006). This database includes a wealth of data on the regime's communication, including memoranda, letters and public communiqués. The final section is a conclusion that addresses both the challenges that such an approach faces as well as some ideas of how these challenges may be addressed.

The Social Language Processing Paradigm

The SLP paradigm is an interdisciplinary research effort that has roots in social and personality psychology, communication, language studies and the cognitive sciences. SLP also integrates automated text analyses that have recently emerged as a result of landmark advances in computational linguistics (Jurafsky & Martin, 2008), discourse processes (Pickering & Garrod, 2004; Graesser, Gernsbacher, & Goldman, 2003), the representation of world knowledge (Lenat, 1995; Landauer, McNamara, Dennis, & Kintsch, 2007) and corpus analyses (Biber, Conrad, & Reppen, 1998). Thousands of texts can be quickly accessed and analyzed on hundreds of measures in a short amount of time. These data are mined and farmed in an attempt to identify how language and discourse have interesting tentacles to social states of people and groups.

The SLP paradigm involves three general stages, each of which informs the other and can be recursive. The first stage is *identifying potential language features* that may reflect a given social dynamic, such as deception. This stage is grounded in theory and requires a deep understanding of the social dynamic under consideration, along with how the social dynamic may be manifest in discourse. For instance, research suggests that deceptive messages involve fewer first person singular pronouns (i.e. 'I' for English) than truthful messages (Hancock Curry, Goorha & Woodworth, 2008; Newman, Pennebaker, Berry & Richards, 2003). The second stage involves developing methods for *automatically extracting the relevant discourse*

features from communication. In our deception example, any program that can count the frequency of first person singular pronouns in a document could be used. Of course, other features of language are much more complex and require more sophisticated tools. Finally, the third stage is *developing statistical classifiers* that use message features to classify a message as belonging to one type of message (e.g. deceptive) or another (e.g. truthful). Below we lay out each of these steps in more detail.

Stage 1: linguistic feature identification

The process for developing fully automated SLP systems begins with the identification, based on theoretical grounds, of linguistic features that are predicted to correlate with the social phenomena under study. These linguistic features may be low-level features (such as individual word-counts), high-level features (such as discourse cohesion) or anywhere on a spectrum in between these extremes. In our deception example, an example of a low-level feature would be the use of first person singular pronouns associated with the motivation to take the spotlight off one's self when lying (Newman et al., 2003). A medium-level feature would be the rate of linguistic style matching observed between the liar and target (Hancock et al., 2008). Finally, an example of a high-level feature would be the degree of cohesion in deceptive texts, since liars are expected either (a) to have preplanned their stories to the point of showing advantages in cohesion over those who may struggle to express the truth or (b) to have difficulty composing coherent messages because of resources used in tracking their lies (Duran, McCarthy, Hall, & McNamara, in press).

Once theoretical or linguistic features have been identified as potential correlates of the social phenomena under study, empirical validation of these features is required. This step begins with an indexing of language data sets with the social features being studied. For example, a deception data set needs to have the deceptive portions of the text identified. Empirical tests are then performed to establish that the theoretically predicted correlations are present and statistically significant (e.g. that the linguistic feature of first person singular is related to deception). In our deception example, we know from multiple laboratory studies where participants have been induced to lie, from court transcripts, and from political leaders' speeches that first person singular pronouns are indeed associated with deceptive portions of texts (see Hancock et al., 2008). Interestingly, we have also found that, when participants are asked to lie to one another, partners tend to use the same pronouns, or more broadly, to linguistically style match *more* when one of the partners is lying than when both partners are telling the truth (Hancock et al., 2008).

Stage 2: linguistic feature extraction

In this stage, existing software can be modified or new programs developed that can automatically identify the linguistic features empirically found to be associated with the social constructs. For low-level features, this may require simple word counts. For higher level features, there is a need for more sophisticated natural language processing techniques, such as syntactic parsing and cohesion computation.

One word counting tool that is increasingly used in the social sciences is Linguistic Inquiry and Word Count (LIWC; Pennebaker, Booth, & Francis, 2007). LIWC reports the percentage of words in a given text devoted to grammatical (e.g. *articles*,

pronouns, *prepositions*), psychological (e.g. *emotions*, *cognitive mechanisms*, *social*) or content categories (e.g. *home*, *occupation*, *religion*). LIWC categories have been shown to be valid and reliable markers of a variety of psychologically meaningful constructs (Pennebaker, Mehl, & Niederhoffer, 2003). The LIWC software has an expansion capability that is especially useful for the linguistic feature extraction step in SLP; a user-defined dictionary made up of words hypothesized or empirically found to be associated with a social construct can be uploaded to LIWC, and LIWC will report the percentage of words in a given text that contains those words. In our deception example, we know from LIWC analyses that self-referencing and exclusive words (e.g. *except*, *but*) tend to decrease, while negative emotion terms and motion words (e.g. *go*, *move*) tend to increase.

For higher level features, one tool that is particularly well suited for SLP is Coh-Metrix (Graesser, McNamara, Louwerse, & Cai, 2004). Coh-Metrix analyzes discourse on different types of text cohesion. It automatically computes whether pairs of sentences/utterances are linked by different foundations of cohesion: co-reference (Clark, 1996; Halliday & Hasan, 1976), connectives and discourse markers (Louwerse & Mitchell, 2003), dimensions of situation models (such as temporality, spatiality, causality and intentionality; Zwaan & Radvansky, 1998) and latent semantic analysis (Landauer et al., 2007). Several dozen cohesion measures have been validated in a program of research that compares Coh-Metrix cohesion measures (or what is called text cohesion in the reports) with various forms of psychological data, such as expert annotations of cohesion, recall, and reading time (McNamara, Louwerse, McCarthy, & Graesser, in press). In our deception example, Duran et al. (in press) reported that deceptive senders of messages in conversation used more concrete words, had more complex syntax, asked more questions and had higher redundancy and cohesion in an analysis with Coh-Metrix.

AQ3

Stage 3: classifier development

The ultimate objective of this stage of SLP is to take the features extracted in the previous step to classify messages into relevant social dynamic categories. The first step in achieving this goal is using *supervised* machine learning techniques to build classifiers which categorize individuals or groups according to the social dynamic of interest (e.g. status, group cohesion, deception). Supervised machine learning builds classifiers by letting learning algorithms consult with human-annotated examples that serve as the gold standard. Unsupervised machine learning detects clusters inductively, without any comparison to a standard. The input of a classifier consists of both low- and high-level linguistic features. The output of a classifier is information about a social category of a person or an aspect of conversational interaction. In our deception example, a classifier would assess the probability of a text being deceptive or not, based on features such as first-person singular pronoun use, language style matching, and the cohesion of utterances in the texts. If the classifier is trained using supervision, the learning algorithm will have access to the deception ratings that humans assigned to each of the texts.

The next step involves automatic feature identification. Here, we apply more sophisticated computational techniques that involve *unsupervised* machine learning in order to automatically discover additional linguistic features that are predictive of the social features under study. That is, unsupervised machine learning achieves the same goal of classification without or with less human annotated data. Apart from their

theoretical interest, the automatic identification of features has two benefits. First, it may improve classifier performance. Second, it can enable the automatic transduction of classifiers in languages for which we have not previously hand-identified the relevant linguistic features. This can greatly speed up the development of SLP systems for new languages.

For any classifier, it is important to know for what range of domains and contexts the performance of the classifier is reliable. In particular, caution is essential when applying the classifier outside of the domain of the original training set (e.g. for other languages, for other modalities of communication such as written to spoken language). While there are some empirical techniques, such as error analysis, which can help in understanding the limits of classifier performance, theoretical insight is required to evaluate the extent to which the results of the classifier can be extrapolated beyond the original domain used in training. Thus, a theoretical understanding of the nature and distribution of social and linguistic features must provide guidance as to where the classifiers can safely be used, for example, the extent to which a deception classifier trained on Arabic chat room data could be applied to email or military transmissions.

It is informative to contrast SLP with fields that analyze language and discourse with a delicate palate that is sensitive to context and subtle interpretations of language contributions. For example, the field of Discourse Processing (Graesser et al., 2003) has a foundation in rigorous scientific methodologies, but is also sensitive to the linguistic, social, and contextual foundations of language. Those working in Conversation Analysis (e.g. Sacks, 1995; Sacks, Schegloff, & Jefferson, 1974) and Discourse Analysis (e.g. Blommaert, 2005; Schiffrin, Tannen & Hamilton, 2001) employ a fine-grained word-by-word manual analysis of the language, which requires interpretations of the words by human experts in the specific discourse context. In contrast, the SLP paradigm makes no prior assumptions about context other than whatever contextual features can be automatically extracted from text. This radical position makes SLP unlike any other current work on social dynamics in sociolinguistics, anthropology, psychology or communication studies, and makes it particularly suitable for analyzing the communication of terrorist groups and other aggressive political entities for which contextual information may be difficult to ascertain.

There are other characteristics of SLP that set it apart from sister fields. SLP concentrates on gross, automatically extractable statistical properties of segments of text. Most traditional linguistic researchers have adopted analytical methods drawn from linguistic, social and cultural theories, whereas SLP interprets data from the lens of social and personality psychology, a field that tends to be more quantitatively and statistically based and lends itself to automated analysis. Once again, these properties make SLP a useful approach in understanding terrorist organizations. Frequently, the interests in these groups are to understand their psychological characteristics, but given that assessment can only be indirectly observed through their communication and behavior, SLP can tie the discourse of individuals in these groups to psychological measures.

In summary, SLP is a multi-disciplinary approach to automatically detecting and classifying statistical properties from text, and relating them to social and psychological states of individuals and groups in the broader framework of discourse processing theories. Much of the SLP work reviewed in this paper focuses on Stages 1 and 2. Although the classification techniques for Stage 3 are in use by computational linguists for purposes other than SLP, the degree to which Stage 3 has been applied to

the understanding of social dynamics within an SLP approach is limited. In the present paper we demonstrate how these stages can be applied to understanding three social dynamics, status, cohesion and deception, by analyzing the captured and public communications of terrorist organizations and other political entities.

A Primer on Language Features

A key aspect to our approach is that we do not try to interpret the meaning of what is said in order to determine relationships within and between texts: predictions are made directly in terms of observable elements or near surface elements. As previously mentioned, our predictors range from low-level (e.g. word counts of function words) to high-level (e.g. text cohesion) linguistic features. Examples of classes of features that SLP employs are:

(1) word-level features, such as the words themselves and morphemes such as *-ed* and *-ing*;
(2) features referring to hand-crafted classes of words, such as positive affect;
(3) features that generalize over individual words to semantic groups (e.g. clusters resulting from statistical patterns of words in documents);
(4) linear word order, as with word *n*-grams;
(5) features that encode sentence structure (syntax); and
(6) discourse-level features identifying speech acts, given vs new information, and text cohesion.

Below we describe several features more fully in order of low-level to high-level features, and sketch out some of the ways these discourse types may be linked to social dynamics.

Function words

Function words are at a very low level of language analysis that is highly suitable for computational applications. Function words include pronouns, prepositions, articles, conjunctions and auxiliary verbs. This deceptively trivial percentage (less than 0.04%) of our vocabulary accounts for over half of the words we use in daily speech (Rochon, Saffran, Berndt, & Schwartz, 2000).

Given that function words are difficult for people to deliberately control, examining the use of these words in natural language samples has provided a non-reactive way to explore social and personality processes. In fact, most of the language samples that are analyzed for function word use come from sources in which natural language is recorded for purposes other than linguistic analyses, and therefore have the advantage of being more externally valid than the majority of studies involving implicit measures. Computerized text analyses in the last five years has helped to understand function words and their links to individual differences, psychological states and social processes (for a review, see Chung & Pennebaker, 2007).

Some of the basic findings of the work on function words have revealed demographic and individual differences in function word production. There are sex, age and social class differences in function word use (e.g. Newman, Groom, Handleman, and Pennebaker, in press; Pennebaker & Stone, 2003). For example, first-person singular pronouns (e.g. *I*, *me*, *my*) are used at higher rates among women, young people and

among people of lower social classes. Pronouns have also reliably been linked to psychological states, such as depression and suicide across written text, natural conversations and in published literature (Stirman & Pennebaker, 2001; Rude, Gortner, & Pennebaker, 2004; Weintraub, 1989).

Presupposition

Presupposition is the conventionalized marking of information as being assumed to be true, uncontroversial and taken for granted by the speaker. Presupposition occupies a unique place on the boundary between semantics and pragmatics. Perhaps because of this status as an interface between meaning levels, presupposition theory has been one of the active areas in semantics and pragmatics in the last few decades (Beaver, 1997, 2001; Beaver & Zeevat, 2007; Hempelmann et al., 2005). Presupposition is a ubiquitous feature of all text and conversation across all languages, and is signaled by function words, such as pronouns, definite articles, discourse connectives, and politeness morphemes and by many open class expressions, such as cognitive factive verbs (e.g. *realize*, *know*, *discover*), emotive factive verbs (e.g. *regret*, *be glad that*), and implicatives (e.g. *manage*, *succeed*).

There are features of presuppositions that make them particularly suitable for use in SLP. Presuppositions play a central role in establishing and marking the common ground of discourse participants (Beaver & Zeevat, 2007; Stalnaker, 1974). It is therefore anticipated on theoretical grounds that presuppositions should indicate group cohesion, and that there should be significant differences between use of presupposition among in-group and non-in-group interlocutors. We predict higher rates of presupposition use in discourse-initial segments of in-group conversation than discourse-initial segments of non-in-group conversation, and higher rates of presupposition use when participants are conversing about a topic of high salience to the group. Presupposition may also play a role in deception as liars attempt to introduce novel information as given.

Speech acts

Social states are expected to be manifested in the distributions of speech act categories in conversations. D'Andrade and Wish (1985) have identified a set of speech act categories that are both important classes of expressions in conversational pragmatics and that also can be identified by trained judges with a satisfactory interjudge agreement. The categories that they identified were command, indirect request, assertion, promise/denial, expressive evaluation, short responses, declaration, greeting, question and response to question. The leadership status and style of would presumably be manifested in their speech act categories. For example, leaders who act like prototypical drill sergeants should have a high percentage of commands whereas democratic leaders should have more questions and indirect requests. Followers and low-status individuals, in contrast, should have a higher percentage of speech acts in the response-to-questions, promises, and short verbal-response categories.

Discourse patterns can be measured by observing adjacent speech act categories between speakers. These adjacency pairs have been extensively analyzed in the field of conversational analysis, starting with the pioneering publication of Sacks et al. (1974), who identified the common adjacency pairs in Western cultures. For example, it is conceivable that effective leaders are responsive to the followers. If so, then a

question by a follower should have a high likelihood of being answered by the leader (i.e. a high question → response-to-question percentage). Groups that follow the conventional adjacency pairs of a culture would presumably function more productively and have higher group cohesion. Individuals in a group who are ignored or who have low status would not receive responses by other group members that follow the conventional adjacency pairs. In summary, we hypothesize that a number of social states (such as leadership, group cohesion) could be predicted by the distribution and sequencing of speech act categories.

An automated detection of speech act categories has been one of the goals of the SLP paradigm. One hypothesis is that there is a language universal that the speech act categories are telegraphed by the initial words of main clauses in sentences or utterances. An early speech act classification would assist comprehenders in performing pragmatic and semantic analyses of messages. Indeed, we have found that the first four words in English and in Arabic are very diagnostic of speech act category. An analysis of over 400 speech acts in Arabic newspapers and television programs revealed that the speech act categories of 88% of the sentences/utterances could be identified by the first four words (Graesser et al., 2009).

Cohesion

As reviewed in the SLP introduction, Coh-Metrix identifies the cohesion of texts using multiple measures. Social states are expected to be predicted by cohesion measures. For example, we predict that the leadership status of speech participants who interact by email, letters or oral conversation can be identified. It is plausible that the leader of a group might have the least cohesion and language quality because every other member of group is working hard to get their messages across. A leader has a more complex agenda, is less concerned with justifying their statements, and is more succinct. Cohesion within individuals and between individuals would be expected to predict *commitment*, as well as the *familiarity* of participants in a conversation.

Given–new

Given–new has been developed as a statistical measure that computes the amount of new information vs given (old) information in the discourse context (Hempelmann et al., 2005). Given information is recoverable either explicitly or inferentially from the preceding discourse, whereas new information is not recoverable (Halliday, 1967; Haviland & Clark, 1974; Prince, 1981). The Coh-Metrix measure for given–new was based on a variant of latent semantic analysis (LSA), a statistical representation of world knowledge that is manifested in text in very large corpora of 10 million words or more (Landauer et al., 2007). The LSA-based computation for given–new is called *span* (Hu et al., 2003). The span method statistically segregates the extent to which an incoming sentence presents new information vs old information when compared with the previous discourse history in a conversation or text. Hempelmann et al. (2005) reported that the span method has a high correlation with given vs new information in a sample of texts annotated by human experts applying Prince's given–new scheme.

Using the span method, social states can be predicted by the newness of information in discourse. People who take on important roles or have a high leadership status are predicted to contribute higher newness scores. A group may be more productive

when the newness scores are higher. The LSA-based metrics of given–new can be computed for any language as long as an LSA space has been created for that language.

In summary, LIWC and Coh-Metrix are promising tools for automatically analyzing the discourse in print and in oral conversation. Together they provide a comprehensive analysis of language at all levels, from words to syntax to discourse cohesion. Following the SLP paradigm, there are systematic relations between the features of language and discourse on the one hand to social states of people and groups on the other.

SLP applied to status, cohesion and deception

In the following section, we describe the SLP paradigm for the study of status, cohesion, and deception in terrorist and authoritarian regimes. Much of the SLP work has been in establishing foundations in Stage 1 of SLP, linguistic feature identification, so the work reviewed here is heavily focused on those foundations. Stage 2 of SLP, linguistic feature extraction, is showcased in an analysis of one sample of Arabic discourse. Finally, Stage 3, classifier development, is discussed with respect to the status and deception social dynamics.

Status and group dynamics

SLP Stage 1: linguistic feature identification

Although each has a distinct meaning, the terms ‘authority’, ‘dominance’ and ‘status’ are often used interchangeably; they are all determinants of power (Keltner, Gruenfeld, & Anderson, 2003). Several studies suggest that we can identify the status of group members by examining their language use. Across four studies with both laboratory manipulations of status as well as natural status from the analysis of emails, we have consistently found that the person with the higher status uses fewer I-words (Kacewicz, Pennebaker, David, Jeon, & Graesser, submitted). Similar patterns of effects have been found in the analyses of several of the Watergate tapes between Nixon and his aides (Chung and Pennebaker, 2007). Those aids (e.g. H.R. Haldeman) with the most egalitarian relationships with Nixon used I-words at comparable rates; those aids who were viewed as more subservient (e.g. John Dean and John Ehrlichman) used I-words at rates of over 50% above Nixon (see also study 3 from Niederhoffer & Pennebaker, 2002).

The linguistic correlates of status have been examined in two of al-Qaeda’s leaders bin Laden and al-Zawahiri (Pennebaker & Chung, 2008). First, the 58 translated al-Qaeda texts were compared with those of other terrorist groups from a corpus created by Smith (2004), which included the Sicarii group of ancient Palestine, the Front du Liberation du Québec, the Shining Path, Hamas and the Army of God. Compared with other extremist groups, the texts from al-Qaeda were more emotional, angry and oriented towards other groups and governments as evidenced by the use of third-person plural pronouns.

Similarly, the function word use by bin Laden and al-Zawahiri was tracked over time (Pennebaker & Chung, 2008). Overall, bin Laden evidenced an increase in the rate of positive emotion words as well as negative emotion words, especially anger words. He also showed higher rates of exclusive words over the last decade, which

often marks cognitive complexity in thinking. Particularly interesting, the data from al-Zawahiri evidenced a surprising shift in his use of first person singular pronouns over the last two years (see Figure 1). Such a pattern in the use of first person singular pronouns (Pennebaker et al., 2003) is indicative of greater insecurity, feelings of threat and, perhaps, a shift in al-Zawahiri's relationship with bin Laden (see Kacewicz et al., submitted; Tausczik & Pennebaker, in press). Overall, al-Zawahiri tended to be slightly more positive and significantly less negative and less cognitively complex than bin Laden in his statements.

SLP Stage 2: linguistic feature extraction

The analyses of status reviewed above, both quantitative and qualitative, show that easily extractable textual features are indicative of status and relationship, but should it be expected that these insights will transfer to other domains, such as memos produced by the Iraqi army? And if they can, will this allow us to develop practical tools to tap them? By performing a series of initial studies, we are convinced that the answer to both these questions is yes.

At this point we turn to the results from an initial study that included a sample of 60 letters we drew from the Iraqi Perspectives Project (IPP; Woods et al., 2006), which we refer to as the IPP Memo Corpus. Condition 1 included 20 letters from low-status senders to high-status recipients (low–high letters). Examples of these were letters from a low-status member to a deputy supervisor, a low-status member to a director, a staff officer to a chief of staff, or a director to a secretary of the president.

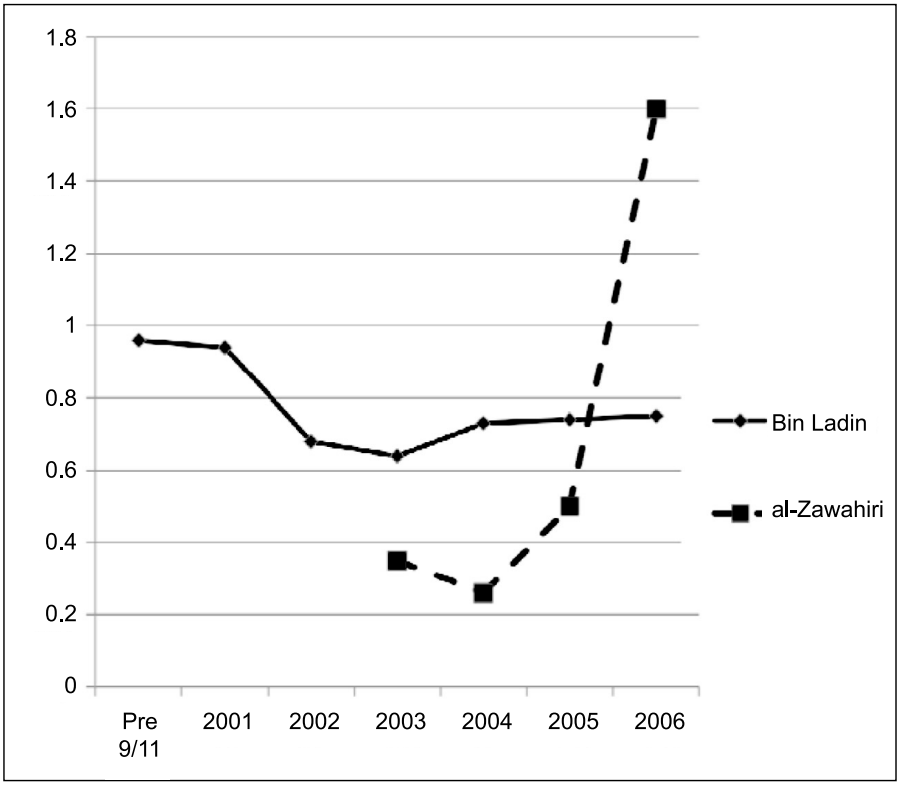

Figure 1. First person singular pronoun use by bin Laden and Al Zawahiri pre-9/11 to 2006.

Table 1. Mean and standard errors of language features across status conditions (high-status to low-status messages, same status messages, low-status to high status messages).

	High–low M (SE)	Same status M (SE)	Low–high M (SE)
Word count	63.85 (10.07)	88.70 (19.62)	187.95 (25.70)
First person singular 'I'	0 (0)	0.55 (0.29)	0.38 (0.23)
Second person 'You'	2.20 (0.48)	1.32 (0.28)	0.38 (0.16)
Cognitive mechanisms	11.56 (0.73)	14.36 (0.96)	14.18 (0.97)

Condition 2 included 20 letters from high-status senders to low-status recipients, which was the flip side of the above. Condition 3 had 20 letters between senders and receivers of the same status.

We first performed an analysis using LIWC in order to calculate the frequency of a range of words and word classes. Overall, high-status people used fewer words when writing to lower status, $F(2,57) = 11.28$, $p < 0.001$. A number of textual features were significantly correlated with the relative status of the sender and receiver, and these results are described in Table 1. We conducted contrast analyses comparing the means of the three different statuses against each other for each of the relevant linguistic dimensions. First, low-status individuals used marginally more first person singular subjects ('I') than high-status [$t(57) = -1.86$, $p = 0.07$], while relatively high-status individuals used 'you' significantly more than low-status [$t(57) = 3.82$, $p < 0.001$]. Second, high-status individuals used far fewer cognitive mechanism words (words indicating cause, discrepancy and inclusion) than relatively low-status individuals [$t(57) = -2.07$, $p < 0.05$], suggesting lower cognitive complexity. These results, which are completely in line with the predictions of our prior results on status, demonstrate not only that our techniques are applicable, but also that the techniques apply to documents that have been hand-translated.

SLP Stage 3: developing classifiers for determining status

We used statistical machine learning techniques to automatically classify speakers as low, medium, or high status. Classification models were trained using shallow, non-domain-specific features (e.g. number of pronouns and non-pronouns, whether the paragraph contained numeric words, whether the paragraph contained date/time keywords, punctuation) in combination with the speaker–hearer status combination. The techniques as applied, crucially, do not have or need access to any additional information (e.g. whether a sentence was a command or question or what kinds of communications are tied to speaker and hearer status). Rather, we are able to automatically learn the relative importance of variables like those discussed in Pennebaker et al. (2003) and Pennebaker & Chung (2008), which allows us to accurately determine speaker status using text alone.

The two machine learning techniques we used were based on Maximum Entropy models (Berger, Della Pietra, & Della Pietra, 1996) and vector-space models, specifically support vector machines (hence SVMs) (Joachims, 1998). Here we treated the problem as one of building an automatic document classifier, with the goal of automatically classifying documents according to the relative status of the writer and intended recipient. Both machine learning techniques are in wide use across computer science. Maximum entropy, the machine learning equivalent of what in statistics is

known as logistic regression models, is used in a wide range of natural language processing tasks. SVMs are known to be both efficient and effective in many document classification tasks (Joachims, 1998). In a nutshell, SVMs treat bunches of complex features (like word frequencies) as separate dimensions, so that each document is represented as a point in an abstract multi-dimensional space. The classification task then amounts to figuring out how the position of a point corresponds to the classification of a document. SVMs make this positional identification by finding the best way to draw a line (or, more generally, a hyperplane) between points such that everything on one side falls in one class, and everything on the other side falls in the other class.

We performed multiple runs for each experiment, each time dividing the 60 letter corpus into a training set and a test set by selecting letters randomly from the total pool of letters. We varied the size of the training set up to 45 letters, for which case the test set was 15 letters, in order to see how performance was related to corpus size. We also ran two separate tasks. For the *multiclass* task, the problem was to classify letters according to a three-way split based on the relative status of the sender of a message and the recipient, so the classes were: *high–low*, *same*, or *low–high*. For the *binary* classification task, the problem was to classify the subset of transmissions for which there was a status difference as *high–low* or *low–high*.

The results are shown in Figures 2 and 3. These results report classification accuracy, which is the ratio of correct classification decisions over all classification decisions. Accuracy, here, can be viewed as an indicator of whether the classifier will make correct predictions for hypothetical letters in some future corpus. In all graphs, the low, solid line is the classification baseline, which is the classification accuracy that would be expected from naive, probability-based guesswork using the proportions of different document classes in the sample, but without looking at the documents. Thus for the binary classifiers, the baseline hovers around 50%, and for the three way classifiers, the baseline is around 33%. The baseline varies because the letters used for testing were chosen randomly for each trial, hence test conditions vary slightly for each trial. Importantly, for all types of classifiers, performance is well above the baseline for training sets of 20 or more letters.

Figure 2 shows performance on the binary classification task, including Maximum Entropy models and SVMs. The non-solid lines in the graphs represent the average performance of classifiers of various types across multiple runs. Each of these lines represents classifiers that were trained using different features of the documents. The *Word* classifiers had access only to (a limited range of) word frequency information, such as overall word count and pronoun frequency; the *Complex* feature classifiers had access to features such as the author's use of person prefix words (Dr, Mr, etc.), numbers and calendar expressions; and the *All* classifiers had access to both basic word frequency information and complex features. For both Maximum Entropy and SVMs, the best classifiers using a maximally sized training set are those which have access to all features. For these, average performance is close to 80% for Maximum Entropy models, and above 80% for SVM models, with accuracy reaching 90% accuracy on some runs. In general, SVM performed better than Maximum Entropy for all our experiments, but the differential is not large.

Further, the graphs show that, for smaller data sets, classifier performance is often better with fewer features, but extrapolating the trends in the graphs suggests that with a larger data set the best classifiers would continue to be those using a wide range of features. In other words, it does not seem to be the case that status can be determined

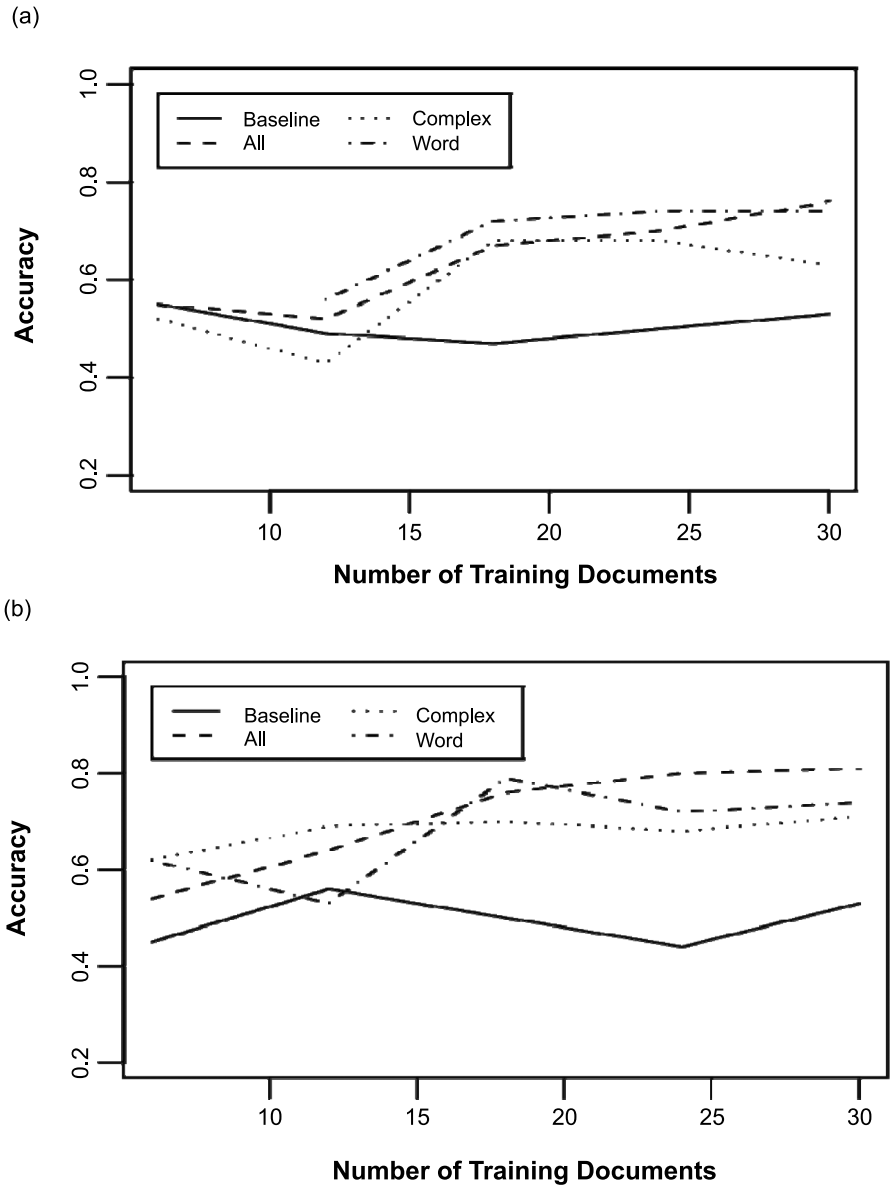

Figure 2. Two-way status classifier performance for high–low vs same vs low–high status in military transmissions in the IPP database. (a) Performance of binary maximum entropy classifiers. (b) Performance of binary SVM classifiers.

by using just one or two features, like address forms: a wide range of features of the text carries information about status, and given a sufficiently large data set it is possible to build classifiers that are sensitive to all of them.

Figure 3 shows performance on the multiclass task for Maximum Entropy and SVM models. The best average performance was obtained using SVMs with the combined (All) feature set, with average performance at above 75% for larger training sets. This performance is obviously much greater than the 33% baseline performance, but in addition, the upward trend of the graph suggests that significantly higher levels of performance could be attained using larger datasets: no clear performance ceiling is in sight for this task.

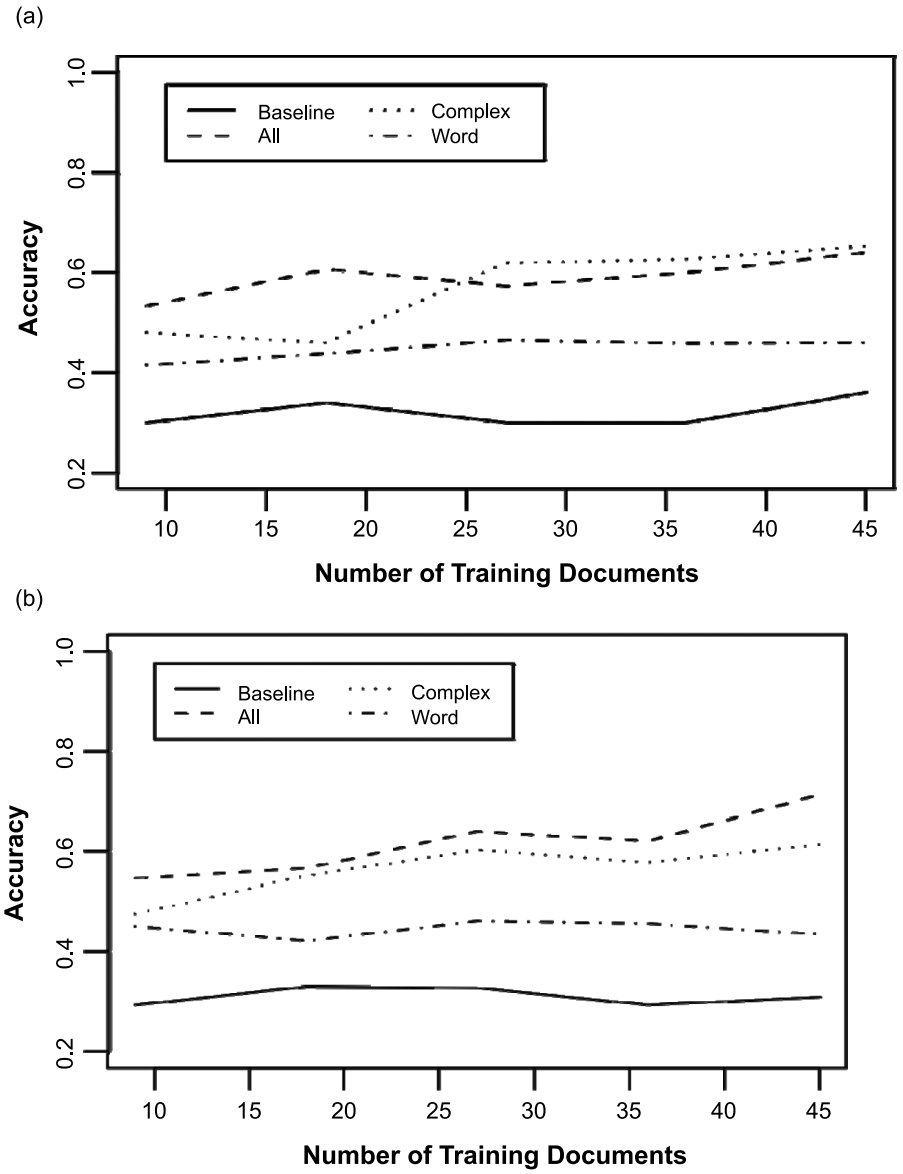

Figure 3. Three-way status classifier performance for high–low vs same vs low–high status in military transmissions in the IPP database. (a) Performance of three-way maximum entropy classifiers. (b) Performance of three-way SVM classifiers.

The fact that all the classifier lines are well above the baseline shows graphically that all our machine learning classifiers performed significantly better than chance. This result would already suffice to show the potential of the SLP approach for status classification, but in fact, our results go well beyond showing the possibility of building status classifiers: many of our classifiers operate at a sufficiently high accuracy that they could effectively be used for automatic status classification information of large numbers of documents. As mentioned, for the binary task our best classifiers were correct 90% of the time. Although these best-classifier results were not robust in our experiments, it is plausible that at least this accuracy could be achieved reliably,

even on the multiclass task, given further optimization of feature sets and the machine learning models.

In the future we plan on testing much richer feature sets. These include (a) the more linguistically sophisticated features that can be obtained by running part of speech taggers and parsers over the texts, and textual cohesion features from the Coh-Metrix toolkit, and (b) demographic and social information obtained by our expert Iraqi consultants, including ideally longitudinal information on the personal histories of individual authors represented in the corpus.

Increasing the corpus to include more letters should significantly improve performance and reliability. Including more documents types can sometimes negatively impact classification accuracy, since a small sample may just happen to contain documents that have very clear differences while those lines can sometimes be blurred in a large corpus. Still, larger corpora do often lead to higher classification accuracy, and the resulting classifiers should be more robust in the face of documents of unknown provenance.

Cohesion in text and social processes

SLP Stage 1: linguistic feature identification

The cohesion of language and discourse is diagnostic of the affiliation between speech participants, and may indicate the status and personality of particular speech participants. According to the linguistic and discourse *style matching* hypothesis, a speech participant will copy the linguistic style of another individual who is a close friend, a respected person, a leader, a celebrity, and so on. If this hypothesis is correct, then we should be able to observe the language profiles of particular speakers being a mirror image of other speakers. Group cohesiveness should be a direct function of the similarity in style matching according to the language-based group cohesion (LGC) hypothesis. These predictions address relations between speakers. Predictions can also be made on the language and discourse cohesion of individual speakers. For example, the cohesion of a message, letter or document of a particular speaker should reflect his/her status, personality and the leaders the person admires.

Our research team has reported analyses of pronouns that are compatible with the style matching and LGC hypothesis. The supporting evidence can be found in the Boeing 727 cockpit simulator studies (Sexton & Helmreich, 2000), published literature of the Weatherman Underground (McFadden, 2005) and the Watergate tapes (Mullen, Chapman, & Peaugh, 2001). In a series of studies investigating groups working on a cooperative task (Gonzales, Hancock, & Pennebaker, in press), style matching scores positively predicted both cohesiveness ratings and objective performance measures, such as the effectiveness of the groups' decisions. In the context of military, political or terrorist groups, these data suggest that analyzing the discourse cohesion of a group can reveal important information about how well-functioning that group is.

SLP Stage 2: linguistic feature extraction

In order to examine how cohesive (or how well-functioning) members of an authoritarian regime were towards other members, we conducted a cohesion analysis on the IPP Memo Corpus using Coh-Metrix (Graesser et al., 2009). Compared with

high-status letters, low-status letters had a higher volume of language (words, sentences, paragraphs), higher semantic cohesion (based on latent semantic analysis scores of adjacent sentences), lower lexical diversity, higher cohesion at the conceptual level (causal, intentional, and temporal cohesion), a higher incidence of connectives that link clauses, and a higher incidence of logical operators. In essence, the low-status letter writers expended more effort in writing the letters and produced more coherent letters. Quite clearly, the status of the letter writers and the recipients had a substantial impact on the language and discourse of the letters.

These results illustrate how CohMetrix can be used to dissect the language and discourse of texts at a fine-grained level. In future research we plan on analyzing the texts in the Iraqi Perspectives Project, as well as other corpora, to explore a number of research questions. For example, the status of the letter writers can be scaled on military, religious and ideological criteria. Military rank is the obvious criterion, but it is likely that the status of the individuals can be ranked on religious and ideological criteria. Also, the style matching of letter writers can be assessed by analyzing the stream of letters in exchanges. If letter $N + 1$ has a similar language and discourse profile to letter N, then there is evidence of linguistic and discourse style matching. Finally, if the IPP database has conversations between military personnel and Saddam Hussein, we can assess style matching of the language of others with the language of Hussein. The style matching hypothesis predicts that a positive correlation with Hussein's language may predict group success.

A different approach investigates the speech acts expressed by conversational participants. As discussed earlier, it is expected that the distribution of speech act categories of a person in a group will be diagnostic of social status and the leadership style of leaders. For example, do leaders control their soldiers with direct commands, or do they do so more indirectly with questions and indirect requests? Are there differences between task leaders and social leaders? As discussed earlier, the sequences of speech act categories between conversational participants are also expected to be diagnostic of social status, leadership and group cohesion. For example, individuals who are rejected in a group may not have questions answered and their indirect requests may be denied.

In a recent analysis of speech act classification, we collected a sample of sentences in Arabic from the Internet. There was a total of 261 Arabic sentences sampled from the Aljazeera TV channel and a total of 148 Arabic sentences sampled from the *Alhayat* newspaper. Each sentence was classified by two native speakers in Arabic based on D'Andrade and Wish's (1985) speech act classification: *command*, *indirect request*, *assertion*, *promise/denial*, *expressive evaluation*, *short responses*, *declaration*, *greeting*, *question* and *response to question*. There was high inter-judge reliability in the classification ($\kappa = 0.89$), which confirms D'Andrade and Wish's claim that these categories can be reliably classified by trained judges. The next step is to improve our existing automated classifiers of speech acts to match the high reliability of humans. Some speech act categories, such as indirect requests, will be difficult to identify because they rely on a deeper discourse context for classification. However, our current speech act classifiers (Graesser et al., 2009) are quite respectable for most of D'Andrade and Wish's speech act categories. A detailed analysis of speech acts (as well as other linguistic information) will enable us to perform comparative studies in both controlled and relatively free media outlets across Arabic nations. For example, to what extent do military personnel and citizens produce speech acts and other language patterns that match the style of political leaders in those countries? Does a

successful leader talk like Saddam Hussein? The SLP project should yield qualitative and quantitative measures of governmental influence and control on national and local media, providing a new way of studying the relation between cultural and political institutions in authoritarian and non-authoritarian regimes.

To date we have not moved into the third stage of building automated classifiers for cohesion social dynamics, although the process will be similar to the classifier development described in the section on status described above. Nonetheless, the analyses reported in this section demonstrate how leadership, social familiarity and group cohesion have systematic tentacles to language and discourse. These social states are diagnostically manifested in pronouns, style matching, speech acts, cohesion and other features of language and discourse. Moreover, LIWC and Coh-Metrix can automatically perform the analyses on large text corpora, as opposed to relying on the thoughtful judgments of language experts. This ability allows for the analysis of communication of terrorist organizations to identify, for example, members who may be more or less cohesive with the group, or identifying more or less with the leader of the group. Those that are less cohesive may be more willing to defect from the group, while those that are more cohesive may be more dangerous.

Deception and misinformation

SLP Stage 1: linguistic feature identification for deception

Deception can be defined as an intentional attempt to create a false belief in the receiver (Buller & Burgoon, 1996). A recent but growing body of research suggests that linguistic and discourse traces can be automatically extracted from deceptive language (Hancock et al., 2008). Can these traces be used in our SLP paradigm to assess deception or deceptive intent in political, military or other national security contexts?

The majority of previous research has been grounded in theories that focus primarily on the non-verbal cues associated with deception (Ekman, 1985; Zuckerman, DePaulo, & Rosenthal, 1981). For example, non-verbal 'leakage' cues are assumed to reveal hidden emotions that are manifest in unconscious and uncontrolled movements of the face or body. This approach tended to ignore verbal cues presumably because non-verbal behavior is assumed to be uncontrollable, while speech is assumed (incorrectly) to be controlled (Vrij, 2008). Overall, this approach has not lead to promising results, with meta-reviews indicating that there are very few non-verbal cues reliably related to deception (DePaulo et al., 2003), and that humans, who primarily rely on non-verbal cues, are notoriously poor at detecting deception, with a recent meta-analysis (Bond & DePaulo, 2008) indicating only slightly better than chance accuracy (54%).

Since the early 1990s, however, theories of deception have begun to consider the verbal and linguistic aspects of deception. For example, Information Manipulation Theory (McCornack 1992) draws on Grice's cooperative principle and assumes that when people lie they violate one of the cooperative principle's four maxims of quality (veridicality of an utterance), quantity (amount of information in an utterance), relevance (relatedness to prior utterances) and manner (clarity of an utterance). These violations are assumed to have detectable linguistic manifestations.

Other theories have begun to emphasize the cognitive and motivational consequences of deception on language use, such as Criteria-Based Content Analysis (CBCA, Köhnken, 1996) and Reality Monitoring Theory (Johnson & Raye, 1981).

For example, CBCA provides 18 verbal cues that are associated with the cognitive and motivational aspects of truthful accounts, which are assumed to include more detail, be more logically coherent, contain more spontaneous corrections and include more quoted speech. Similarly, Reality Monitoring Theory assumes that descriptions of real memories of an event differ from imagined or fabricated memories, such that descriptions of real memories will contain more perceptual and contextual information than false memories (e.g. Vrij, Edward, Roberts, & Bull, 2000).

Taken together, these theories all suggest that deception should be reflected in language. Recent empirical evidence using human coding techniques (i.e. non-automated) provides consistent support for this assumption (see Vrij, 2005, for reviews of CBCA research and Masip, Sporer, Garrido, & Herrero, 2005; Vrij, 2008, for reviews of Reality Monitoring research). A recent meta-review of studies examining verbal features of deception revealed that the majority of studies found support for speech content differences across deceptive and truthful language (Vrij, 2008), prompting the author, a leading deception scholar, to conclude that attending to language features can lead to more reliable deception detection than non-verbal cues.

Nonetheless, there are a number of challenges in applying SLP to deception, especially in terrorist or political contexts. First, the vast majority of research on deception has focused only on English from Western cultures (see Bond, Omar, Mahmoud, & Bonser, 1990). The scant research that has examined deception patterns across cultures suggests that there are some culture-specific differences in how deception is perceived, but that there are also some principles of deception that may be universal across cultures (Zhou & Lutterbie, 2005). For example, the collectivism–individualism dynamic can affect whether pro-social lying designed to maintain harmonious relationships is perceived as acceptable or not, with people from collectivist cultures viewing these pro-social lies as more acceptable than people from individualistic cultures (Lee et al., 1997). This is significant in view of the hypothesis that Arab cultures adhere to an ideal of collectivism in social interaction.

The second challenge to bringing deception research to bear on questions concerning terrorism and authoritarian regimes is that most deception research takes place within fairly static and highly controlled laboratory studies that may or may not generalize to the real world. Most studies involve people writing essays that are either truthful or not, telling true or false stories to a small audience, or being interrogated about a mock crime (see DePaulo et al., 2003; Newman et al., 2003). Critically, the lies generated in most laboratory studies are mocked up, and liars have little motivation to succeed in their lies.

One way around this problem is to examine language and deception that takes place in the 'real world.', such as the Iraq War Claims database that catalogues all the false statements made by the Bush administration when making the case for the Iraq war (Hancock, Bazarova & Markowitz, submitted). Similarly, the tremendous amount of ground truth established regarding Iraq's major military and political actions in the IPP database described earlier provides another example to which the SLP paradigm can be applied in analyzing deception. By comparing (a) the actual events and actions of the Iraqi political and military leadership with (b) their stated actions and intentions, we can identify real-world, motivated deceptions and campaigns of misinformation. A number of key questions can then be asked, including can we detect features of deceptive language in translations from Arabic or from the original Arabic, and if so can we build tools that can be used to automatically extract and classify messages as deceptive? In the next phase of SLP, however, we

must first develop techniques to automatically extract linguistic features theoretically linked with deception.

SLP Stage 2: linguistic feature extraction

The review above suggests that there are some extractable features related to language. One example is the empirically-derived Newman–Pennebaker (NP) model of deception, which finds that several language features predict deception, including fewer first person singular, fewer instances of exclusive conjunctions (words such as except, but, without), and more negative emotion terms (Newman et al., 2003). While this model was derived from controlled laboratory studies, this linguistic pattern has also been observed in deception by prison inmates (Bond & Lee, 2005) and most recently in courtroom testimonies of defendants who were found guilty of a crime and of perjury (Pennebaker & Huddle, 2008, Detecting deception with courtroom transcripts, unpublished data).

Hancock has found similar patterns within laboratory studies (Hancock et al., 2008) and in political speeches (Hancock et al., submitted). For example, Hancock and his colleagues compared false statements (e.g. claims that Iraq had WMD or direct links to al-Qaeda) and non-false statements (e.g. that Hussein had used gas on his own people) produced by officials in the Bush administration in the run-up to the Iraq war. Consistent with the NP model of deception's predictions, false statements contained substantially reduced rates of first-person singular ('I') and exclusive terms ('except', 'but') but more negative emotion terms and action verbs.

Moving forward, we believe that a similar approach and methodology can be applied to other political groups. As noted, the effort that has gone into understanding Iraqi military and political actions and organization prior to and during the war provides substantial opportunities to examine linguistic traces of deception. Consider, for example the case of Abu al-Abbas, a foreign agent who worked actively for the Iraqi Intelligence Service (IIS). In one report by an IIS officer, Abu al-Abbas claimed success in several missions (e.g. burning of the Japanese Embassy in Manila, Philippines; placing an explosive device near an American base in Izmir). The conclusion of the analysts, however, was that 'the possibility is good that either the IIS or Abu al-Abbas himself embellished, overstated, or even falsely reported some exploits' (p. 30, *Iraqi Perspectives Project*). Is it possible to detect such false statements within the military organization from extracts in the IPP database?

A second example is the statements by Iraqi political leadership about supporting terrorism, including the following (from p. xiii, Iraqi Perspectives Project):

> when they say anything about Iraq – [like] Iraq supports terrorism – then they have to say that Iraq has documents on this issue and [we] don't. (Saddam Hussein, 1993)

> It has never [been] proven that Iraq participated in a terrorist operation. (Tariq Aziz, 1996)

In contrast to these claims, the Iraqi Perspectives Project (IPP) uncovered substantial documentary evidence indicating that, although Saddam had no direct ties to al-Qaeda, the regime funded and supported pan-Arab terrorist causes and emerging pan-Islamic radical movements during the time of these statements. As such, these statements can be annotated and indexed as deceptive in the IPP database and analyzed using the SLP paradigm for uncovering linguistic and discursive traces of deception.

An important question to be addressed in future research is whether the linguistic markers identified in English hold or change across cultures. There are important linguistic differences, such as differences in obligatory evidentiality between English and Arabic (e.g. Isaksson, 2000), that may affect how deception is signaled in language. In our work we take the approach advocated by Zhou and Lutterbie (2005), in which bottom-up language patterns are identified statistically without reference to psychological expectations, while top-down approaches will guide specific analyses (e.g. speakers psychologically distance themselves from their lies, which theoretically should be expressed in languages across cultures).

SLP Stage 3 – developing classifiers for deceptive messages

Our group's work on this stage of the SLP program is limited to simple, supervised classification techniques relying on logistic regression. Newman et al. (2003) took the features described in the NP model (e.g. first person singular, exclusive words, negative emotion words, motion words) and used them to classify truthful and deceptive messages about abortion attitudes produced by college students. The logistic regression model, using the LIWC-extracted features, predicted deceptive messages at a rate of 67%, which was significantly above chance (50%) and better than human judges (52%). In a second study by one of our group, Hancock et al. (2008) used a similar logistic regression classification approach to identify deception in student conversations. This model, which included word count, first person singular and third person pronouns, causal terms (e.g. because, so), negations and terms related to the senses (e.g. see, hear, feel), correctly classified 66.7% of the messages as deceptive. Once again, the classification model outperformed chance (50%), and was a similar rate to the Newman et al. (2003) results noted above. A third study involved the classification of the Bush administration's false vs non-false statements about Iraq and its possession of weapons of mass destruction and links to al-Qaeda (Hancock et al., submitted). As described above, this classification task involved using the NP linguistic features (i.e. first person singular, exclusive terms, negative emotion words and motion terms) to identify statements as false or non-false. The logistic regression model classified statements with 76% of the statements, which was a significant improvement over chance (51.4%).

Taken together, these data suggest that very simple classifiers, based on logistic regression, can outperform chance and human judges in classifying deceptive in discourse. Nonetheless, these classification models are far from perfect, making errors on approximately one-third of the decisions. Is it possible that more advanced statistical classifiers, such as the models described in our analysis of status in the IPP Memo Corpus, can improve our ability to classify deception?

Our group has only begun to address this problem. However, others have reported on the performance of more advanced statistical classifiers on deception detection tasks. Zhou, Burgoon, Twitchell, Qin, and Nunamaker (2004) report on a comparison of various statistical classifiers for a deception detection, including discriminant analysis, logistic regression, decision trees and neural networks. The classifiers were applied to deceptive and truthful texts derived from two experiments. All of the classification techniques performed relatively well and in line with the logistic regression models reported by our group, ranging from a classification rate of 55.3% to a high of 66%. Of the techniques, neural networks performed the most consistently across the two datasets, suggesting that this approach might be the most gainful moving forward.

Clearly, however, more work with the kinds of models described above, such as SVMs, is required.

Challenges and conclusions

The SLP paradigm provides a promising way to act as a remote sensor for group dynamics of terrorist organizations and authoritarian regimes. Extensive empirical work has shown that linguistic features correlate robustly with the social features of interest across a wide range of languages and media. Pilot classifiers for *Status* and *Deception* have been successfully piloted for English texts, and we see no principled reason preventing generalization to the other languages. Indeed, since terrorist organizations and authoritarian regimes exist and operate in English and in other languages, it is of great importance to understand the linguistic features of social dynamics for the language in which communications are produced.

A related issue is that of translation. What are the effects on SLP when messages have been translated from the original language to English? We believe there is promise for SLP on translation documents given that our successful SLP analysis of the IPP database for status and cohesion described above relied on messages translated from Arabic.

Despite the evidence provided throughout the manuscript that there are linguistic features related to social dynamics that can be automatically extracted from discourse, it must be borne in mind that the classifiers we have described here are just intended as a proof-of-concept, demonstrating how speech and text can be used to automatically identify social or psychological aspects of speakers and writers. The real next step, given appropriate resources and text collections, should be to look at more subtle aspects of the social structure of groups of interest, be they (suspected) terrorist cells, or units, factions and political institutions operating as part of an authoritarian regime.

Acknowledgments

Preparation of this manuscript was aided by funding from the Department of Defense (H9C104-07-C0019), the Army Research Institute (W91WAW-07-C0029), the Department of Homeland Security/START and the National Science Foundation (NSF0904822).

Notes on contributors

Jeffrey T. Hancock is an Associate Professor at Cornell University. Research interests are in computer-mediated communication and cognitive and social dynamics, including discourse, deception and trust, self-presentation and identity, and social cognition.

David I. Beaver conducts research on formal semantics and pragmatics of natural language. His major interests are presupposition and intonational meaning, and work on anaphora resolution, temporal connectives and Optimality Theoretic Semantics.

Cindy K. Chung is a Post Doctoral Fellow at the University of Texas, Austin. Her research is on how word use reflects personality, psychological states and social groups. One line of research focuses on the use of words that people cannot readily manipulate. A second line of research focuses on extracting statistical patterns of word use in order to track topics over time, across cultures and in multiple languages.

Joey Frazee is a doctoral student at the University of Texas, Austin. His research interests are in computational linguistics, discourse and co-reference, mathematical linguistics, and semantics.

James W. Pennebaker is a Professor at the University of Texas, Austin. His research explores the links between traumatic experiences, expressive writing, natural language use and physical and mental health. His more recent research focuses on the nature of language and range of psychological dynamics in the real world, including emotion, deception and social status.

Art Graesser is a Professor at the University of Memphis. His research interests are in cognitive science, discourse processing and the learning sciences. More specific interests include knowledge representation, question asking and answering, tutoring, text comprehension, inference generation, conversation, reading, education, memory, expert systems, artificial intelligence and human–computer interaction.

Zhiqiang Cai is a Professor at Purdue University. His main research interests are numerical analysis, applied mathematics and computational mechanics

References

Beaver, D. (1997). Presupposition. In J. van Bentham & A. ter Meulen (Eds.), *The handbook of logic and language* (pp. 939–1008). Amersterdam: Elsevier.

Beaver, D. (2001). *Presupposition and assertion in dynamic semantics.* Stanford, CA: CSLI.

Beaver, D., & Zeevat, H. (2007). Accomodation. In G. Ramchand & C. Reiss (Eds.), *The Oxford handbook of linguistic interfaces* (pp. 533–538). Oxford: Oxford University Press.

Berger, A.L., Della Pietra, S.A., & Della Pietra, V.J. (1996). A maximum entropy approach to natural language processing. *Computational Linguistics, 22,* 39–71.

Biber, D., Conrad, S., & Reppen, R. (1998). *Corpus linguistics: Investigating language structure and use.* Cambridge: Cambridge University Press.

Blommaert, J. (2005). *Discourse.* Cambridge: Cambridge University Press.

Bond, C.F., & DePaulo, B.M. (2008). Accuracy of deception judgments. *Personality and Social Psychology Review, 10,* 214–234.

Bond, C.F., Omar, A., Mahmoud, A., & Bonser, R.N. (1990). Lie detection across cultures. *Journal of Nonverbal Behavior, 14,* 189–204.

Bond, G.D., & Lee, A.Y. (2005). Language of lies in prison: Linguistic classification of prisoners' truthful and deceptive natural language. *Applied Cognitive Psychology, 19,* 313–329.

Buller, D.B., & Burgoon, J.K. (1996). Interpersonal deception theory. *Communication Theory, 6,* 203–242.

Chung, C.K., & Pennebaker, J.W. (2007). The psychological functions of function words. In K. Fiedler (Ed.), *Social communication* (pp. 343–359). New York: Psychology Press.

Clark, H.H. (1996). *Using language.* Cambridge: Cambridge University Press.

D'Andrade, R.G., & Wish, M. (1985). Speech act theory in quantitative research on interpersonal behavior. *Discourse Processes, 8,* 229–259.

DePaulo, B.M., Lindsay, J.J., Malone, B.E., Muhlenbruck, L., Charlton, K., & Cooper, H. (2003). Cues to deception. *Psychological Bulletin, 129,* 74–118.

Duran, N.D., McCarthy, P.M., Hall, C., & McNamara, D.S. (in press). The linguistic correlates of conversational deception: Comparing natural language processing technologies. *Applied Psycholinguistics.*

Ekman, P. (1985). *Telling lies: Clues to deceit in the marketplace, politics and marriage.* New York: Norton.

Gonzales, A.L., Hancock, J.T., & Pennebaker, J.W. (in press). Language indicators of social dynamics in small groups. *Communications Research.*

Graesser, A.C., Gernsbacher, M.A., & Goldman, S. (2003). Introduction to the handbook of discourse processes. In A. C. Graesser, M. A. Gernsbacher, & S. Goldman (Eds.), *Handbook of discourse processes* (pp. 1–23). Mahwah, NJ: Lawrence Erlbaum.

Graesser, A.C., Han, L., Jeon, M., Myers, J., Kaltner, J., Cai, Z., McCarthy, P., Shala, L., Louwerse, M., Hu, X., Rus, V., McNamara, D., Hancock, J., Chung, C., & Pennebaker, J. (2009). *Cohesion and classification of speech acts in Arabic discourse.* Paper presented at the Society for Text and Discourse, Rotterdam.

Graesser, A.C., McNamara, D.S., Louwerse, M.M., & Cai, Z. (2004). Coh-Metrix: Analysis of text on cohesion and language. *Behavioral Research Methods, Instruments, and Computers, 36,* 193–202.

Halliday, M. (1967). *Intonation and grammar in British English.* The Hague: Mouton,.
Halliday, M., & Hasan, R. (1976). *Cohesion in English.* London: Longman.
Hancock, J.T., Bazarova, N.N., Markowitz, D. (submitted). A linguistic analysis of Bush administration statements on Iraq. *Political Psychology.*
Hancock, J.T., Curry, L., Goorha, S., & Woodworth, M.T. (2008). On lying and being lied to: A linguistic analysis of deception. *Discourse Processes, 45,* 1–23.
Haviland, S.E., & Clark, H.H. (1974). What's new? Acquiring new information as a process in comprehension. *Journal of Verbal Learning and Verbal Behavior, 13,* 515–521.
Hempelmann, C.F., Dufty, D., McCarthy, P., Graesser, A.C., Cai, Z., & McNamara, D.S. (2005). Using LSA to automatically identify givenness and newness of noun-phrases in written discourse. In B. Bara (Ed.), *Proceedings of the 27th Annual Meeting of the Cognitive Science Society* (pp. 941–946). Mahwah, NJ: Erlbaum.
Hu, X., Cai, Z., Louwerse, M., Olney, A., Penumatsa, P., Graesser, A.C., & the Tutoring Research Group (2003). A revised algorithm for latent semantic analysis. *Proceedings of the 2003 International Joint Conference on Artificial Intelligence* (pp. 1489–1491). San Francisco, CA: Morgan Kaufmann.
Joachims, T. (1998). Text categorization with support vector machines: Learning with many relevant features. In *Proceedings of the European Conference on Machine Learning (ECML),* pp. 137–142.
Johnson, M.K., & Raye, C.L. (1981). Reality monitoring. *Psychological Review, 88,* 67–85.
Jurafsky, D., & Martin, J.H. (2008). *Speech and language processing: An introduction to natural language processing, computational linguistics, and speech recognition.* Upper Saddle River, NJ: Prentice-Hall.
Kacewicz, E., Pennebaker, J.W., Davis, D., Jeon, M., & Graesser, A.C. (submitted). Pronoun use reflects standings in social heirarchies.
Keltner, D., Gruenfeld, D.H., & Anderson, C. (2003). Power, approach, and inhibition. *Psychological Review, 110,* 265–284.
Kintsch, W. (1988). The role of knowledge in discourse comprehension: A construction–integration model. *Psychological Review, 95,* 163–182. AQ5
Köhnken, G. (1996). Social psychology and the law. In G.R. Semin & K. Fiedler (Eds.), *Applied social psychology* (pp. 257–282). London: Sage.
Landauer, T., McNamara, D., Dennis, S., & Kintsch, W. (2007). *Handbook of latent semantic analysis.* Mahwah, NJ: Erlbaum.
Lee, K., Cameron, C.A., Fen, X., Genyue, F., & Board, J. (1997). Chinese and Canadian children's evaluations of lying and truth telling: Similarities and differences in the context of pro and antisocial behavior. *Child Development, 68,* 924–934.
Lenat, D.B. (1995). CYC: A large-scale investment in knowledge infrastructure. *Communications of the ACM, 38,* 33–38.
Louwerse, M.M., & Mitchell, H.H. (2003). Towards a taxonomy of a set of discourse markers in dialog: a theoretical and computational linguistic account. *Discourse Processes, 35,* 199–239.
Masip, J., Sporer, S.L., Garrido, E., & Herrero, C. (2005). The detection of deception with the Reality Monitoring approach: A review of the empirical evidence. *Psychology, Crime, & Law, 11,* 99–122.
McCornack, S.A. (1992). Information manipulation theory. *Communication Monographs, 59,* 1–16.
McFadden, D. (2005) *Analysis of the Weather Underground transcripts.* Unpublished thesis.
McNamara, D.S., Louwerse, M.M., McCarthy, P.M., & Graesser, A.C. (in press). Coh-Metrix: Capturing linguistic features of cohesion. *Discourse Processes.*
Mehl, M.R., & Pennebaker, J.W. (2003). The social dynamics of a cultural upheaval: Social interactions surrounding September 11, 2001. *Psychological Science, 14,* 579–585.
Mullen, B., Chapman, J.G., & Peaugh, S. (2001). Focus of attention in groups: A self-attention perspective. *The Journal of Social Psychology, 129,* 807–817.
Newman, M.L., Groom, C.J., Handleman, L.D., & Pennebaker, J.W. (in press). Sex differences in language use: An analysis of text samples from 70 studies. *Discourse Processes.*
Newman, M.L., Pennebaker, J.W., Berry, D.S., & Richards, J.M. (2003). Lying words: Predicting deception from linguistic style. *Personality and Social Psychology Bulletin, 29,* 665–675.

Niederhoffer, K.G., & Pennebaker, J.W. (2002). Linguistic style matching in social interaction. *Journal of Language and Social Psychology, 21,* 337–360.

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

Pennebaker, J.W., & Chung, C.K. (2008). Computerized text analysis of al-Qaeda statements. In K. Krippendorff & M. Bock (Eds.), *A content analysis reader* (pp. 453–466). Thousand Oaks, CA: Sage.

AQ7 Pennebaker, J.W., Mayne, T.J., & Francis, M.E. (1997). Linguistic predictors of adaptive bereavement. *Journal of Personality and Social Psychology, 72,* 863–871.

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.

Pennebaker, J.W., & Stone, L.D. (2003). Words of wisdom: Language use over the life span. *Journal of Personality and Social Psychology, 85,* 291–301.

Pickering, M.J., & Garrod, S. (2004). Toward a mechanistic psychology of dialogue. *Brain and Behavioral Sciences, 27,* 169–190.

Prince, E.F. (1981). Toward a taxonomy of given–new information. In P. Cole (Ed.), *Radical pragmatics* (pp. 223–255). New York: Academic Press.

Rochon, E., Saffran, E.M., Berndt, R.S., & Schwartz, M.F. (2000). Quantitative analysis of aphasic sentence production: Further development and new data. *Brain and Language, 72,* 193–218.

Rude, S.S., Gortner, E.M., & Pennebaker, J.W. (2004). Language use of depressed and depression-vulnerable college students. *Cognition & Emotion, 18,* 1121–1133.

Sacks, H. (1995). *Lectures on conversation.* Oxford: Blackwell.

Sacks, H., Schegloff, E.A., & Jefferson, G. (1974). A simplest systematics for the organization of turn-taking for conversation. *Language, 50,* 696–735.

Schiffrin, D., Tannen, D., & Hamilton, H.E. (2001). *Handbook of discourse analysis.* Oxford: Blackwell.

Sexton, J.B., & Helmreich, R.L. (2000). Analyzing cockpit communications: The links between language, performance, error, and workload. *Performance in Extreme Environments, 5,* 63–68.

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.

Stalnaker, R. (1974). Pragmatic presuppositions. In M. Munitz & P. Unger (Eds.), *Semantics and philosophy* (pp.197–214). New York: New York University Press.

AQ8 Steyvers, M., & Griffiths, T.L. (2007). Probabilistic Topic Models. In T. Landauer, D. McNamara, S. Dennis, & W. Kintsch (Eds.), *Handbook of latent semantic analysis.* Mahwah, NJ: Lawrence Erlbaum.

Stirman, S.W., & Pennebaker, J.W. (2001). Word use in the poetry of suicidal and non-suicidal poets. *Psychosomatic Medicine, 63,* 517–522.

Tausczik, Y.R., & Pennebaker, J.W. (in press). The psychological meaning of words: LIWC and computerized text analysis methods. *Journal of Language and Social Psychology.*

Vrij, A. (2005). Criteria-based content analysis: A qualitative review of the first 37 studies. *Psychology, Public Policy, and Law, 11,* 3–41.

Vrij, A. (2008). *Detecting lies and deceit: pitfalls and opportunities.* Chichester: Wiley.

Vrij, A., Edward, K., Roberts, K.P., & Bull, R. (2000). Detecting deceit via analysis of verbal and nonverbal behavior. *Journal of Nonverbal Behavior, 24,* 239–263.

Weintraub, W. (1989). *Verbal behavior in everyday life.* New York: Springer.

Woods, K., Pease, M.R., Stout, M.E., Murray, W., & Lacey, J.G. (2006). *The Iraqi Perspectives Report: Sadaam's senior leadership on Operation Iraqi Freedom from the Official U.S. Joint Forces Command Report.* Annapolis, MD: Naval Institute Press.

Zhou, L., J.K. Burgoon, D. Twitchell, T. Qin, and J.F. Nunamaker (2004), A comparison of classification methods for predicting deception in computer-mediated communication. *Journal of Management Information Systems, 20,* 139–165.

Zhou, L., & S. Lutterbie (2005). Deception across cultures: Bottom-up and top-down approaches. In P. Kantor et al. (Eds.), *IEEE international conference on intelligence and security informatics (IEEE ISI-2005),* Atlanta, GA, 19–20 May 2005, LNCS 3495, pp. 465–470. Springer: Berlin.

Zuckerman, M., DePaulo, B.M., & Rosenthal, R. (1981). Verbal and nonverbal communication of deception. In L. Berkowitz (Ed.), *Advances in experimental social psychology. Vol. 14.* New York: Academic Press.

Zwaan, R.A., & Radvansky, G.A. (1998). Situation models in language comprehension and memory. *Psychological Bulletin, 123,* 162–185.