A corpus-based analysis of potential linguistic indicators of corporate deception in tobacco industry documents

Cati Brown-Johnson, Donald Rubin

Abstract Introduction: To more fully understand the impact of specific language attributes on deception in corporate communication, we used a stratified random sample of tobacco industry documents to test a small number of individual potential automated linguistic indicators of corporate deception: cognitive-emotional verbs, allness and superlative terms, nonbinding verbs, and group mentality. Methods: Texts from the Tobacco Documents Corpus were categorized by audience addressed (industry internal/external) and company of origin, and compared for incidence of the deceptive language indicators. Results: Cognitive-emotional verbs were strongly associated with likely deception and most prevalent for external audiences and documents produced by the Tobacco Institute, an industry front group. Cognitive-emotional verbs include *believe, think, seem, feel* and *realize,* in opposition to action verbs (e.g. *throw*). Discussion: Linguistic measurements of cognitive-emotional verbs may help pinpoint deceptive and misleading corporate communication. Verb choice and the use of different classes of verbs may reflect corporate linguistic deception.

A corpus-based analysis of potential linguistic indicators of corporate deception in tobacco industry documents

Running head: Linguistic Indicators of Deception

Cati G Brown-Johnson, PhD
Stanford Prevention Research Center, Stanford University, Stanford, California, USA

Donald L Rubin, PhD
Center for Health and Risk Communication, University of Georgia, Athens, Georgia, USA

Corresponding Author: Cati Brown-Johnson, Stanford Prevention Research Center, 1265 Welch Rd, Palo Alto, CA 94305; Email:catibj@stanford.edu; Work telephone: 650-497-4397.

Abstract

Introduction: To more fully understand the impact of specific language attributes on deception in corporate communication, we used a stratified random sample of tobacco industry documents to test a small number of individual potential automated linguistic indicators of corporate deception: cognitive-emotional verbs, allness and superlative terms, nonbinding verbs, and group mentality.

Methods: Texts from the Tobacco Documents Corpus were categorized by audience addressed (industry internal/external) and company of origin, and compared for incidence of the deceptive language indicators.

Results: Cognitive-emotional verbs were strongly associated with likely deception and most prevalent for external audiences and documents produced by the Tobacco Institute, an industry front group. Cognitive-emotional verbs include *believe, think, seem, feel* and *realize,* in opposition to action verbs (e.g. *throw*).

Discussion: Linguistic measurements of cognitive-emotional verbs may help pinpoint deceptive and misleading corporate communication. Verb choice and the use of different classes of verbs may reflect corporate linguistic deception.

INTRODUCTION

Evidence from tobacco industry documents warrants the conclusion that the tobacco industry has used deception and unethical business and marketing practices to perpetrate a variety of activities that are illegal and endanger public health, including marketing to youth (Cummings et al. 2002a), marketing "light" cigarettes as healthful (Pollay & Dewhirst 2002), and controlling international tobacco markets through black-market trade (Lee et al. 2004). Some recent studies have linked tobacco industry practices that negatively affect public health to specific language strategies. Low-income women have been targeted by tobacco companies with more easily-comprehensible language (Brown 2012). Tobacco companies have engineered associations between menthol and health with "health reassurance" language such as the following tagline, "Breath easy, clean smoke" (Anderson 2011). Language choice was an important aspect of successfully launching "light" and "ultra-light" cigarettes (Pollay & Dewhirst 2002); words used to market cigarettes in this genre ("light," "low," and "mild") have been deemed so misleading that the Family Smoking and Tobacco Control Act creating oversight of tobacco by the US Federal Drug Administration banned their use (Sec. 911 of the Federal Food, Drug, and Cosmetic Act).

Corporations can be expected to try to cover up their deceptive practices. Their attempts to hide unsavory business practices will at some point most certainly include deceptive or misleading communication with people or organizations outside their organization. Previous work in corpus linguistics has shown that language variables can be crafted to uncover subtle stylistic differences between texts, for example in attributing unknown authorship or automatically determining document genre (Stamatatos et al. 2000). Furthermore, linguistic indicators have been shown to pick up on

"leaked" language cues of deception, but results of specific indicators vary; for example, modal verbs have been shown to predict deception (Burgoon et al.) and hedging verbs have been associated with fraudulent financial statements (Humpherys 2009).

Automated assessment of deception has been a holy grail within numerous fields - criminal justice, accounting, business ethics - and great strides have been made in the last fifteen years towards automated linguistic lie detectors. The computationally straightforward Linguistic Inquiry Word Count computation program (Pennebaker & King 1999) was subsequently found to predict truthfulness or deception in texts almost 70% of the time (Newman et al. 2003). More recently, linguistic assessments of investment house telephone calls and emails successfully identified deceptive executives at a rate greater than chance; deceptive executives used more extreme positive emotions, fewer anxiety words, and talked less about shareholder value (Larcker & Zakolyukina 2012). Overall, a meta-analysis of computer-assisted studies of deceptive language showed small but statistically significant effect sizes (Hauch et al. 2012). Individual deceivers, then, whether petty thieves or CEOs, can likely be identified by their language choices.

To test whether potential individual linguistic indicators of deception might be associated with deceptive corporate documents, we assessed the instance of four linguistic indicators (allness and superlative terms, group mentality, nonbinding verbs, and cognitive-emotional verbs) between likely nondeceptive and likely deceptive text groups derived from publicly available tobacco industry documents.

Linguistic indicators of deception

Deception has been defined as a "communicative act that is intended to induce in the addressee a particular belief, by manipulating the truth and falsity of information" (Galasiânski 2000:20). The

quintessential type of deception is outright lying, in which a speaker actively represents the opposite of the truth (e.g., US President Richard Nixon's famous statement, "I am not a crook."), but other forms of deception include evasion (not providing answers to a question), deception by omission (withholding or concealing relevant information), and deception by obfuscation (overuse of jargon to restrict comprehension) (Buller et al. 1994; Galasiânski 2000). Any of these manifestations of deception may be present in corporate contexts. Little research has been conducted on corporate deception language, however. Therefore, a review of interpersonal deception literature (for details of that review see Author1, 2006) was undertaken. This review identified four proposed indicators of corporate fraud and deception: (1) allness and superlative terms, (2) group mentality, (3) nonbinding verbs, and (4) cognitive-emotional verbs. These initial indicators test whether this type of approach would be worth pursuing further within the corporate document repositories such as the Legacy Tobacco Documents Library (LTDL), but they do not constitute an exhaustive set of deception indicators.

Allness and superlative terms

Deception research asserts that superlative terms and superlatives often mark efforts to deceive via exaggeration. *Always*, *never*, *nobody*, *worst* and *everybody* are mentioned as words that may indicate deception (Lebaron 1996; Vrij 2000) Superlative terms mark communication that has an absolutist position aligning with tobacco industry perception of having been 'singled out' for legislative and judicial punitive attention (1993).

Group mentality

According to guidelines used by the United States Federal Bureau of Investigation agents in profiling perpetrators of crimes, deceivers often speak in ways that try to enlarge the group they are in

(Hess 1997). They use more group inclusion words and fewer self-references, possibly in order to distribute responsibility and distance or cushion themselves from individually assignable blame.

Linguistically, an enlargement of the group that you are in is most readily expressed by the use of first person plural pronouns: *we, us, our*. Of course many referentially appropriate uses of first person plural pronouns are present in everyday communication and speech, and may be commonly used in corporate communication. However, a non-specific *we* can also diffuse responsibility for negative outcomes and be associated with an identity of victimization (Zhou et al. 2004), as is evident in some tobacco industry documents. For example, in a statement responding to a proposed cigarette advertising ban, the President of Wholesale Tobacco Distributors of New York makes use of *we* to enlarge his group's interest to include all people interested in personal freedom: "*We* are particularly sensitive about restrictions on *our* freedom and liberties" (Bloomrosen 1969).

Nonbinding verbs

Nonbinding verbs refers to verb forms that are nonbinding from a legal sense. These verbs are in opposition to *commissives* (Searle 1969), which by contrast are a form of performative speech-act that *commits* or promises future action. Commissive verbs such as *will* and *shall* create an environment of obligation (Searle 1969). Corporations involved in deception may avoid obligation and obscure their future intentions so that they cannot be held responsible. Thus, language that does not promise and is nonbinding from a legal sense is preferred.

Cognitive-emotional verbs

Cognitive and emotional language prioritizes *emotional* and *cognitive state* information as opposed to *action-oriented* information. A heightened use of cognitive-emotional verbs allow the tobacco industry to assert a position with respect to an issue from an emotive or opinion-centered

instead of factual point of view. The use of *believe* and other cognitive verbs such as *consider* (Lebaron 1996) steers tobacco companies away from making any statements about actions or behaviors that could subsequently be challenged (Derry & Brown 2005). For example, RJR invoked a science-denier position in a advertisement entitled "Of cigarettes and science": "We *believe* in science… but we do not *believe* that there should be…a different set of experiments involving cigarettes" (Mitchell & Womble 1983).

Research aims and hypotheses

The goal of this study is to track how four proposed linguistic indicators of deception behave in the corporate texts of the Tobacco Documents Corpus (TDC) (Kretzschmar et al. 2004), a stratified sample subset of the previously secret internal tobacco industry documents housed through the LTDL (Kretzschmar et al. 2004). We reviewed deception studies from diverse fields (discourse analysis, organizational studies, corpus linguistics, experimental psychology) to create the four proposed linguistic indicators of corporate deception, which we test on tobacco industry documents to assess in which document sets these indicators occur.

Using these theoretically based indicators of deception, we aim to determine whether there are differences between external (public) and internal language used by the tobacco industry and to explore any differences between companies.

H1: Internal versus external audience documents. Compared with documents addressed to audiences internal to tobacco companies, documents addressed to external audiences will exhibit higher relative frequencies of linguistic indicators of deception.

H1 is based on a body of evidence showing that the industry engaged in communication campaigns deliberately designed to deceive the public (Cataldo & Malone 2008; Cummings et al.

2002), and that internal information was frequently at odds with information released to the public (Bero 2005).

H2a: Linguistic indicators of deception should vary significantly by industry source;

H2b: Documents from the Tobacco Institute and Center for Tobacco Research are predicted to have higher relative frequencies of indicators compared to other companies.

For source, we predicted that material originating from two organizations, the Tobacco Institute (TI) and the Council for Tobacco Research (CTR), would house more deception in documents compared to the five major corporate defendants who were party to the Master Settlement Agreement: the American Tobacco Company (ATC), Brown and Williamson (BW), Lorillard (LL), Philip Morris (PM), and RJ Reynolds (RJR). In contrast to manufacturing, TI and CTR did not create cigarettes themselves; instead they promoted tobacco industry agendas (TI) and protected tobacco manufacturers from liabilities with respect to research (CTR). Material originating from these two tobacco industry umbrella organizations was intentionally more misinformative and deceitful than documents from the other tobacco industry company sources (Yach & Bettcher 2000). These "front groups set up by industry to block public health policies" were disbanded as part of the Minnesota Settlement due to the pervasive misinformation and deceptive content they consistently espoused (Yach & Bettcher 2000). TI and CTR strategically misled and deceived the public (Campbell & Balbach 2009; Campbell & Balbach 2011) from 1954 until their court-ordered disbanding in 1998.

MATERIALS & METHODS

Indicators of deception

Allness and superlative terms were calculated as the total number of non-adjective, non-adverbial allness words, superlative adjectives and superlative adverbs (e.g., greatest, best, longest, worst, fastest, etc.) divided by the total number of adjectives and adverbs. Allness words include always, never, nobody, forever, everybody and everyone.

Group mentality was calculated as the total of plural first and third person pronouns (us, our, ours, we, them, their, they, theirs) relative to the total number of personal pronouns, possessive pronouns and plural nouns (Hess 1997). The group agency pronouns were denominated by the total number of personal pronouns.

The indicator for nonbinding verbs was operationalized as the number of present and past tense base form verbs over the total base form verbs using the MontyTagger part-of-speech tagging software package (Liu 2004).

The indicator for cognitive-emotional verbs was operationalized as the total cognitive-emotional verbs divided by the total number of base form verbs. Cognitive-emotional verbs included stative and relational verbs: believe, consider, think, doubt, intend, know, appear, become, appreciate, etc. (Kiparsky & Kiparsky 1970; Lebaron 1996; Shuy 2003).

The Tobacco Documents Corpus (TDC)

The Tobacco Document Corpus (TDC) (Kretzschmar et al. 2004) was created as a stratified random sample of the LTDL. The TDC used hand encoding (retyping all documents) to reduce errors that can be created by Optical Character Recognition (OCR) in the electronic text. The general tobacco

industry document files are available electronically thanks to scanning techniques adopted by the LTDL in 2004 that translate hard paper copies into electronic text via OCR (Anderson et al. 2011). Unfortunately, due to image quality, hand written notations, and font differences in the hard copies, OCR output is imprecise and error prone. Accordingly, the TDC was manually keyed in as full text. In addition, as part of this retyping process, staff coded additional metadata to represent potentially important information about texts as well as divisions within a single text. The documents in the TDC contain fully represented text as well as various differentiated metatext including headers, pretext, marginalia, main text, post text, lineouts, etc.

Additional metadata associated with each TDC document of interest for this study included audience (industry internal vs. industry external addressees) and industry source (ATC, BW, LL, PM, RJR, CTR, TI). Industry internal audience was defined as documents addressed to any individual or organization known to have received financial support from any tobacco source; industry external audience encompassed all other documents. We used audience affiliation and source as variables among which the indicators of deception could be compared.

Statistical analyses

Because distributions for all four of the indicators violated assumptions of normality, nonparametric tests were performed. For audience (industry Internal versus External), the Mann-Whitney Rank-Sum test (a nonparametric equivalent of the unpaired t-test -- U is the test statistic) was used. For source, chi square (χ^2) analyses were performed. Posthoc Mann-Whitney Rank-Sum tests were performed between pairs of sources where the overall chi square result was significant in order to determine which individual contrasts between industry sources produced the significant effects.

RESULTS

The means and standard deviations broken down by audience and source for the four indicators of deception are presented in Table 1 and Table 2.

H1 was supported for cognitive-emotional verbs. Nonparametric tests indicated that documents intended for internal audiences had a significantly lower score for this linguistic feature than did documents addressed to external audiences (U=70985.0, Z=-2.829, p=0.005). Results were marginally significant for nonbinding verbs (U=74612.0, Z=-1.932, p=0.053); documents addressed to industry external audiences likewise displayed a higher frequency of these features. Differences between internally and externally directed documents did not emerge on the other two indicators of deception.

Sources differed significantly for cognitive-emotional verbs (χ^2=23.051, df=6, p=0.001). Partially confirming our second hypothesis, pairwise comparisons among the seven industry sources showed that TI documents were associated with greater cognitive-emotional verbs compared to four of the other six sources, including ATC (U=2423.5, z=-4.540, p<0.000), LL (U=4137.5, z=-2.001, p=0.045), PM (U=16425.0, z=-2.636, p=0.008), and RJR (U=11142.0, z=-2.989, p=0.003). BW documents were associated with greater cognitive-emotional verb use compared to RJR (U=14177.0, z=-2.065, p=0.039) and ATC (U=3421.5, z=-3.206, p=0.001). By contrast, ATC documents were associated with lower cognitive-emotional verb use compared not only to BW and TI (as above), but to all other companies (CTR U=1363.0, z=-2.251, p=0.024; LL U=3061.0, z=-2.610, p=0.009; PM U=12592.0, Z=-2.896, p=0.004; RJR U=9366.5, z=-2.193, p=0.028). No other pairwise contrasts were statistically significant for this indicator.

Differences among sources were marginally significant for group mentality (χ^2=12.519, df=6, p=0.051). For the group mentality indicator, differences among source based on pair-wise comparisons

were mixed with respect to our hypothesis. CTR documents were associated with significantly lower group mentality compared to ATC (U=1354.5, z=-2.304, p=0.021), while TI was associated with higher group mentality compared to LL (U=3938.5, z=-2.497, p=0.013), CTR (U=1599.5, z=-2.784, p=0.005), PM (U=16867.5, z=-2.302, p=0.021), and RJR (U=11585.0, z=-2.528, p=0.011). No other pairwise contrasts were statistically significant for this indicator.

No significant differences among industry sources emerged for nonbinding verbs or allness and superlative terms.

DISCUSSION

Cognitive-emotional verbs seemed to be especially active as an indicator of potential deceptiveness in tobacco industry documents. The indicator for cognitive-emotional verbs tracked mental, cognitive, stative or relational verbs (e.g., *believe*, *consider*, etc.) that express a position on a topic as opposed to any action. External documents included significantly more cognitive-emotional verbs compared to internal audience documents. This indicator also varied by industry source. Documents produced by the Tobacco Institute, the tobacco industry front group that was deliberately created to shield manufacturers from negative consequences of their products, revealed higher levels of this indicator compared to four of the other six corporate entities. Thus, relatively high frequencies of cognitive-emotional verbs may be a useful indicator of corporate deception. Lastly, the significance of cognitive-emotional verbs suggests that verbs in particular may be a fruitful site for future linguistic investigations of corporate deception.

The three other proposed indicators did not uniformly follow hypothesized distributions. Within this sample, nonbinding verbs and group mentality were marginally significant (in each case $p<.055$) in differentiating between audience and source document subsets, respectively. Further testing within the

full LTDL is needed to determine whether these marginal results were an artifact of the specific sample examined here, or evidence of the usefulness of the indicators.

Our results revealed some problems with our underlying assumptions of deception. For instance, we had assumed that TI and CTR documents would both manifest deception, but only TI documents (and not CTR documents) included higher amounts of cognitive-emotional verbs and group mentality compared to other sources. Although both organizations were found in previous research to have contributed to public misinformation and corporate fraud (Cummings et al. 2002), our results hint that they accomplished this task in different ways. Cursory reviews of a number of TI and CTR documents included in our sample show that TI took on more of the role of public interface, while CTR operated scientific granting operations. Verification of these roles and any potential impact they may have had on the corporate culture and language should be undertaken using both quantitative indicator-based investigations and indepth qualitative analysis of documents.

Corporate deception is enacted by individuals who are directly influenced by the structure and norm, or "corporate culture," of the organization and corporate deception is often encouraged by organizational leadership (Fleming & Zyglidopoulos 2008). Examples within large organizations demonstrate that bad behavior of 'rogue' employees is at minimum supported by a culture of willful negligence, and often all levels of an organization are implicated in fraudulent practices and their cover up (Zyglidopoulos & Fleming 2008). Indeed, within the tobacco industry companies and organizations, a large body of evidence exposed within previously confidential tobacco industry documents demonstrates that decisions were made in the highest corporate offices to deceive by discouraging public access to information (Cummings 2003; Cummings et al. 2002; Cummings & Pollay 2002; Farren 2004; Friedman 2007). This study demonstrates that although individual and corporate deception do not follow the same pattern on all accounts, there are potentially overlaps between the

two, including the indicator of cognitive-emotional verbs, derived from individual deception studies and shown here to be associated with corporate deception in tobacco documents. This successful test case indicates that what little homogeneity exists within written corporate deception can be used to differentiate between deceptive and truthful texts, at least within a highly organized and cohesive corporate culture such as that of the tobacco industry.

One limitation of this study is the small number of indicators, which may not represent the full breadth of linguistic mechanisms for deception. Additionally, external audience, TI and CTR documents were assumed a priori based on the literature to be deceptive, but obviously not every document of any of these document sets would be deceptive, which could introduce significance noise into our a priori deceptive document sets.

CONCLUSIONS

Taken as a whole, this study shows that one linguistic indicator —cognitive-emotional verbs— may be robust enough to red flag deceptive documents for further examination within the context of the tobacco industry. This conclusion needs to be explored in the larger body of documents from the LTDL. Additionally, since two productive indicators of deception (i.e., cognitive-emotional verbs and nonbinding verbs) revolved around verb choice, further research could consider focusing on verb use in particular in the context of corporate deception. One reason why common verbs may be good candidates to function as indicators of deception is that they are so routine as to be used in an unconscious manner by writers. Indeed, authorship attribution studies have similarly found that patterning in nonsalient function words is often the best indicator of an author's individual "wordprint" (Stamatatos et al. 2000). These findings have important implications for public health, tobacco control and applied linguistics. Public health practicioners and activists, who are often under-funded compared

to corporations, may now have another tool with which to track and quickly assess possible deception in corporate language. With respect to tobacco control, automated linguistic analysis offers an additional route beyond typical keyword searches to conduct surveillance on tobacco industry activities. The present findings remind us that while tobacco companies may try to control their rhetorical strategies, they may have less control of the linguistic features by which they "leak" deceptive and misleading intentions, thus providing another possible mode for identification of documents of interest for tobacco control and public health.

References

1993. Rogers makes the case for kentucky tobacco families, says clinton's tobacco taxes are punitive and unfair. *Lorillard*. Bates: 89735041/89735042. Accessed on 14 Nov 2012. http://legacy.library.ucsf.edu/tid/gue01e00

Anderson SJ. 2011. Marketing of menthol cigarettes and consumer perceptions: A review of tobacco industry documents. *Tobacco Control* 20:ii20-ii28.

Anderson SJ, McCandless PM, Klausner K, Taketa R, and Yerger VB. 2011. Tobacco documents research methodology. *Tobacco Control* 20:Ii8-Ii11.

Bero LA. 2005. Tobacco industry manipulation of research. *Public Health Reports* 120:200-208.

Bloomrosen M. 1969. Statement of milton bloomrosen, president , wholesale tobacco distributors of new york, inc. In opposition to proposed bill no. 987 inc. No. 834 banning cigarette advertising in public places and on public conveyances. *Philip Morris*. Bates: 1005098873/1005098879. Accessed on 14 Nov 2012. http://legacy.library.ucsf.edu/tid/bev38e00

Brown CG. 2012. Tobacco industry marketing to low income women. American Psychological Association - Disparities Meeting. Washington, DC.

Buller DB, Burgoon JK, White CH, and Ebesu AS. 1994. Interpersonal deception .7. Behavioral profiles of falsification, equivocation, and concealment. *Journal of Language and Social Psychology* 13:366-395.

Burgoon JK, Hamel L, and Tiantian Q. Predicting veracity from linguistic indicators.

Campbell RB, and Balbach ED. 2009. Building alliances in unlikely places: Progressive allies and the tobacco institute's coalition strategy on cigarette excise taxes. *American Journal of Public Health* 99:1188-1196.

Campbell RB, and Balbach ED. 2011. Manufacturing credibility: The national energy management institute and the tobacco institute's strategy for indoor air quality. *American Journal of Public Health* 101:497-503.

Cataldo JK, and Malone RE. 2008. False promises: The tobacco industry, "low tar" cigarettes, and older smokers. *Journal of the American Geriatrics Society* 56:1716-1723.

Cummings KM. 2003. A promise is a promise. *Tobacco Control* 12:117-118.

Cummings KM, Morley CP, Horan JK, Steger C, and Leavell N-R. 2002a. Marketing to america's youth: Evidence from corporate documents. *Tobacco Control* 11 Suppl 1:I5-17.

Cummings KM, Morley CP, and Hyland A. 2002c. Failed promises of the cigarette industry and its effect on consumer misperceptions about the health risks of smoking. *Tobacco Control* 11:110i-117.

Cummings KM, and Pollay RW. 2002. Exposing mr butts' tricks of the trade. Introduction. *Tobacco Control* 11 Suppl 1:I1-4.

Derry R, and Brown C. 2005. Strategic trust-building: The use of linguistic devices to induce trust. IABS: International Association for Business and Society Annual Conference. Sonoma Valley, California.

Farren C. 2004. Gasp: Picking off the pack of lies. *Tobacco Control* 13:100-101.

Fleming P, and Zyglidopoulos SC. 2008. The escalation of deception in organizations. *Journal of Business Ethics* 81:837-850.

Friedman LC. 2007. Philip morris's website and television commercials use new language to mislead the public into believing it has changed its stance on smoking and disease. *Tobacco Control* 16:e9.

Galasiânski D. 2000. *The language of deception : A discourse analytical study*. Thousand Oaks, California: Sage Publications.

Hauch V, Masip J, Blandón-Gitlin I, and Sporer SL. 2012. Linguistic cues to deception assessed by computer programs: A meta-analysis. Proceedings of the workshop on computational approaches to deception detection: Association for Computational Linguistics. p 1-4.

Hess JE. 1997. *Interviewing and interrogation for law enforcement*. Cincinnati, OH: Anderson Pub. Co.

Humpherys SL. 2009. Discriminating fradulent financial statements by identifying linguistic hedging. AMCIS: American Conference on Information Systems.

Kiparsky P, and Kiparsky C. 1970. Fact. In: Bierwisch EM, and Heidolph K, eds. *Progress in linguistics*. The Hague: Mouton, 143-173.

345 Kretzschmar WA, Jr., Darwin C, Brown C, Rubin DL, and Biber D. 2004. Looking for the smoking gun: Principled
346 sampling in creating the tobacco industry documents corpus. *Journal of English Linguistics* 32:31-47.
347 Larcker DF, and Zakolyukina AA. 2012. Detecting deceptive discussions in conference calls. *Journal of*
348 *Accounting Research* 50:495-540.
349 Lebaron CD. 1996. Looking for verbal deception in clarence thomas's testimony. In: Ragan SL, Bystrom D, and
350 Kaid LL, eds. *The lynching of language : Gender, politics, and power in the hill-thomas hearings*. Urbana:
351 University of Illinois Press, 113-132.
352 Lee K, Gilmore AB, and Collin J. 2004. Breaking and re-entering: British american tobacco in china 1979-2000.
353 *Tobacco Control* 13 Suppl 2:ii88-95.
354 Liu H. 2004. Monylingua: An end-to-end natural language processor with common sense.
355 Mitchell WGC, and Womble C. 1983. We're not just raising questions. We're seeking answers. *RJ Reynolds*.
356 Bates: 503115487-503115494. Accessed on 11 Apr 2013. http://legacy.library.ucsf.edu/tid/ppg87h00
357 Newman ML, Pennebaker JW, Berry DS, and Richards JM. 2003. Lying words: Predicting deception from
358 linguistic styles. *Personality & Social Psychology Bulletin* 29:665-675.
359 Pennebaker JW, and King LA. 1999. Linguistic styles: Language use as an individual difference. *Journal of*
360 *Personality & Social Psychology* 77:1296-1312.
361 Pollay RW, and Dewhirst T. 2002. The dark side of marketing seemingly "light" cigarettes: Successful images
362 and failed fact. *Tobacco Control* 11 Suppl 1:I18-31.
363 Searle J. 1969. *Speech acts: An essay in the philosophy of language*. Cambridge, England: Cambridge University.
364 Shuy RW. 2003. Tobaccospeak: Image repair as a variety of american english. American Dialect Society. Atlanta,
365 Georgia.
366 Stamatatos E, Kokkinakis G, and Fakotakis N. 2000. Automatic text categorization in terms of genre and author.
367 *Computational Linguistics* 26:471-495.

Vrij A. 2000. *Detecting lies and deceit : The psychology of lying and the implications for professional practice*. Chichester, England ; New York: John Wiley.

Yach D, and Bettcher D. 2000. Globalisation of tobacco industry influence and new global responses. *Tobacco Control* 9:206-216.

Zhou L, Burgoon JK, Nunamaker JF, and Twitchell D. 2004. Automating linguistics-based cues for detecting deception in text-based asynchronous computer-mediated communication. *Group Decision and Negotiation* 13:81-106.

Zyglidopoulos SC, and Fleming PJ. 2008. Ethical distance in corrupt firms: How do innocent bystanders become guilty perpetrators? *Journal of Business Ethics* 78:265-274.

Table 1 (on next page)

Means by audience of the four normalized potential indicators of deception (standard deviation in parentheses)

*p<.05; †p<.055

Table 1. Means by audience of the four normalized potential indicators of deception (standard deviation in parentheses)

Linguistic Indicator	***Internal Audience***	***External Audience***
Allness and superlative terms	0.0239 (0.0413)	0.0236 (0.0321)
Group mentality	0.2813 (0.239)	0.2523 (0.212)
Nonbinding verbs†	0.9506 (0.0815)	0.9544 (0.0686)
Cognitive-emotional verbs*	0.1299 (0.094)	0.152 (0.1028)

*p<.05; † p<.055

Table 2(on next page)

Mean by source of the five potential indicators of deception (standard deviation in parentheses)

**p<0.001; †p<.055

Table 2. Mean by source of the five potential indicators of deception (standard deviation in parentheses)

Linguistic Indicator	*Lorillard*	*Tobacco Institute*	*Center for Tobacco Research*	*Philip Morris*	*R.J. Reynolds*	*Brown & Williamson*	*American Tobacco Company*
Allness and superlative terms	0.0273 (0.05469)	0.0209 (0.02155)	0.0164 (0.04259)	0.0255 (0.04664)	0.0202 (0.02702)	0.0274 (0.04467)	0.0267 (0.04366)
Group mentality†	0.26 (0.241)	0.32 (0.209)	0.22 (0.206)	0.27 (0.237)	0.27 (0.236)	0.27 (0.246)	0.32 (0.241)
Nonbinding verbs	0.9329 (0.1103)	0.9583 (0.05493)	0.9352 (0.08795)	0.9475 (0.07611)	0.9342 (0.08243)	0.9373 (0.08206)	0.9537 (0.05418)
Cognitive-emotional verbs*	0.1315 (0.08374)	0.1544 (0.09366)	0.1589 (0.12976)	0.1322 (0.09704)	0.1254 (0.08935)	0.1482 (0.10411)	0.1061 (0.08859)

**p<0.001; † p<.055