

Talking Heads:  
Measuring Elite Personality Using Speech

Jonathan D. Klingler, Gary E. Hollibaugh, Jr. and Adam J. Ramey



# Talking Heads: Measuring Elite Personality Using Speech \*

Adam J. Ramey<sup>†</sup>  
Jonathan D. Klingler<sup>‡</sup>  
Gary E. Hollibaugh, Jr.<sup>§</sup>

July 8, 2015

## Abstract

Political scientists have long considered ideology, partisanship, and constituency in determining how members of the United States Congress make decisions. Meanwhile, psychologists have held that personality traits play central roles in decision-making. In this paper, we apply recent advances in machine learning (Mairesse et al. 2007) to measure Congressman personality traits using floor speeches from 1996–2014. We show that these estimates are robust to concerns about strategic behavior and generally conform with findings in the behavior literature linking ideology with the Big Five (e.g. Gerber et al. 2010). We also provide two examples of the utility of our method, one examining the impact of personality on elite survey non-response and the other showing how the Conscientiousness of members of Congress affects the contents of bill proposals.

---

\* Authors are listed in reverse alphabetical order. All contributed equally to the paper. Support through ANR - Labex IAST is gratefully acknowledged. We thank John Jost, Jonathan Nagler, Richard Bonneau, Drew Dimmery, Larry Rothenberg, Dave Primo, Maya Sen, Matt Blackwell, Michael Peress, Karine van der Straeten, Elena Panova, Andy Harris, Pablo Hernandez-Lagos, Arthur Spirling, Molly Roberts, Michael Gill, Ken Benoit, Slava Mihkaylov, Jo Silvester, Jeff Mondak, Conor Dowling, and participants at the 5th Annual Text as Data Conference and the Rooney Center for the Study of American Democracy for comments and feedback. All remaining errors are our own.

<sup>†</sup>Corresponding author. Assistant Professor of Politics, New York University Abu Dhabi, PO Box 129188, Abu Dhabi, United Arab Emirates. Email: adam.ramey@nyu.edu

<sup>‡</sup>Postdoctoral Fellow, Institute for Advanced Study in Toulouse, 21 allée de Brienne, Toulouse 31000, France. Email: jonathan.klingler@iast.fr

<sup>§</sup>Assistant Professor, Department of Political Science, University of Notre Dame, Notre Dame, IN 46556. Email: gholliba@nd.edu

# 1 Motivation

Recently, political scientists have taken notice of the role of personality in political behavior. The Five Factor, or “Big Five,” model has been applied widely in the behavior literature to the study of participation, ideology, and vote choice (e.g. Caprara et al. 2006, Mondak et al. 2010, Gerber et al. 2010, Gerber et al. 2011). This model decomposes individual personality into five traits—Openness to Experience, Conscientiousness, Extraversion, Agreeableness, and Emotional Stability. Though growing rapidly, this literature has eschewed a focus on elite behavior.

A key reason is measurement, as measuring elite personality is more problematic than it may seem at first, since most political science applications using these metrics involve surveys of voters (Caprara, Barbaranelli & Zimbardo 2002, Gerber et al. 2011). While useful, these studies tell us little about the traits of elites. One solution might be to survey legislators; unfortunately, this approach would prevent us from being able to study the deceased and would be difficult to perform for retired legislators. Moreover, even if we restrict ourselves to the contemporary legislative period, most legislators would be unwilling to participate, and any responses would likely be subject to selection bias and strategic decisions.

We are not the first to recognize this limitation. To our knowledge, only one study has attempted to apply survey-based inventories with legislators (Dietrich et al. 2012).<sup>1</sup> Response rates were low—ranging from 17% to 26%. These rates, coupled with the inability to get estimates for past time periods, has prevented scholars of behavior in institutions from incorporating personality into their analyses.

In this paper, we present a method for measuring elite personality without surveys. Drawing on recent advances in machine learning, we present an approach to measuring legislator personality using floor speeches. Using this method, we estimate Big Five personality traits for all incumbent members of the United States House of Representatives from 1996–2014. We show

---

<sup>1</sup>Though see Silvester, Wyatt & Randall (2014).

that these estimates parallel findings in the behavior linking personality with ideology (Gerber et al. 2010, Jost, Nosek & Gosling 2008) and that our estimates are robust to authorship and speechwriting effects. We then use our estimates for two empirical applications. First, we investigate how our text-based estimates of personality can shed light on the propensity of legislators to respond to surveys. Second, we examine the relationship between our estimates and the types of bills that legislators propose. The results of these applications have important implications for the study of elite behavior; namely, we show that speech-based personality has strong effects on the probability of survey response and that legislators with certain personality profiles are more likely to propose substantive legislation than others. The former suggests that studies employing survey-based measures of legislator personality might be fraught by personality-induced selection bias. The latter demonstrates that personality has an important effect on legislator behavior, even after accounting for traditional ideological and office-based constraints on legislative decision-making.

## **2 Measuring Personality: From Speeches to Scores**

Given the issues with survey-based measurement of personality identified above, studying the role of personality in legislator behavior—or, more generally, elite behavior—requires a method different from direct surveys. We therefore draw on a recent literature in machine learning that uses traditional psychometric personality inventories in conjunction with written texts and auditory transcriptions to train predictive models for personality (Mairesse et al. 2007, Mairesse & Walker 2008, Schuller et al. 2013). Once the known personalities of a subset of authors are calibrated with their linguistic usage, “virgin texts” can be assessed for personality content, even in the absence of responses to traditional inventories.

Using Pennebaker & King’s (1999) corpus of over 1.9 million words from laboratory experiments and Mehl, Gosling & Pennebaker’s (2006) corpus of almost 100,000 words from recorded

conversations, Mairesse et al. (2007) train several machine learning models to predict personality from linguistic usage. Words are not treated as independent entities, but are first processed according to one of more than 70 categories developed by Pennebaker, Francis & Booth (2001) in the 2001 version of *Linguistic Inventory and Word Count* (LIWC), as well as Coltheart's (1981) *MRC Psycholinguistic Database* (MRCPD).<sup>2</sup> This allows for generalizability to topics different than those in test samples. The classification schemes for both LIWC and MRCPD search for linguistic features in a text, such the number of first person pronouns, six letter words, and more. After processing the Pennebaker & King (1999) and Mehl, Gosling & Pennebaker (2006) corpora and the Coltheart (1981) database, Mairesse et al. (2007) train several machine learning algorithms on a random sample, finding Support Vector Machines for Regression (SMOreg) best recovers personality measures of respondents in written trials.<sup>3,4</sup>

Applying this approach to Congress requires legislator texts to feed to the models. To that end, the most systematic data available are speeches on the floors of the House and Senate, which can be found in the *Congressional Record*. That said, there exist concerns with using these data, as legislators might use floor speeches to strategically convey their policy preferences, constituency preferences, or (remote, but possibly) personality or leadership profiles. However, we show below that our measures of personality, while correlating with ADA ideological scores (and generally in the expected directions), explain only a small fraction of the variance, suggesting that personality is not equivalent to ideology and our estimates have meaning independent of it. As for the third concern, this is only a problem if legislators try to convey “fake” profiles. Relative to one-time self-administered personality surveys, the Mairesse et al. (2007) method should result in more accurate profiles, as it will be harder to maintain a “fake” profile over the course of a career than over the span of a ten-minute survey. Nevertheless, this effect would simply attenuate

---

<sup>2</sup>The categories are found in the appendix.

<sup>3</sup>JAVA code for Mairesse et al.'s (2007) Personality Recognizer program is found at <http://people.csail.mit.edu/francois/research/personality/recognizer.html>.

<sup>4</sup>However, for our application, the choice of model used (SMOreg versus M5' model trees with linear models) seems to make little substantive difference, presumably due to the size of our data.

our results. Furthermore, if some legislators were speaking sincerely and others strategically, this would result in further attenuation, whereby our estimates would have little predictive power.

Last, we might be concerned that the subject matter reflected in the Pennebaker essay corpus might not map well to legislator floor speeches. Specifically, the kinds of words used by laboratory participants are almost certainly different from those employed by legislators in their floor speeches. To assess this concern, Figure 1 plots the hyperbolic arcsine-transformed mean LIWC category usage from the Pennebaker data against the hyperbolic arcsine-transformed mean usage of members of the 114th House of Representatives.<sup>5</sup> Each point represents a LIWC category; a version of this plot with category labels is found in the online Appendix. As we see, there is a close correspondence between the two; the correlation on the untransformed scale is approximately 0.99. Additionally, the  $\tau_B$  rank correlation coefficient for the rank orderings across the two is 0.72 ( $p < 0.001$ ). Thus, while the subject matter of the texts are different, legislators' usage of the various categories does not differ substantively from the laboratory essay writers.

Since none of these concerns seem to be particularly damning, we apply Mairesse et al.'s (2007) SMOReg model to the corpus of every legislative speech by every sitting member of the House of Representatives in the 104th–113th Congresses (1996–2014). The procedure is as follows:

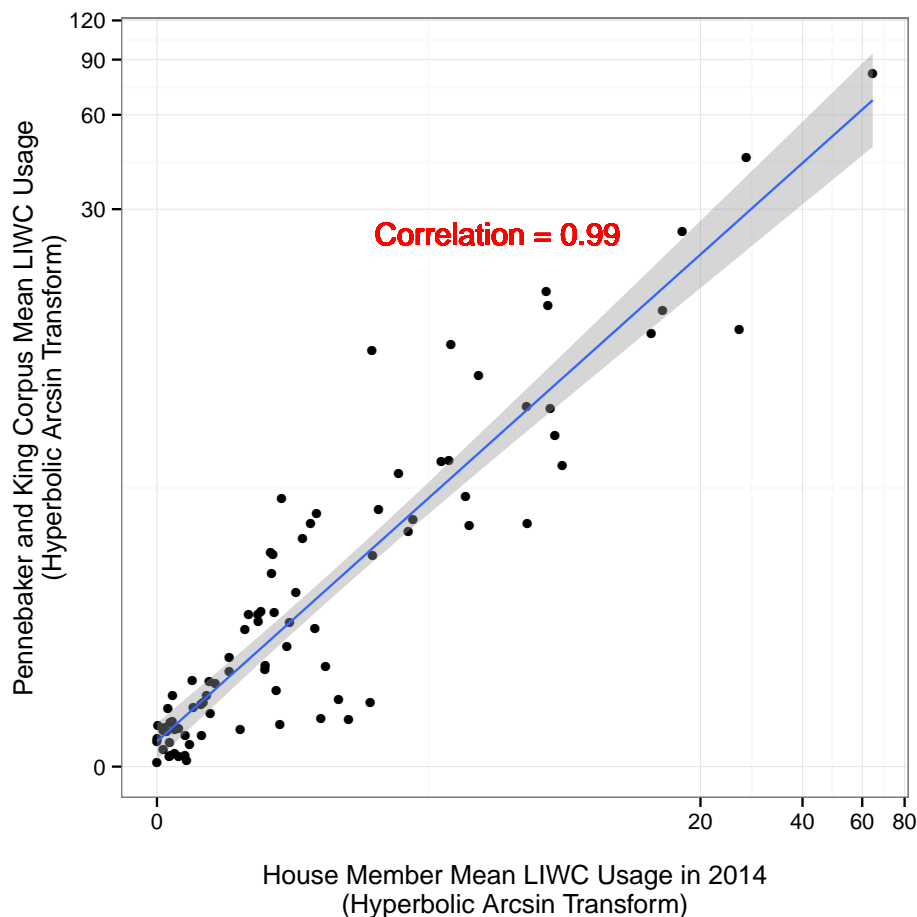
1. For each set of speeches by a given legislator in a fixed time span (in our case, by year, though Congress-level estimates can be similarly estimated), process the raw text through LIWC and MRCPD to get counts of word usage across all LIWC and MRCPD categories.<sup>6</sup>
2. Once all speeches have been processed for the given span, feed the directory of texts through the Mairesse et al. (2007) models in corpus mode. Specifically, all LIWC and MRCPD features are standardized and then fed in to the standardized SMOReg models. Though

---

<sup>5</sup>This transformation is necessary for presentation, as some LIWC categories have larger observed values than others.

<sup>6</sup>Our choice to do the estimation by year is not related to the debate over whether personality is static or dynamic. Rather, since language changes from Congress to Congress, it is important to ensure language is measured relative to the time period in which it is delivered.

Figure 1: Comparing LIWC (2001) Usage between the Pennebaker Corpus and Floor Speeches



we showed above that the category usage of the training (essay) sample is substantively similar to the legislator sample, using the standardized models offers additional leverage on estimating the personality traits within the legislative domain.

3. The output from Step 2 (Big Five estimates for each legislator) is saved and the process is repeated for as many years as there are available.
4. (Optional correction) A legislator  $i$ 's personality on dimension  $d$  in year  $t = 1, 2, \dots, T_i$  ( $T_i$  is the number of years in which legislator  $i$  serves) is rescaled as the jackknifed aver-

age of his/her personality estimates on that dimension for all years except  $t$ , i.e.,  $\tilde{\theta}_{it}^d = \frac{1}{T_i-1} \sum_{t' \neq t} \hat{\theta}_{it'}^d$ . This ensures that personality trait estimates in session  $t$  are not corrupted by language associated with particular actions in that session (e.g., for example, legislators that cosponsor more might be perceived as more Agreeable, not because they cosponsor more legislation *per se*, but simply because the act of cosponsorship inherently entails using Agreeable language).<sup>7,8,9</sup> In short, this correction alleviates concerns of endogeneity.

5. To generate measures of uncertainty, we follow Lowe & Benoit (2011) in using a sentence-level bootstrap. Specifically, suppose legislator  $i$  utters  $N_{it}$  sentences in time period  $t$ . For all  $i$ , we resample from the set of their sentences  $N_{it}$  times with replacement (Efron & Tibshirani 1994). At the year level, we conduct 100 bootstrap replications per member and compute the empirical 95% confidence interval.<sup>10</sup> To measure uncertainty in the jackknifed estimates, we take the legislator’s estimates by year across each of the 100 bootstraps, calculate the jackknife per Step 4, and then compute the 95% confidence interval.

We refer to these estimates as **Elite LingUistiC Individual Difference EstimATIOn (ELUCIDATION)** scores.<sup>11</sup> Importantly, we are agnostic as to whether ELUCIDATION scores are measures of “sincere” legislator personality. Like with ideal-point estimates based on roll calls, we simply consider these estimates as *revealed* and potentially strategic preferences. Estimates of the traits and confidence intervals are presented for key House members in Figure 2. The scale for each trait ranges from 1 to 7. As we see, the estimates are stable and precisely estimated.

Additionally, Figure 3 presents the relationship between word count and confidence interval

---

<sup>7</sup>Absent strategic concerns, the raw scores or simple arithmetic means across years may be used.

<sup>8</sup>Another way to estimate the jackknifed scores would be to pool all speeches excluding the current year, perform the estimation, and repeat for each year. We avoid this for two reasons. First, the agenda changes substantially from year-to-year and feature usage patterns can change over time. Second, it is computationally expensive to do so.

<sup>9</sup>Congress-level jackknife estimates can be estimated in a similar way. Let  $c_i = 1, 2, \dots, C_i$  be the number of Congresses in which member  $i$  served. Since each Congress is a two year period, denote the years associated with Congress  $c_i$  by  $t(c_i)$ . Member  $i$ ’s jackknifed score on dimension  $d$  in Congress  $c$  is thus  $\hat{\theta}_{ic}^d = \frac{1}{T_i-2} \sum_{t' \notin t(c_i)} \hat{\theta}_{it'}^d$ .

<sup>10</sup>The number of bootstrap replications has little effect on the estimates.

<sup>11</sup>Upon publication, the scores will be released.



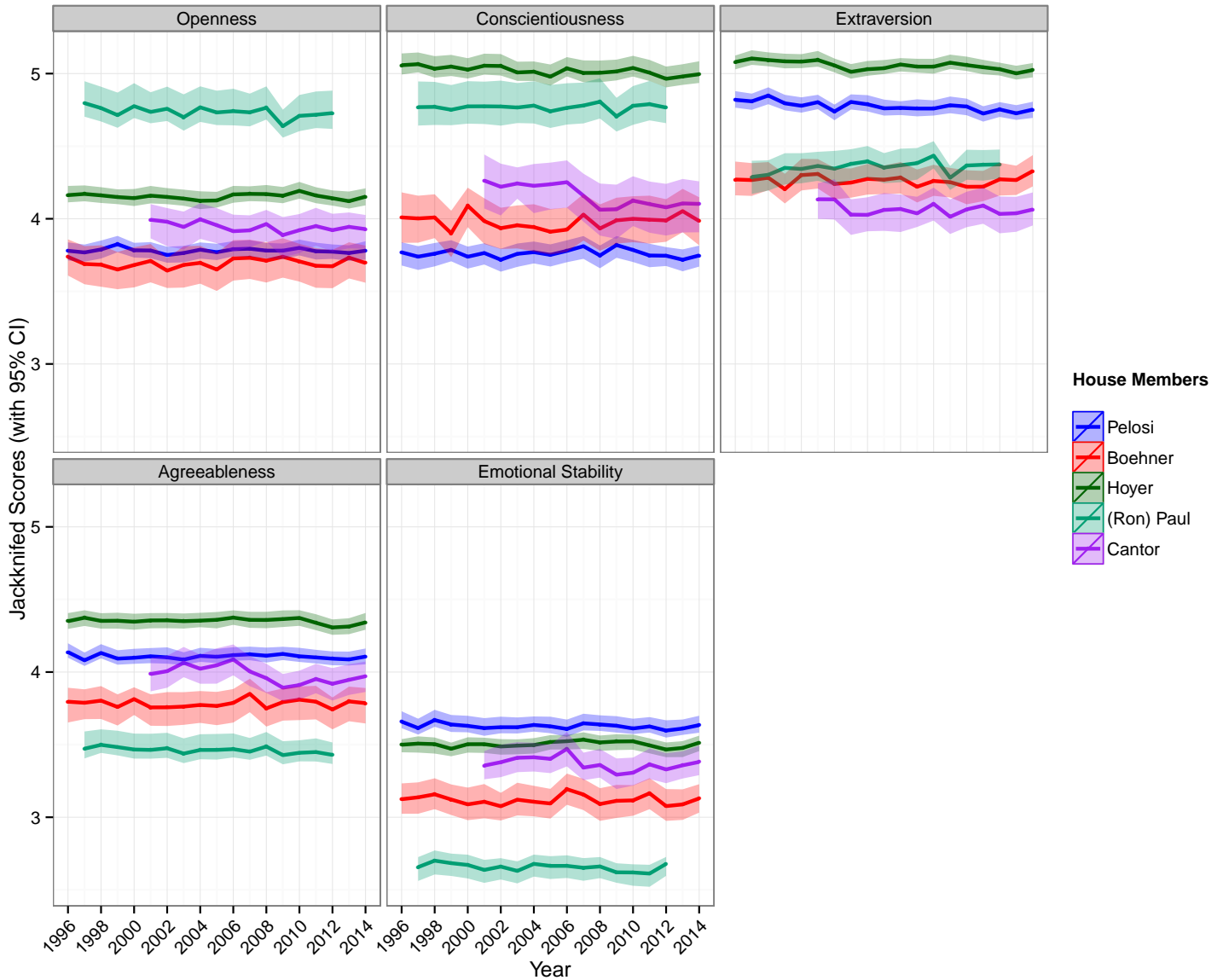


Figure 2: House Scores over Time (selected members)

width. The median member utters around 11,000 words per year. Once the word count exceeds about 5,000 words, the confidence interval width is about 0.3 or less, meaning the 95% interval is the point estimate  $\pm 0.15$  on the 7-point scale. Thus, the estimates are relatively precise for the vast majority of legislators. Moreover, the estimates make intuitive sense. Former Representative Ron Paul (R-TX) is significantly more Open and less Emotionally Stable than the rest of the

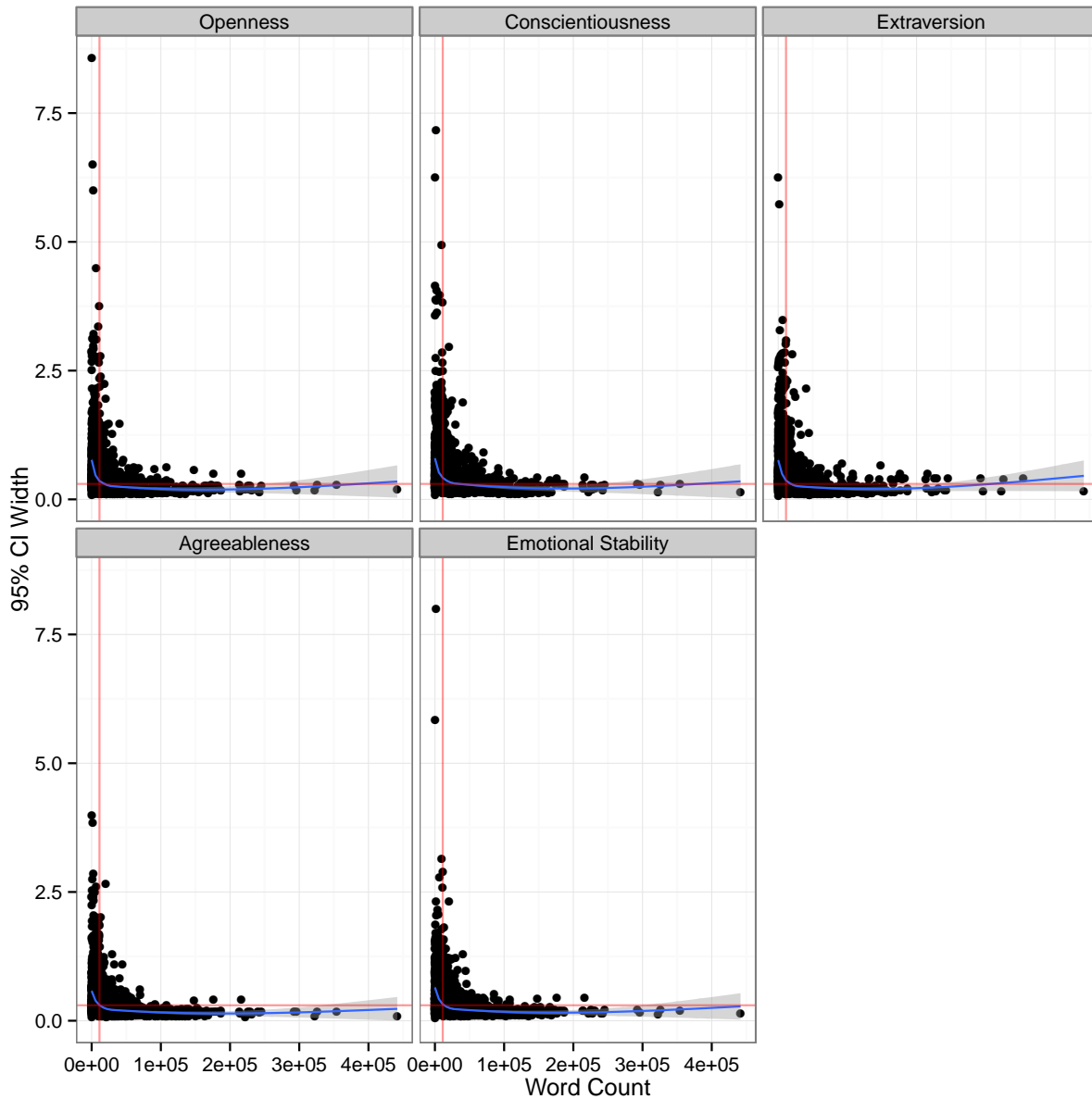


Figure 3: Word Count and Precision. Points are individual members. The smoothed line is a loess-smoothed trend. The vertical red lines are the median member’s word count (approximately 11,000 words per year). Horizontal red lines are the average confidence interval width (approximately 0.3). Thus, most members’ personality estimates are their point estimates  $\pm 0.15$ .

legislators in this subsample. Representative Paul’s out-of-the-box libertarian ideology and his frequent diatribes against the status-quo politics embodied by the two major parties are in line

with common characterizations of both of these traits (McCrae & John 1992, Jost, Nosek & Gosling 2008).

### 3 Validity Checks

#### 3.1 Strategic Concerns

We acknowledge legislators may have incentives to misrepresent their “true” personality traits and adopt false personas. In this case, their behavior and language would be affected to portray the desired traits rather than those they hold with conviction. While we acknowledge this concern as a serious issue, based on the evidence currently available, we do not believe that it is a serious impediment to the measurement of legislators’ beliefs and preferences, exhibited through the Big Five personality traits.

Indeed, a common objection to the standard NEO-PI-R personality inventory (Costa & McCrae 1992) is that the “desired” trait or answer is transparent to respondents. Critics of the inventory argue the average respondent would certainly prefer to think of themselves as extraverted rather than introverted, emotionally stable rather than neurotic, agreeable rather than selfish, conscientious rather than lazy, and open to experience rather than insular.<sup>12</sup> This concern appears to gain support from personality inventories administered to state legislators (Dietrich et al. 2012). The Big Five personality trait questionnaires show that the responding legislators’ responses are substantially skewed toward these “favorable” personality trait values, suggesting personality traits serve as valence characteristics.

Legislators may consciously adopt speech patterns in order to signal these valence characteristics. There are likely many verbal cues that both indicate personality types to legislators and

---

<sup>12</sup>While this may seem to bias the results, there is substantial variation in sample responses that allow the measure to retain meaning. This is likely due to efforts to maintain credibility, as when candidates try to pivot to the center following a primary campaign and cannot with credibility adopt the position of the median voter. If acting against one’s preferences and belief structures is costly, it would be difficult to credibly present a false personality

voters, and are known to all and can be voluntarily suppressed or performed insincerely. However, we also assume there exists some subset of these cues tied to personality, which are unknown to legislators, but may be captured by the personality recognizer described above.

We suppose legislators also have unconscious verbal cues that signal personality and that the algorithm can read them. In that case, it is in the interests of legislators to misrepresent their personalities by appearing to hold personality trait values as favorable as possible throughout their public speech. While all known cues should be utilized to develop valence trait values, the unknown subset of cues should exhibit variation as legislators are simply unaware of their link with personality. Combining these known and unknown subsets of speech elements should result in measures that appear skewed toward the valence characteristics, but with substantial variation reflecting the true variation in the underlying trait values.

### **3.2 Face Validity**

Theoretical concerns aside, it is important to ensure our estimates have some degree of face validity. Traditionally, personality psychologists have relied on surveys to develop measures on the Big Five dimensions. However, as we note above, surveying current and former federal legislators is not likely to yield adequate responses given the difficulty other scholars have faced in reaching currently serving state legislators. Therefore, we will instead rely on more indirect validity checks based on associations repeatedly uncovered in the literature. Our analysis here will focus on the relationship between personality and ideology.

As mentioned, much recent research has focused on the relationship between personality and partisan attitudes (Gerber et al. 2010, Gerber et al. 2011, Mondak 2010). Strong and consistent relationships have been uncovered between ideology and Openness, Conscientiousness, and Emotional Stability, with the former being associated with liberalism, and the latter two associated with conservatism. Effects for the other two traits are more mixed; Mondak (2010) finds no relationship between Extraversion and ideology, and only a weak—and inconsistent—relationship

between Agreeableness and liberalism. Gerber et al. (2011) find the same relationships between ideology and Openness, Conscientiousness, and Emotional Stability. Moreover, they find no relationship between Agreeableness and self-reported ideology, but uncover divergent relationships between Agreeableness and economic attitudes, and Agreeableness and social attitudes; more Agreeable people are more economically liberal and socially conservative. Finally, they find a positive relationship between Extraversion and conservatism, though the relationship is weaker than those between ideology and the other four traits. Similarly, Gerber et al. (2010) find consistent relationships between ideology and Openness, Conscientiousness, and Emotional Stability in the aforementioned directions, but the relationships between ideology and Agreeableness and Extraversion are more nuanced, as mentioned.

Nonetheless, we can use these findings to perform an indirect test of the validity of our estimates. Using Groseclose, Levitt & Snyder (1999) inflation-adjusted ADA scores, we can examine the direction of the relationships between the Big Five traits and ideology.<sup>13</sup> From these findings, we should expect a positive relationship between Openness and the ADA score (recall, ADA scores range from conservative to most), and negative relationships between Conscientiousness, Emotional Stability, and the ADA score. Additionally, as ADA and DW-NOMINATE scores tap into the underlying economic conflict between the two dominant parties (roughly approximate to the contemporary liberal-conservative dimension), we should find a positive relationship between Agreeableness and the ADA score (Crespin & Rohde 2010, Poole & Rosenthal 1997).<sup>14</sup> Finally, the literature is much more mixed on the relationship between ideology and Extraversion; as such, we remain agnostic regarding the expected sign (or significance). Table 1 presents several OLS models of ideology as a product of personality and demographic traits.

All four traits with expected relationships have statistically significant coefficients in the expected directions in each model. Additionally, in line with previous literature, the coefficients on

---

<sup>13</sup>We opt for this instead of DW-NOMINATE because our estimates are at the year level. Substantively similar results are found aggregating our scores over lifetime and using DW-NOMINATE scores.

<sup>14</sup>No other trait has these divergent effects.

Table 1: OLS Models of Personality and ADA Score (1996-2008)

	Model 1	Model 2	Model 3	Model 4
Openness	18.57*** (1.34)	18.36*** (1.32)	17.36*** (1.34)	17.36*** (1.33)
Conscientiousness	-19.20*** (1.17)	-17.02*** (1.16)	-19.39*** (1.16)	-17.28*** (1.16)
Extraversion	12.51*** (0.84)	11.50*** (0.84)	12.74*** (0.84)	11.74*** (0.83)
Agreeableness	10.70*** (2.31)	9.36*** (2.28)	11.30*** (2.30)	9.93*** (2.27)
Emotional Stability	-17.79*** (1.45)	-14.95*** (1.44)	-18.26*** (1.44)	-15.49*** (1.44)
Male	—	-17.08*** (1.38)	—	-16.21*** (1.38)
Age	—	—	0.35*** (0.05)	0.29*** (0.05)
Constant	11.49 (6.53)	18.83** (6.47)	-4.21 (6.85)	5.43 (6.82)
R <sup>2</sup>	0.09	0.11	0.10	0.12
Adj. R <sup>2</sup>	0.09	0.11	0.09	0.12
Num. obs.	5252	5252	5252	5252

Standard errors in parentheses.

Two-tailed tests: \*\*\*  $p < 0.01$ , \*\*  $p < 0.05$ , \*  $p < 0.1$

Extraversion are of smaller magnitudes than those for Openness, Conscientiousness, and Emotional Stability, and this holds for all four models (though the relationship between ideology and Extraversion is still a point of contention). Moreover, these results allay one natural concern with using potentially ideologically-tinged legislative speeches to estimate personality, in that our personality estimates may be simply summaries of legislator ideology. However, our model  $R^2$ s in Table 1 are not large, suggesting personality traits alone do not account for large proportions of the variance in ideology. All together, the results are in line with expectations from the literature (Jost, Nosek & Gosling 2008, Gerber et al. 2010). Specifically, Jost, Nosek & Gosling (2008) find that while there are correlations between the Big Five and ideology, the magnitudes of these are not large. Our adjusted  $R^2$  is very much consistent with this finding. This suggests whatever

theoretical concerns about the dependence of the Big Five on ideology or the ideological content of the legislative record are not a problem.

### 3.3 Personality Traits and Linguistic Features

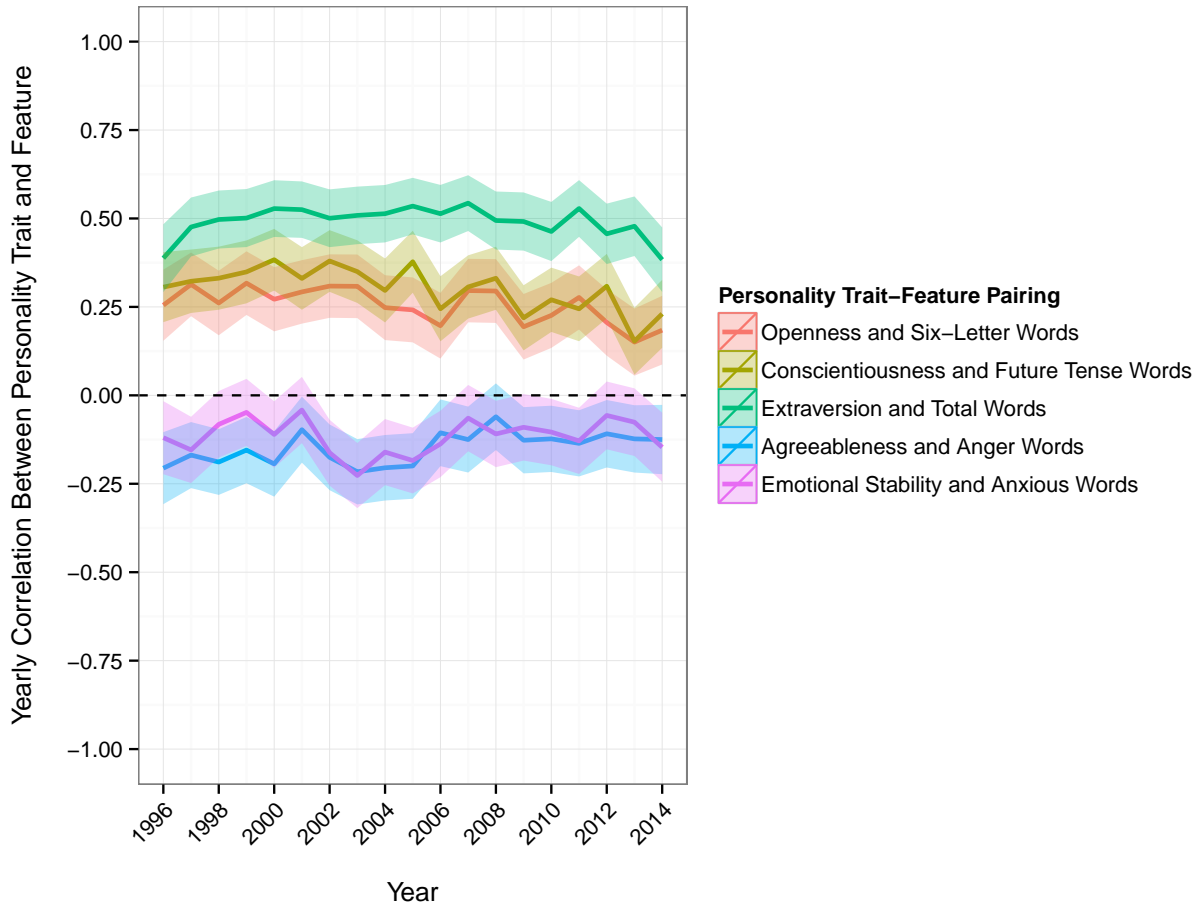
We can also examine the relationships between LIWC/MRCPD features and the estimated personality traits to see if they have face validity. While a full feature exploration is beyond the scope of this paper due to reasons of length (as there are over 80 features), we can examine the stability of selected relationships. To wit, Figure 4 presents the yearly correlations between the Big Five and one feature per trait.<sup>15</sup> The correlations are remarkably stable over time and generally are distinct from zero at the 95% level.

Perhaps more importantly, these relationships make sense. As Openness is associated with an appreciation for aesthetic beauty, intellectual pursuits, a tendency for creativity and imagination, we might expect more Open members to express themselves in more conventionally “intellectual” ways; one rough proxy for intellect is the length of words used (Tupes & Christal 1992, Borgatta 1964). As such, it should be unsurprising that more Open members use longer words at higher rates. Additionally, more Conscientious individuals tend to be more goal-oriented, and should focus on the future more than those who are less Conscientious (Ozer & Benet-Martinez 2006). As Figure 4 shows, this is also supported, as more Conscientious members use future tense words at higher rates. Similarly, more Extraverted members should be more sociable and talkative, and they do speak more in our data. Finally, more Agreeable legislators should use angry, off-putting language at lower rates, and more Emotionally Stable members should exhibit less outward anxiety; again, these relationships are found in our data, though are somewhat less strong than those for the other three traits. Nonetheless, the relationships between the linguistic features and the estimated personality traits are not only stable, but they are generally significant in the expected directions, thus providing additional face validity to our

---

<sup>15</sup>All yearly correlations for all trait-feature pairs are in the Appendix.

Figure 4: Correlations Between Personality Traits and Selected LIWC Features



estimates.

### 3.4 Authorship and Speechwriter Effects

An additional important concern to address is the potentiality for speechwriting effects. Specifically, since at least some of members' speeches are written by their staff members, our estimates might be reflective of the writers' personalities and not those of the Congressmembers. This turns out to be a non-issue for a number of reasons. Few members have professional speechwrit-



ing staff. A cursory look at the House disbursement records demonstrates that, outside of the Speaker and a few key members, speechwriters are simply not hired by most members.<sup>16</sup> This does not mean that other staffers do not write speeches. However, since staff turnover is notoriously high,<sup>17</sup> it is reasonable to assume that *many* staffers would contribute to many speeches and, if so, our estimates should be unstable or clustered for all members. As Figure 2 shows, for some of the highest ranking and/or most prominent members of Congress during our time period, there is marked stability in the jackknifed personality estimates. Perhaps most assuring, since neither Pelosi or Boehner were Speaker at the start of the data and the estimates remain consistent once they become Speaker, their speechwriters clearly maintained the same linguistic patterns used by those politicians earlier.

Indeed, the Congressional Research Service admonishes those individuals charged with speechwriting for Congressmembers to make every effort to mimic the language and diction of the member for whom they are writing (Neale 1998):

“Congressional speechwriters should make every effort to become familiar with the speaking style of the Member for whom they are writing, and adjust their drafts accordingly.”

These results somewhat contradict those of Sigelman (2002), who argues that President Reagan’s revealed persona was different in ghostwritten and off-the-cuff remarks. However, he notes that while differences exist between Reagan with speechwriters and Reagan without speechwriters, they are simply variations on a theme. Since presidents have arguably the most incentive and ability to shape their public personalities, and the difference is still minimal, there should be even less of an effect for members of Congress. This is precisely what we observe.

---

<sup>16</sup>See, e.g., [http://disbursements.house.gov/2013q4/2013q4\\_singlevolume.pdf](http://disbursements.house.gov/2013q4/2013q4_singlevolume.pdf).

<sup>17</sup>Jaffe, Alexandra. 2012. “Report: High Turnover, Low Pay Make Hill Jobs Difficult.” *The National Journal*, June 7, 2012, <http://www.nationaljournal.com/congress/report-high-turnover-low-pay-make-hill-jobs-difficult-20120607>

## 4 Application 1: Speeches Versus Surveys

Despite the advantages proffered by the speech-based approach, there may be circumstances in which a survey-based approach might be preferable, at least for contemporaries. Indeed, surveys might be preferable for these types of legislators if they were perfectly responsive. However, missingness would not necessarily be a problem, as personality traits could conceivably be imputed using existing algorithms (*e.g.*, Gelman, King & Liu 1998, Rubin 1976, Rubin 1977, Rubin 1987). This is a potentially appealing route, as this would allow us to recover estimates for those who did not respond. However, this approach has limitations. Most important for our purposes here is that the missingness must obey the so-called missing at random assumption. Following King et al. (2001), let  $D_{obs}$  denote observed data,  $D_{mis}$  denote missing data,  $D$  denote the total data, and  $M$  represent missingness. Data are missing at random (MAR) if  $\Pr(M|D_{obs}, D_{mis}) = \Pr(M|D_{obs})$ . That is, data satisfy MAR if missingness can be modeled as a function of observed data. A canonical example of this sort of missingness is the case of high wage-earners who fail to report their income in surveys. While the income may be unobserved, several known correlates (*e.g.*, education) are not missing. By conditioning on these observed values, we may model missingness using existing algorithms. However, if the missing observations cannot be predicted from observed covariates, MAR is not satisfied and multiple imputation is no longer an option (Weisberg 2009).

This is most problematic if the missingness is in part due to personality. In these cases, missingness is a function of the underlying latent traits the surveys seek to recover, meaning that several characteristics that help predict missingness are themselves missing. Therefore, imputation cannot be used with any degree of certainty. Moreover, the distribution of responsive legislators is subject to selection bias, and any inferences drawn from these data will themselves be problematic. Using predicted values from Heckman (1976) selection models will not suffice either, as important covariates that predict selection into the sample (*i.e.*, personality traits) will be missing.

To check whether this is a problem, we examine the relationship between our recovered ELU-

CIDATION scores and response rates to the National Political Awareness Test (NPAT), a survey developed by Project Vote Smart that asks candidates for Congressional elections to provide answers to a series of political issue questions.<sup>18</sup> This survey has received a great deal of attention by political scientists and has been used with great fruit to estimate the ideology of legislators and candidates (e.g. Shor & McCarty 2011, Ansolabehere, Snyder Jr & Stewart III 2001, Battista, Peress & Richman 2012). Response rates to the NPAT are higher than those from previous (anonymous) studies of personality, presumably because the NPAT questions ask about policy positions, which are much more in line with the roles played by legislators, as opposed to the arguably more personally intrusive personality surveys. However, both types of surveys have potential consequences for giving the “wrong” response. As Dietrich et al. (2012) rightly note about their surveys, “respondents may have been concerned about social desirability, leaving them unwilling to admit to possessing certain traits” (201). Similarly, Project Vote Smart notes that “[m]ost candidates, fearing their opponents might use their positions in attack ads, refuse or only respond to a few questions that their consultants stamp as safe.”<sup>19</sup> Indeed, this attitude is institutionalized within the major party leadership; in 2006, a Democratic leader in the Florida House of Representatives told a *Wall Street Journal* reporter that “We tell our candidates not to do it. It sets them up for a hit piece.”<sup>20</sup>

We estimate a series of logistic regression models where the dependent variable equals one if the member in question ever responded to the NPAT, and zero otherwise. This is arguably a very low bar, and therefore biases against the possibility of finding significant results, as this approach minimizes the amount of variance in the dependent variable. As such, any significant results should provide strong evidence of a relationship between personality and nonresponse.

---

<sup>18</sup>Project Vote Smart was discussed only twice in the *Congressional Record* during the years under analysis, and only by Reps. Henry Hyde and Alcee Hastings. Therefore, the likelihood of discussion of the survey or the larger organization affecting the estimates in any meaningful sense is effectively zero.

<sup>19</sup>See <http://votesmart.org/about/political-courage-test>

<sup>20</sup>Grant, Peter. 2006. “Politicians Grow Wary Of Survey as Internet Spreads Attack Ads.” *The Wall Street Journal*. October 25, 2006. Page B1.

Table 2: Predicting Response to the NPAT At Least Once

	Model 1	Model 2	Model 3	Model 4
Openness	−0.15 (0.23)	−0.08 (0.25)	−0.14 (0.25)	−0.08 (0.25)
Conscientiousness	0.17 (0.20)	0.26 (0.21)	0.38* (0.22)	0.29 (0.22)
Extraversion	−0.09 (0.14)	−0.10 (0.15)	−0.17 (0.16)	−0.20 (0.16)
Agreeableness	−0.39 (0.39)	−0.53 (0.40)	−0.61 (0.41)	−0.57 (0.41)
Emotional Stability	0.47* (0.25)	0.49* (0.27)	0.62** (0.28)	0.58** (0.28)
Democratic Normal Vote	—	0.02** (0.01)	0.02* (0.01)	−0.02 (0.01)
South	—	−0.17 (0.19)	−0.14 (0.19)	−0.13 (0.20)
Female	—	0.22 (0.26)	0.19 (0.26)	0.10 (0.27)
Seniority	—	−0.05* (0.03)	−0.06** (0.03)	−0.07** (0.03)
Legislator Voteshare <sub>t−1</sub>	—	0.01 (0.01)	0.01 (0.01)	0.02* (0.01)
Leader	—	−0.49 (0.91)	−0.36 (0.91)	−0.60 (0.90)
Ideological Extremism	—	—	−0.04* (0.02)	1.08*** (0.27)
Majority Party	—	—	—	−11.38*** (2.76)
Constant	−0.12 (1.09)	−1.89 (1.28)	−1.72 (1.29)	4.79** (2.04)
BIC	881.23	869.07	871.99	860.51
Log Likelihood	−421.32	−396.17	−394.44	−385.50
Num. obs.	622	598	598	598

Notes: Standard errors in parentheses. Observations are pooled by Congressman.

Two-tailed tests: \*\*\*  $p < 0.01$ , \*\*  $p < 0.05$ , \*  $p < 0.1$

Independent variables include all personality traits, as well as host of other demographic information on legislators. These include dummy variables for gender (*Female*), region (*South*), whether or not legislators are *Leaders*, and whether or not they are in the *Majority Party*. Additionally, we control for district partisanship (*Democratic Normal Vote*), *Seniority*, electoral vulnerability (*Legislator Voteshare* in the last election), and *Ideological Extremism* (absolute value of their NOMINATE score). Because we are estimating the likelihood of *ever* responding to the NPAT, each member has but one observation in our data; as such, the independent variables are actually set

to the Congressman-level means.<sup>21</sup> Results are in Table 2.<sup>22</sup>

As Table 2 shows, the propensity of nonresponse is a function of personality. Perhaps most notably, Emotional Stability is positive and significant at conventional levels at all models, regardless of which controls are included. This finding provides strong evidence that the propensity to respond to surveys on which social desirability may be a factor is at least in part a function of personality traits. Given that Dietrich et al. (2012) note that social desirability likely plays a role in the decision to respond (and, perhaps more worryingly, the revealed traits conditional on response), we have every reason to believe that data gathered through survey methods alone are inappropriate for investigating the relationships between personality traits and legislative behavior, as well as that of elites in general.

Figure 5 makes these effects even clearer. In this figure, we plot the predicted probability of NPAT response varying Emotional Stability while holding other variables at their means (using the results from Model 4 in Table 2). As we see, increasing Emotional Stability from two standard deviations below its mean to two standard deviations above increases the predicted probability of response from about 30% to about 53%. This increase is enormous and suggests estimates of personality from surveys of elites are highly susceptible to personality-induced selection bias.<sup>23,24</sup>

---

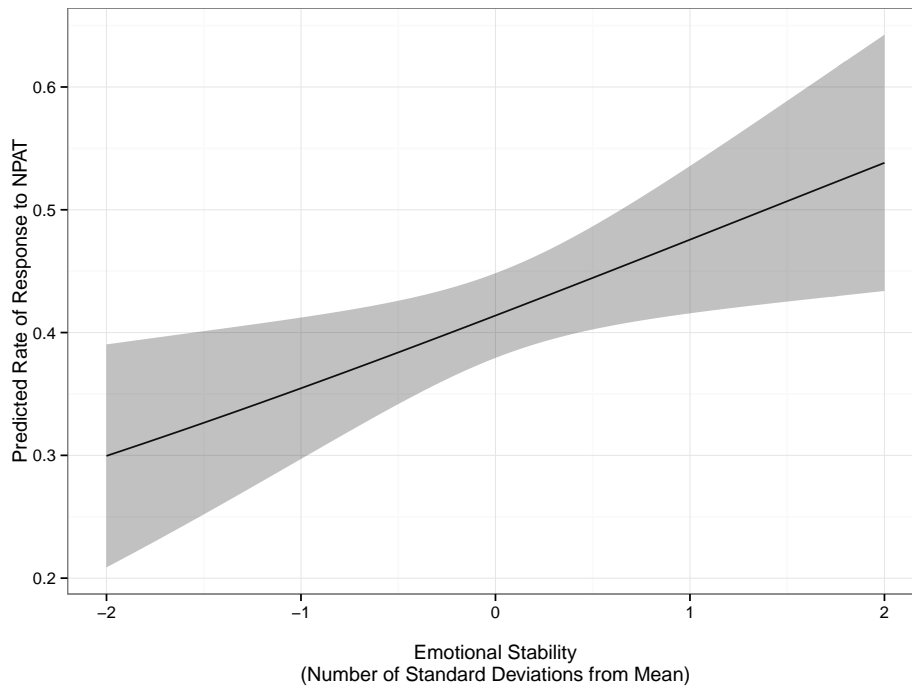
<sup>21</sup>Project VoteSmart only reports a legislators most recent survey. Thus, at best, we can only tell if a legislator in our data has *ever* responded to the survey. This necessitates our pooling strategy.

<sup>22</sup>We also tested models including fixed effects for states in which the delegation size is larger than 5. Results are substantively similar.

<sup>23</sup>That said, surveys of legislators do have potential applications in the measurement of legislator—and, more generally, elite—personality traits. For starters, Mairesse et al.’s (2007) Personality Recognizer was trained on non-political speech. One concern might be that, despite the similarity in word frequencies and distributions noted earlier, political and nonpolitical speech are fundamentally different from one another. In this case, elite surveys—coupled with spoken word transcripts—could be used to train elite- and/or domain-specific recognizers. Additionally, it might be worthwhile to use the ELUCIDATION scores to examine the relationships between legislator personality traits and the likelihood of response (or, alternatively, the types of responses) on personality surveys.

<sup>24</sup>While surveys of legislators may prove problematic for the study of elite personality, surveys of experts may prove to be a fruitful “crowd-sourced” avenue for improving our machine learning approach. For example, Rubenzer & Faschingbauer (2004) administer a Big Five inventory to a series of experts so as to measure personality traits for U.S. presidents. One could use the expert ratings to retrain the machine learning algorithm considered herein, using these ratings in conjunction with presidential speeches. This would undoubtedly help to hone in on language pertinent to the political domain.

Figure 5: Probability of Responding to the NPAT



## 5 Application 2: Symbolic Bill Proposals

Finally, we present a brief substantive application of these scores to the legislative domain—determining the types of bills legislators propose. Do they propose symbolic bills intended to rename post offices after local figures (for example), or substantive bills geared toward effecting major policy changes? Different types of bills should reflect different underlying goals and should also require varying degrees of effort to successfully shepherd through the legislative process. However, different types of bills also face different hurdles, and more substantive bills are at greater risk of being prevented from reaching the floor due to partisan concerns (*e.g.*, Aldrich 1995, Cox & McCubbins 2005). Given these hurdles, we expect more Conscientious legislators to propose fewer ceremonial/symbolic bills, as people who are more Conscientious are described as having tendencies toward hard work, responsibility, and planning (VandenBos 2007),

and they tend to be more driven, goal-oriented, uptight, organized and to have more willpower (Ozer & Benet-Martinez 2006).

To estimate the importance of Conscientiousness, we analyzed all bills proposed in the House during the 104th–112th Congresses. Each bill was categorized using Volden & Wiseman’s (2014) criteria as being ceremonial/symbolic or of substantive importance.<sup>25</sup> We then estimated a series of binomial logistic regressions; for these, the proportion of proposed bills that are ceremonial/symbolic in nature is the dependent variable.<sup>26,27</sup> The independent variables include our jackknifed ELUCIDATION scores, as well as several other independent variables that might affect the decisions. *Ideology* is parameterized as the member’s first-dimension DW-NOMINATE score (Poole & Rosenthal 1997), and *Ideological Extremism* is the squared first-dimension DW-NOMINATE score. *Majority Party* is an indicator variable equaling one if the member and Speaker are of the same party, and zero otherwise. *Seniority* denotes the number of terms a member has served. *Legislator Voteshare*<sub>*t*-1</sub> denotes the percentage of the vote (on a 0–100 scale) the member received in the election to the current Congress. *Committee Chair* and *Subcommittee Chair* are indicator variables equaling one if the member served in the relevant committee leadership role in the Congress under analysis, and zero otherwise. *Power Committee* is an indicator variable that captures whether the member in question sat on at least one of the three most powerful committees in the House—Appropriations, Rules, and Ways and Means; if the member sat on at least one of these committees, the *Power Committee* variable equals one, and it equals zero otherwise. Finally, we account for the importance of party leadership by including indicator variables for *Speaker*, *Majority Leadership*, and *Minority Leadership*, which are indicator variables equaling one if the member served in the relevant role in the Congress under analysis,

---

<sup>25</sup>Volden & Wiseman (2014) classify bills as ceremonial/symbolic if any of the following occur in the title: “commemoration, commemorate, for the private relief of, for the relief of, medal, mint coins, posthumous, public holiday, to designate, to encourage, to express the sense of Congress, to provide for correction of, to name, to redesignate, to remove any doubt, to rename, and retention of the name” (21).

<sup>26</sup>We drop all observations where zero bills were proposed.

<sup>27</sup>Results are substantively identical if negative binomial models—with logged offsets of one plus the total number of bills proposed per member-Congress dyad—are used instead.

and zero otherwise. We account for previously-established correlations between personality and personal demographics by including variables for *Age* and whether or not the member is *Female*; the former denotes the member's age at the end of the first session of each Congress, and the latter is an indicator variable equaling one if the member identifies as female and zero otherwise. Additionally, we include interactions between the ELUCIDATION scores and *Majority Party*.

Results are in Table 3, and Figure 6 presents the predicted proportions of ceremonial bill proposals, conditional on Conscientiousness and majority party status, along with 95% confidence intervals.<sup>28</sup> More Conscientious members propose fewer ceremonial/symbolic bills, instead dedicating their energies to more substantive arenas with longer-term payoffs. Moreover, the effect is mitigated by structural factors. For example, Conscientiousness seems to least affect those in the majority—that is, those most likely to see bills they agree with pass without their own influence. These members will be able to free ride on the efforts of the majority party leadership with some assurance that substantive bills with which they agree will come to the floor. Conversely, those in the minority will be more inclined to propose substantive bills—because there will be fewer fellow members of the House on whose efforts they can free ride—so long as they are sufficiently Conscientious.

---

<sup>28</sup>For Figure 6, we use the results from Model 4 from Table 3. All continuous variables are set their means, and all categorical variables are set to their modes.



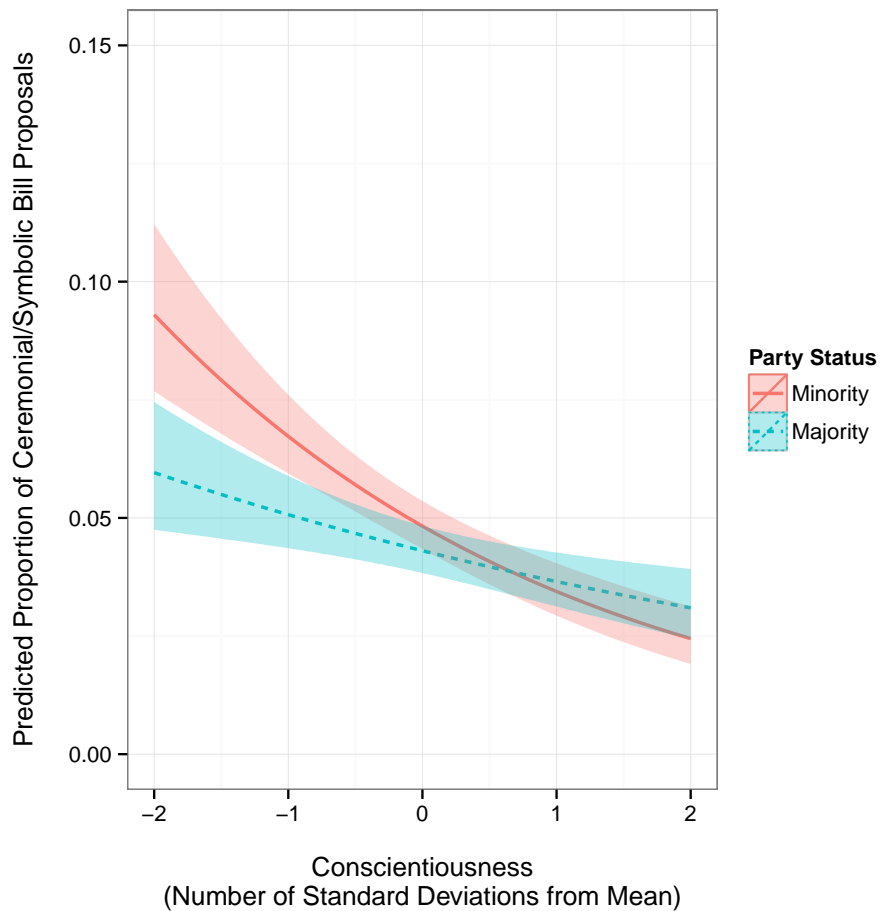
Table 3: Personality and Ceremonial Bill Proposals

	Model 1	Model 2	Model 3	Model 4	Model 5
Openness	0.21*** (0.06)	0.13* (0.07)	0.18*** (0.07)	0.20* (0.10)	0.17 (0.10)
Conscientiousness	-0.52*** (0.05)	-0.43*** (0.05)	-0.39*** (0.06)	-0.53*** (0.08)	-0.58*** (0.08)
Extraversion	0.10*** (0.04)	0.03 (0.04)	0.02 (0.04)	0.19*** (0.06)	0.22*** (0.06)
Agreeableness	0.24** (0.10)	0.18* (0.10)	0.13 (0.11)	0.20 (0.16)	0.28* (0.15)
Emotional Stability	0.11* (0.07)	0.19*** (0.07)	0.19*** (0.07)	0.24** (0.10)	0.21** (0.10)
Majority Party	—	-0.24*** (0.05)	-0.18*** (0.06)	1.35** (0.59)	1.40** (0.60)
Majority Party × Openness	—	—	—	-0.07 (0.14)	-0.07 (0.14)
Majority Party × Conscientiousness	—	—	—	0.27** (0.11)	0.29*** (0.11)
Majority Party × Extraversion	—	—	—	-0.33*** (0.08)	-0.35*** (0.08)
Majority Party × Agreeableness	—	—	—	-0.18 (0.21)	-0.23 (0.21)
Majority Party × Emotional Stability	—	—	—	-0.09 (0.14)	-0.05 (0.14)
Ideology	—	-0.27*** (0.05)	-0.24*** (0.06)	-0.27*** (0.06)	-0.28*** (0.06)
Ideological Extremism	—	0.02 (0.13)	-0.14 (0.14)	-0.12 (0.14)	-0.00 (0.14)
Age	—	—	0.01** (0.00)	0.01** (0.00)	0.01*** (0.00)
Female	—	—	-0.22*** (0.07)	-0.23*** (0.07)	-0.25*** (0.07)
Seniority	—	—	-0.03*** (0.01)	-0.03*** (0.01)	-0.03*** (0.01)
<i>Legislator Voteshare</i> <sub>t-1</sub>	—	—	0.01*** (0.00)	0.01*** (0.00)	0.00*** (0.00)
Committee Chair	—	—	-0.29** (0.12)	-0.31** (0.12)	-0.32** (0.12)
Subcommittee Chair	—	—	-0.15** (0.07)	-0.16** (0.07)	-0.16** (0.07)
Power Committee	—	—	-0.07 (0.06)	-0.08 (0.06)	-0.09 (0.06)
Speaker	—	—	0.50 (0.62)	0.62 (0.62)	0.53 (0.62)
Majority Leadership	—	—	0.11 (0.16)	0.13 (0.16)	0.16 (0.16)
Minority Leadership	—	—	0.32** (0.15)	0.30** (0.15)	0.33** (0.15)
Constant	-3.67*** (0.28)	-3.34*** (0.30)	-4.18*** (0.35)	-4.83*** (0.45)	-5.08*** (0.46)
Congress FE?	No	No	No	No	Yes
Wald Test	158.82***	132.28***	104.90***	121.81***	143.22***
BIC	7953.04	7888.17	7784.17	7799.81	7770.90
Log Likelihood	-3951.81	-3907.03	-3814.09	-3801.38	-3754.08
Num. obs.	3736	3732	3648	3648	3648

Notes: Standard errors in parentheses. Observations are at the Congress-member level. Null hypotheses for the Wald Tests are that all coefficients related to the personality traits are zero.

Two-tailed tests: \*\*\*  $p < 0.01$ , \*\*  $p < 0.05$ , \*  $p < 0.1$

Figure 6: Conscientiousness and Symbolic Bill Proposals



## 6 Conclusion

This paper demonstrates measurement of elite personality is possible and that estimates are consistent and robust over time. We have also demonstrated that survey methods for recovering measures of elite personality are problematic at best and that the Big Five do indeed play a role in legislative decision-making. The next step, incorporating these measures in to theoretically-motivated studies of elite behavior, is a trickier matter and requires careful consideration of how

personality fits in to the standard frameworks of legislative decision-making. These efforts are underway and are found elsewhere (Ramey, Klingler & Hollibaugh 2015).

There are several ways in which the methods discussed in this paper may be extended in future work. First of all, while we believe text-based measures of elite personality are more appropriate for the study of legislative behavior than survey-based measures, there are a number of ways in which the strengths of survey-based methods may be used to enrich the ELUCIDATION project, as discussed above. Additionally, while the Congressional Record corpus is very similar to the Pennebaker corpus used to train the Personality Recognizer, this text source is not available for many political elites of interest, including challengers in congressional races. Retraining a personality recognizer model on alternate text sources (e.g., social media corpora, press releases, party manifestos) would allow for the estimation of elite personality scores for these other actors, thereby widening the subject pool. Additionally, digitizing the Congressional Record before 1996 will allow for the estimation of personality trait scores going back for a significant period of time in history.<sup>29</sup>

ELUCIDATION scores offer legislative scholars a powerful tool for the investigation of the Big Five personality traits among U.S. Congressmembers. For the first time, personality scores within the five factor model are available for virtually all members of the U.S. Congress going back to 1996. With these tools, established findings from personality psychology and the growing literature on personality in political behavior may be applied to develop new theories of legislative behavior. Furthermore, there is an opening for a modeling framework which would allow the Big Five to be incorporated into models of political institutions. The personality scores introduced in this paper now allow the implications of such models to be examined empirically.

In sum, text-based measures of legislator personality make significant improvements over existing measures derived from expert ratings or personality inventories administered through sur-

---

<sup>29</sup>This, of course, is conditional on the distribution of linguistic features being reasonably similar to that in the training set. This is yet to be determined.

veys. They represent a significant innovation and serve to fill a critical gap in the burgeoning literature on elite behavior in institutions.

## References

- Aldrich, John H. 1995. *Why Parties?: The Origin and Transformation of Political Parties in America*. Chicago: University Of Chicago Press.
- Ansolabehere, Stephen, James M Snyder Jr & Charles Stewart III. 2001. "The Effects of Party and Preferences on Congressional Roll-Call Voting." *Legislative Studies Quarterly* pp. 533–572.
- Battista, James Coleman, Michael Peress & Jesse Richman. 2012. "Common-Space Ideal Points, Committee Assignments, and Financial Interests in the State Legislatures." *State Politics & Policy Quarterly* pp. 70–87.
- Borgatta, Edgar F. 1964. "The structure of personality characteristics." *Behavioral Science* 9(1):8–17.
- Caprara, Gian Vittorio, Claudio Barbaranelli & Philip G. Zimbardo. 2002. "When Parsimony Subdues Distinctiveness: Simplified Public Perceptions of Politicians' Personality." *Political Psychology* 23(1):77–95.
- Caprara, Gian Vittorio, Shalom Schwartz, Cristina Capanna, Michele Vecchione & Claudio Barbaranelli. 2006. "Personality and Politics: Values, Traits, and Political Choice." *Political Psychology* 27(1):1–28.
- Coltheart, Max. 1981. "The MRC psycholinguistic database." *The Quarterly Journal of Experimental Psychology: Section A (Human Experimental Psychology)* 33(4):497–505.
- Costa, Jr., Paul T. & Robert R. McCrae. 1992. "Normal Personality Assessment in Clinical Practice: the NEO Personality Inventory." *Psychological Assessment* 4(1):5–13.
- Cox, Gary W. & Mathew D. McCubbins. 2005. *Setting the Agenda: Responsible Party Government in the U.S. House of Representatives*. New York: Cambridge University Press.

- Crespin, Michael H. & David W. Rohde. 2010. "Dimensions, Issues, and Bills: Appropriations Voting on the House Floor." *Journal of Politics* 72(4):976–989.
- Dietrich, Bryce J., Scott Lasley, Jeffery J. Mondak, Megan L. Remmel & Joel Turner. 2012. "Personality and Legislative Politics: The Big Five Trait Dimensions Among U.S. State Legislators." *Political Psychology* 33(2):195–210.
- Efron, Bradley & Robert J. Tibshirani. 1994. *An Introduction to the Bootstrap*. Boca Raton: CRC Press.
- Gelman, Andrew, Gary King & Chuanhai Liu. 1998. "Not Asked and Not Answered: Multiple Imputation for Multiple Surveys." *Journal of the American Statistical Association* 93(443):846–857.
- Gerber, Alan S., Gregory A. Huber, David Doherty & Conor M. Dowling. 2011. "The Big Five Personality Traits in the Political Arena." *Annual Review of Political Science* 14(1):265–287.
- Gerber, Alan S., Gregory A. Huber, David Doherty, Conor M. Dowling & Shang E. Ha. 2010. "Personality and Political Attitudes: Relationships across Issue Domains and Political Contexts." *American Political Science Review* 104(1):111–133.
- Groseclose, Tim, Steven D. Levitt & James M. Snyder. 1999. "Comparing Interest Group Scores across Time and Chambers: Adjusted ADA Scores for the U.S. Congress." *American Political Science Review* 93(1):33–50.
- Heckman, James J. 1976. "The Common Structure of Statistical Models of Truncation, Sample Selection and Limited Dependent Variables and a Simple Estimator for Such Models." *Annals of Economic and Social Measurement* 5(4):475–492.
- Jost, John T., Brian A. Nosek & Samuel D. Gosling. 2008. "Ideology: Its Resurgence in Social, Personality, and Political Psychology." *Perspectives on Psychological Science* 3(2):126–136.

- King, Gary, James Honaker, Anne Joseph & Kenneth Scheve. 2001. "Analyzing Incomplete Political Science Data: An Alternative Algorithm for Multiple Imputation." *American Political Science Review* 95(1):49–69.
- Lowe, Will & Kenneth Benoit. 2011. "Estimating Uncertainty in Quantitative Text Analysis." Paper prepared for presentation at the 2011 Annual Meeting of the Midwest Political Science Association, March 31-April 3, Chicago, Illinois.
- Mairesse, François & Marilyn A. Walker. 2008. Trainable Generation of Big-Five Personality Styles through Data-Driven Parameter Estimation. In *Proceedings of the 46th Annual Meeting of the Association for Computational Linguistics*. pp. 165–173.
- Mairesse, François, Marilyn A. Walker, Matthias R. Mehl & Roger K. Moore. 2007. "Using Linguistic Cues for the Automatic Recognition of Personality in Conversation and Text." *Journal of Artificial Intelligence Research* 30(1):457–500.
- McCrae, Robert R. & Oliver P. John. 1992. "An Introduction to the Five-Factor Model and Its Applications." *Journal of Personality* 60(2):175–215.
- Mehl, Matthias R., Samuel D. Gosling & James W. Pennebaker. 2006. "Personality in Its Natural Habitat: Manifestations and Implicit Folk Theories of Personality in Daily Life." *Journal of Personality and Social Psychology* 90(5):862–877.
- Mondak, Jeffery J. 2010. *Personality and the Foundations of Political Behavior*. New York: Cambridge University Press.
- Mondak, Jeffery J., Matthew V. Hibbing, Damarys Canache, Mitchell A. Seligson & Mary R. Anderson. 2010. "Personality and Civic Engagement: An Integrative Framework for the Study of Trait Effects on Political Behavior." *American Political Science Review* 104(1):85–110.

- Neale, Thomas H. 1998. "Speechwriting in Perspective: A Brief Guide to Effective and Persuasive Communication." *CRS Report for Congress* 98-170 GOV.
- Ozer, Daniel J. & Veronica Benet-Martinez. 2006. "Personality and the Prediction of Consequential Outcomes." *Annual Review of Psychology* 57(1):401-421.
- Pennebaker, James W. & Laura A. King. 1999. "Linguistic Styles: Language Use as an Individual Difference." *Journal of Personality and Social Psychology* 77(6):1296-1312.
- Pennebaker, James W., Martha E. Francis & Roger J. Booth. 2001. *Linguistic Inquiry and Word Count: LIWC 2001*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Poole, Keith T. & Howard Rosenthal. 1997. *Congress: A Political-Economic History of Roll Call Voting*. New York: Oxford University Press.
- Ramey, Adam J., Jonathan D. Klingler & Gary E. Hollibaugh, Jr. 2015. "More than a Feeling: Personality, Polarization, and the Transformation of the U.S. Congress." Manuscript.
- Rubenzler, Steven J. & Thomas R. Faschingbauer. 2004. *Personality, Character, and Leadership in the White House: Psychologists Assess the Presidents*. Sterling, VA: Potomac Books, Inc.
- Rubin, Donald B. 1976. "Inference and Missing Data." *Biometrika* 63(3):581-592.
- Rubin, Donald B. 1977. "Formalizing Subjective Notions About the Effect of Nonrespondents in Sample Surveys." *Journal of the American Statistical Association* 72(359):538-543.
- Rubin, Donald B. 1987. *Multiple Imputation for Nonresponse in Surveys*. New York: Wiley.
- Schuller, Björn, Stefan Steidl, Anton Batliner, Felix Burkhardt, Laurence Devillers, Christian Müller & Shrikanth Narayanan. 2013. "Paralinguistics in speech and language—state-of-the-art and the challenge." *Computer Speech & Language* 27(1):4-39.



- Shor, Boris & Nolan McCarty. 2011. "The Ideological Mapping of American Legislatures." *American Political Science Review* 105(3):530–551.
- Sigelman, Lee. 2002. "Two Reagans? Genre Imperatives, Ghostwriters, and Presidential Personality Profiles." *Political Psychology* 23(4):839–851.
- Silvester, Jo, Madeleine Wyatt & Ray Randall. 2014. "Politician personality, Machiavellianism, and political skill as predictors of performance ratings in political roles." *Journal of Occupational and Organizational Psychology* 87(2):258–279.
- Tupes, Ernest C. & Raymond E. Christal. 1992. "Recurrent Personality Factors Based on Trait Ratings." *Journal of Personality* 60(2):225–251.
- VandenBos, Gary R. 2007. *APA Dictionary of Psychology*. American Psychological Association.
- Volden, Craig & Alan E. Wiseman. 2014. *Legislative Effectiveness in the United States Congress: The Lawmakers*. New York: Cambridge University Press.
- Weisberg, Herbert F. 2009. *The Total Survey Error Approach: A Guide to the New Science of Survey Research*. Chicago: University of Chicago Press.

# Online Appendix

## LIWC (2001) Categories

Category	Abbreviation	Examples
<i>Standard Linguistic Dimensions</i>		
Word count	wc	
Words/sentence	wps	
Sentences ending in ?	qmarks	
Unique words	unique	
Dictionary words	dic	
Words>6 letters	sixltr	
Total pronouns	pronoun	I, them, itself
1st pers singular	i	I, me, mine
1st pers plural	we	We, us, our
Total 1st person	self	I, we, me
Total 2nd person	you	You, you'll
Total 3rd person	other	She, their, them
Negations	negate	No, not, never
Assents	assent	Yes, OK, mmhmm
Articles	article	A, an, the
Prepositions	prep	To, with, above
Numbers	number	Second, thousand
<i>Personal Concerns</i>		
Occupation	occup	Work, class, boss
School	school	Class, student, college
Job or work	job	Employ, boss, career
Achievement	achieve	Earn, hero, win
Leisure	leisure	Cook, chat, movie
Home	home	Apartment, kitchen, family
Sports	sports	Football, game, play
Television and movies	tv	TV, sitcom, cinema
Music	music	Tunes, song, cd
Money	money	Cash, taxes, income
Metaphysical issues	metaph	God, heaven, coffin
Religion	relig	Altar, church, mosque
Death and dying	death	Bury, coffin, kill
Physical states and functions	physcal	ache, breast, sleep
Body states, symptoms	body	ache, heart, cough
Sex and sexuality	sexual	lust, penis, sex
Eating, drinking, dieting	eating	eat, swallow, taste
Sleeping, dreaming	sleep	asleep, bed, dreams
Grooming	groom	wash, bath, clean

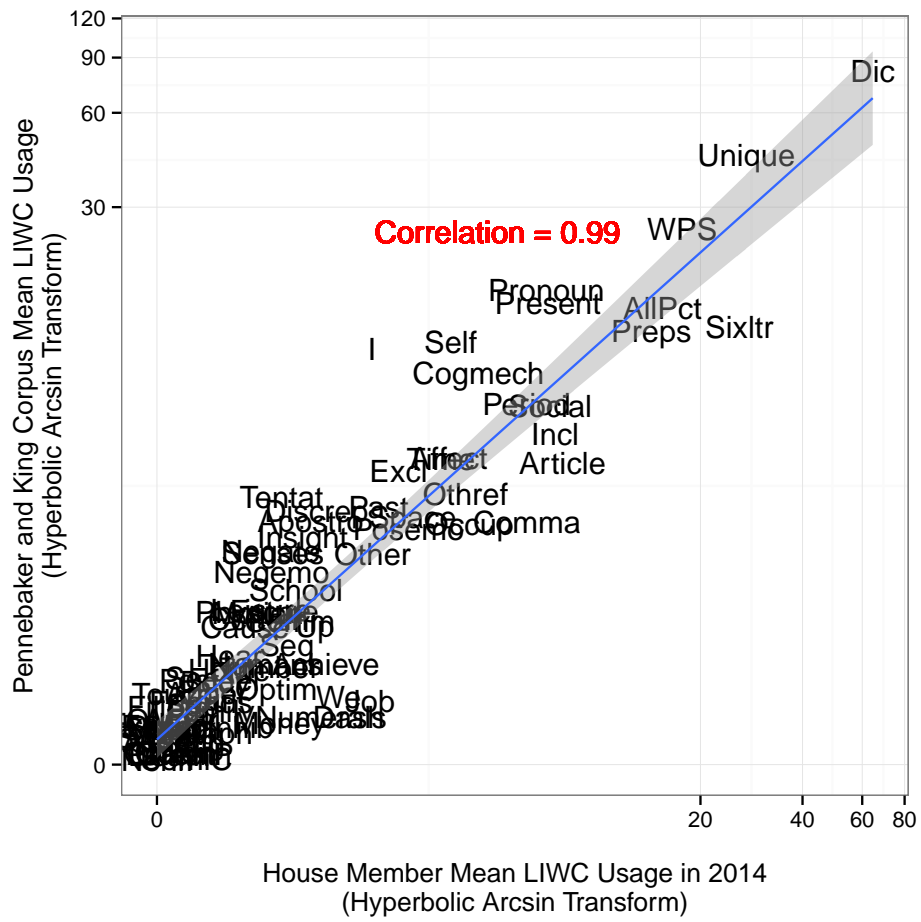
## LIWC (2001) Categories (Continued)

	Category	Abbreviation	Examples
<i>Psychological Processes</i>			
	Affective processes	affect	Happy, cried, abandon
	Positive emotion	posemo	Love, nice, sweet
	Positive feelings	posfeel	Happy, joy, love
	Optimism and energy	optim	Certainty, pride, win
	Negative emotion	negemo	Hurt, ugly, nasty
	Anxiety or fear	anx	Worried, fearful, nervous
	Anger	anger	Hate, kill, annoyed
	Sadness or depression	sad	Crying, grief, sad
	Cognitive processes	cogmech	cause, know, ought
	Causation	cause	because, effect, hence
	Insight	insight	think, know, consider
	Discrepancy	discrep	should, would, could
	Inhibition	inhib	block, constrain, stop
	Tentative	tentat	maybe, perhaps, guess
	Certainty	certain	always, never
	Sensory and perceptual processes	senses	Observing, heard, feeling
	Seeing	see	View, saw, seen
	Hearing	hear	Listen, hearing
	Feeling	feel	Feels, touch
	Social processes	social	Mate, talk, they, child
	Communication	comm	Talk, share, converse
	Other references to people	othref	1st pl, 2nd, 3rd person pronouns
	Family	family	Daughter, husband, aunt
	Friends	friend	Buddy, friend, neighbor
	Humans	human	Adult, baby, boy
<i>Relativity</i>			
	Time	time	Hour, day, o'clock
	Past tense verb	past	Walked, were, had
	Present tense verb	present	Walk, is, be
	Future tense verb	future	Will, might, shall
	Space	space	Around, over, up
	Up	up	Up, above, over
	Down	down	Down, below, under
	Inclusive	incl	And, with, include
	Exclusive	excl	But, without, exclude
	Motion	motion	Arrive, car, go
<i>Experimental Dimensions</i>			
	Swear words	swear	Damn, piss, fuck
	Nonfluencies	nonflu	Er, hm, umm
	Fillers	fillers	Blah, I mean, you know

## MRCPD Categories

Category	Abbreviation
Age of Acquisition	AOA
Brown Frequency	BROWN_FREQ
Concreteness	CONC
Familiarity	FAM
Imagability	IMAG
K F Frequency	K_F_FREQ
K F Number of Categories	K_F_NCATS
K F Number of Samples	K_F_NSAMP
Meaningfulness (Colorado)	MEANC
Meaningfulness (Paivo)	MEANP
Number of letters	NLET
Number of phonemes	NPHON
Number of syllables	NSYL
T L Frequency	T_L_FREQ

# LIWC Usage Comparison with Labels



## Feature Analysis

Table A-1: Correlations Between LIWC/MRC Features and Jackknifed Big Five Traits (1996)

	Openness	Conscientiousness	Extraversion	Agreeableness	Emotional Stability
<b>MRCPD Categories</b>					
Age of Acquisition	0.227***	0.173***	-0.159***	0.189***	-0.082
Concreteness	-0.113**	-0.287***	-0.191***	-0.150***	-0.040
Familiarity	-0.243***	0.051	0.368***	-0.066	0.061
Imagability	-0.169***	-0.368***	-0.159***	-0.227***	-0.073
Meaningfulness (Colorado)	-0.373***	-0.284***	0.160***	-0.306***	-0.080
Meaningfulness (Paivo)	-0.215***	-0.297***	-0.060	-0.236***	0.001
Number of Letters	0.240***	-0.034	-0.386***	0.077	-0.063
Number of Phonemes	0.263***	0.006	-0.352***	0.113**	-0.046
Number of Syllables	0.290***	-0.003	-0.372***	0.123**	-0.056
<b>LIWC Categories</b>					
Word Count	0.239***	0.257***	0.388***	0.142***	0.203***
Words Per Sentence	0.083	0.076	0.036	0.055	0.054
Sentences Ending With ?	-0.119**	0.089*	0.212***	-0.033	0.071
Unique Words	-0.057	-0.225***	-0.328***	-0.083	-0.181***
Dictionary Words	-0.247***	0.082	0.434***	-0.065	0.027
Words > 6 Letters	0.255***	-0.027	-0.404***	0.093*	-0.072
Abbreviations	0.168***	-0.025	-0.212***	-0.020	-0.036
Total Pronouns	-0.261***	-0.038	0.318***	-0.121**	-0.042
First-Person Singular	-0.053	0.001	0.055	0.010	-0.082
First-Person Plural	-0.249***	0.107**	0.409***	-0.091*	-0.076
Total First Person	-0.187***	0.064	0.283***	-0.046	-0.106**
Total Second Person	-0.187***	-0.005	0.171***	-0.076	0.068
Total Third Person	-0.168***	-0.311***	0.060	-0.179***	0.063
Negations	-0.025	0.190***	0.253***	-0.000	-0.004
Assents	0.064	0.079	-0.013	0.141***	-0.000
Articles	0.233***	0.166***	-0.087	0.208***	0.135**
Prepositions	0.169***	0.113**	-0.028	0.184***	0.056
Numbers	0.023	0.053	-0.094*	0.059	0.046
Affective Processes	-0.005	-0.286***	-0.100*	-0.228***	-0.162***
Positive Feelings	0.013	-0.066	0.012	-0.032	-0.056
Optimism and Energy	0.027	-0.253***	-0.172***	-0.120**	-0.069
Anxiety or Fear	-0.007	-0.100*	-0.022	-0.159***	-0.120**
Anger	0.003	-0.143***	0.005	-0.206***	-0.061
Sadness or Depression	-0.086	-0.232***	-0.025	-0.236***	-0.053
Cognitive Processes	-0.061	0.275***	0.264***	0.110**	-0.003
Causation	-0.083	0.070	0.051	-0.007	-0.037
Insight	-0.065	0.159***	0.136**	0.116**	-0.009
Discrepancy	-0.112**	0.126**	0.310***	-0.032	-0.019
Inhibition	0.227***	0.247***	-0.138***	0.262***	0.005
Tentative	0.082	0.294***	0.177***	0.152***	0.070
Certainty	-0.077	0.053	0.214***	-0.041	-0.049
Sensory and Perceptual Processes	-0.165***	0.034	0.238***	-0.101*	0.045

Continued on next page

Continued from previous page					
	Openness	Conscientiousness	Extraversion	Agreeableness	Emotional Stability
<b>LIWC Categories (continued)</b>					
Seeing	-0.124**	0.040	0.183***	-0.084	0.057
Hearing	-0.153***	0.052	0.228***	-0.069	0.045
Feeling	-0.064	-0.129**	0.017	-0.137***	-0.040
Social Processes	-0.273***	-0.140***	0.289***	-0.162***	0.043
Communication	-0.120**	0.090*	0.145***	0.007	0.036
Other References to People	-0.304***	-0.105**	0.338***	-0.178***	0.009
Humans	-0.090*	-0.048	0.183***	0.002	0.121**
Time	-0.048	-0.050	0.025	-0.056	-0.034
Past Tense Verb	-0.056	-0.070	0.079	-0.069	0.061
Present Tense Verb	-0.270***	0.092*	0.373***	-0.075	-0.010
Future Tense Verb	0.154***	0.306***	0.028	0.257***	-0.014
Space	-0.157***	0.068	0.303***	-0.091*	0.093*
Up	-0.036	0.041	0.142***	-0.070	0.018
Down	0.014	0.084	0.067	0.009	0.018
Inclusive	-0.025	-0.140***	0.016	-0.093*	-0.137***
Exclusive	-0.099*	0.248***	0.363***	0.085	0.104**
Motion	-0.186***	0.112**	0.320***	-0.051	0.087
Achievement	-0.115**	-0.125**	-0.025	-0.057	0.021
Religion	-0.115**	-0.288***	-0.043	-0.118**	0.090*
Sex and Sexuality	-0.124**	-0.106**	0.065	-0.070	0.101*
Swear Words	-0.043	-0.075	-0.036	-0.112**	-0.024
Nonfluencies	0.074	0.047	-0.006	0.024	0.020
Fillers	0.009	0.080	-0.018	0.051	-0.021
Number of Periods	0.066	0.011	-0.138***	0.042	-0.062
Number of Commas	-0.077	-0.092*	0.063	-0.075	0.048
Number of Colons	0.112**	0.064	-0.064	0.083	0.069
Number of Semicolons	0.025	0.010	-0.097*	0.024	-0.031
Number of Question Marks	-0.118**	0.067	0.185***	-0.038	0.059
Number of Exclamations	-0.036	-0.098*	-0.106**	-0.023	0.054
Number of Dashes	0.137***	-0.012	-0.227***	0.021	-0.026
Number of Quotation Marks	-0.020	-0.013	-0.025	-0.007	-0.035
Number of Apostrophes	0.042	-0.147***	-0.187***	-0.113**	-0.056
Number of Pairs of Parentheses	0.141***	0.148***	-0.110**	0.147***	0.016
Number of Other Punctuation	0.125**	0.136**	-0.094*	0.122**	0.013
All Punctuation	0.118**	0.053	-0.167***	0.070	-0.010

Note: Pearson correlation coefficients presented. Two-tailed tests: \*\*\*  $p < 0.01$ , \*\*  $p < 0.05$ , \*  $p < 0.1$

Table A-2: Correlations Between LIWC/MRC Features and Jackknifed Big Five Traits (1997)

	Openness	Conscientiousness	Extraversion	Agreeableness	Emotional Stability
<b>MRCPD Categories</b>					
Age of Acquisition	0.167***	0.263***	-0.149***	0.158***	-0.110**
Concreteness	-0.064	-0.318***	-0.150***	-0.031	0.052
Familiarity	-0.288***	-0.064	0.294***	-0.145***	0.011
Imagability	-0.129***	-0.418***	-0.114**	-0.139***	0.013
Meaningfulness (Colorado)	-0.349***	-0.349***	0.176***	-0.289***	-0.069
Meaningfulness (Paivo)	-0.033	-0.248***	-0.084*	-0.080*	0.087*
Number of Letters	0.310***	0.035	-0.344***	0.133***	0.000
Number of Phonemes	0.322***	0.080*	-0.319***	0.186***	0.028
Number of Syllables	0.348***	0.053	-0.349***	0.172***	0.022
<b>LIWC Categories</b>					
Word Count	0.143***	0.227***	0.476***	0.037	0.181***
Words Per Sentence	0.006	0.127***	0.135***	-0.001	0.067
Sentences Ending With ?	-0.111**	0.070	0.202***	-0.101**	0.063
Unique Words	0.041	-0.192***	-0.462***	0.098**	-0.076
Dictionary Words	-0.296***	0.023	0.354***	-0.161***	-0.062
Words > 6 Letters	0.314***	0.032	-0.362***	0.145***	-0.012
Abbreviations	0.145***	0.012	-0.254***	0.044	-0.021
Total Pronouns	-0.317***	-0.151***	0.262***	-0.199***	-0.067
First-Person Singular	-0.095**	-0.026	-0.044	0.057	-0.096**
First-Person Plural	-0.301***	-0.002	0.332***	-0.241***	-0.149***
Total First Person	-0.259***	-0.017	0.196***	-0.126***	-0.158***
Total Second Person	-0.203***	-0.064	0.170***	-0.125***	0.077
Total Third Person	-0.128***	-0.330***	0.092*	-0.125***	0.122**
Negations	-0.123**	0.138***	0.265***	-0.145***	-0.097**
Assents	-0.006	0.126***	0.104**	0.038	-0.066
Articles	0.280***	0.221***	-0.146***	0.287***	0.138***
Prepositions	0.253***	0.166***	-0.069	0.251***	0.113**
Numbers	0.005	0.065	-0.121**	0.097**	0.094*
Affective Processes	0.027	-0.258***	-0.072	-0.213***	-0.179***
Positive Feelings	-0.057	-0.075	0.108**	-0.095**	-0.102**
Optimism and Energy	-0.008	-0.228***	-0.123**	-0.092*	-0.068
Anxiety or Fear	0.094*	-0.004	0.005	-0.144***	-0.154***
Anger	0.092*	-0.108**	-0.074	-0.169***	-0.108**
Sadness or Depression	0.037	-0.088*	0.007	-0.120**	-0.057
Cognitive Processes	-0.128***	0.312***	0.251***	-0.021	-0.161***
Causation	-0.096**	0.194***	0.133***	-0.092*	-0.074
Insight	-0.081*	0.215***	0.092*	0.031	-0.159***
Discrepancy	-0.196***	0.165***	0.283***	-0.099**	-0.154***
Inhibition	0.203***	0.249***	-0.121**	0.212***	-0.034
Tentative	-0.012	0.282***	0.250***	0.017	-0.057
Certainty	-0.059	0.029	0.173***	-0.095*	-0.067
Sensory and Perceptual Processes	-0.185***	0.026	0.231***	-0.091*	0.059
Seeing	-0.011	0.054	0.187***	-0.017	0.067
Hearing	-0.228***	0.044	0.218***	-0.088*	0.042
Feeling	-0.004	-0.169***	-0.067	-0.090*	0.012
Social Processes	-0.279***	-0.209***	0.270***	-0.192***	0.022

Continued on next page



Continued from previous page					
	Openness	Conscientiousness	Extraversion	Agreeableness	Emotional Stability
<b>LIWC Categories (continued)</b>					
Communication	-0.133***	0.085*	0.129***	-0.021	0.033
Other References to People	-0.318***	-0.220***	0.310***	-0.261***	-0.007
Humans	-0.075	-0.007	0.204***	0.044	0.026
Time	-0.034	-0.116**	-0.086*	-0.023	-0.027
Past Tense Verb	-0.134***	-0.190***	0.108**	-0.149***	0.089*
Present Tense Verb	-0.344***	0.064	0.347***	-0.198***	-0.112**
Future Tense Verb	0.139***	0.323***	0.142***	0.157***	-0.059
Space	-0.176***	0.115**	0.190***	-0.068	0.006
Up	-0.056	0.014	-0.005	0.008	0.038
Down	0.092*	0.289***	0.100**	0.103**	0.035
Inclusive	-0.066	-0.163***	-0.002	-0.099**	-0.064
Exclusive	-0.170***	0.272***	0.361***	-0.032	-0.046
Motion	-0.281***	-0.027	0.329***	-0.176***	0.046
Achievement	-0.115**	-0.216***	-0.063	-0.060	-0.052
Religion	-0.087*	-0.291***	-0.083*	-0.104**	0.121**
Sex and Sexuality	-0.055	-0.113**	-0.010	0.008	0.089*
Swear Words	-0.006	0.022	-0.008	0.008	-0.025
Nonfluencies	0.024	0.010	0.087*	-0.021	0.038
Fillers	-0.004	0.077	-0.047	0.005	-0.026
Number of Periods	0.064	-0.085*	-0.213***	0.055	-0.068
Number of Commas	-0.056	-0.149***	0.022	0.009	0.210***
Number of Colons	0.100**	0.032	0.058	0.005	0.055
Number of Semicolons	0.046	0.027	0.006	0.059	0.111**
Number of Question Marks	-0.103**	0.060	0.169***	-0.090*	0.043
Number of Exclamations	-0.002	-0.105**	-0.081*	-0.042	-0.009
Number of Dashes	0.179***	0.018	-0.078	-0.002	-0.012
Number of Quotation Marks	-0.043	0.031	0.047	-0.037	-0.016
Number of Apostrophes	0.103**	-0.181***	-0.222***	-0.015	0.090*
Number of Pairs of Parentheses	0.103**	0.175***	-0.011	0.151***	0.019
Number of Other Punctuation	0.082*	0.149***	-0.023	0.136***	0.063
All Punctuation	0.110**	-0.000	-0.122**	0.110**	0.096**

Note: Pearson correlation coefficients presented. Two-tailed tests: \*\*\*  $p < 0.01$ , \*\*  $p < 0.05$ , \*  $p < 0.1$

Table A-3: Correlations Between LIWC/MRC Features and Jackknifed Big Five Traits (1998)

	Openness	Conscientiousness	Extraversion	Agreeableness	Emotional Stability
<b>MRCPD Categories</b>					
Age of Acquisition	0.202***	0.281***	-0.154***	0.219***	-0.049
Concreteness	-0.083*	-0.381***	-0.264***	-0.119**	-0.019
Familiarity	-0.259***	-0.022	0.335***	-0.110**	0.007
Imagability	-0.141***	-0.454***	-0.227***	-0.183***	-0.041
Meaningfulness (Colorado)	-0.361***	-0.395***	0.112**	-0.295***	-0.097**
Meaningfulness (Paivo)	-0.134***	-0.230***	0.011	-0.118**	0.066
Number of Letters	0.255***	-0.030	-0.327***	0.070	0.071
Number of Phonemes	0.282***	0.017	-0.328***	0.136***	0.075
Number of Syllables	0.294***	-0.002	-0.350***	0.122**	0.057
<b>LIWC Categories</b>					
Word Count	0.261***	0.293***	0.497***	0.140***	0.231***
Words Per Sentence	-0.005	0.087*	0.249***	-0.008	0.116**
Sentences Ending With ?	-0.135***	0.168***	0.305***	-0.024	0.051
Unique Words	-0.088*	-0.286***	-0.484***	-0.039	-0.206***
Dictionary Words	-0.286***	0.054	0.408***	-0.063	-0.034
Words > 6 Letters	0.261***	-0.031	-0.364***	0.083*	0.047
Abbreviations	0.200***	0.035	-0.279***	0.086*	-0.031
Total Pronouns	-0.288***	-0.122**	0.264***	-0.156***	-0.119**
First-Person Singular	-0.015	0.002	-0.066	0.016	-0.203***
First-Person Plural	-0.255***	0.036	0.368***	-0.137***	-0.143***
Total First Person	-0.180***	0.025	0.203***	-0.081*	-0.227***
Total Second Person	-0.276***	-0.128***	0.192***	-0.168***	-0.005
Total Third Person	-0.160***	-0.375***	0.039	-0.135***	0.146***
Negations	-0.098**	0.229***	0.275***	0.002	-0.059
Assents	0.109**	0.085*	-0.125***	0.115**	-0.071
Articles	0.197***	0.266***	-0.101**	0.283***	0.127***
Prepositions	0.182***	0.137***	-0.036	0.205***	0.029
Numbers	-0.000	0.007	-0.117**	0.092*	0.054
Affective Processes	-0.060	-0.312***	-0.087*	-0.232***	-0.151***
Positive Feelings	-0.024	-0.147***	0.026	-0.096**	-0.081*
Optimism and Energy	-0.109**	-0.236***	-0.040	-0.125***	-0.002
Anxiety or Fear	0.104**	-0.011	0.026	-0.114**	-0.082*
Anger	0.041	-0.081*	0.065	-0.189***	-0.081*
Sadness or Depression	0.066	-0.082*	-0.073	-0.058	-0.059
Cognitive Processes	-0.068	0.324***	0.307***	0.072	-0.053
Causation	-0.135***	0.133***	0.207***	-0.063	0.022
Insight	-0.057	0.199***	0.111**	0.084*	-0.069
Discrepancy	-0.149***	0.184***	0.349***	-0.032	-0.046
Inhibition	0.289***	0.374***	-0.074	0.248***	-0.048
Tentative	0.047	0.336***	0.265***	0.130***	0.031
Certainty	0.006	0.124***	0.155***	0.013	-0.037
Sensory and Perceptual Processes	-0.198***	0.069	0.303***	-0.068	0.042
Seeing	-0.145***	-0.007	0.249***	-0.101**	0.050
Hearing	-0.170***	0.122**	0.296***	-0.030	0.038
Feeling	-0.104**	-0.135***	-0.052	-0.065	-0.027
Social Processes	-0.312***	-0.250***	0.289***	-0.177***	0.011

Continued on next page

Continued from previous page					
	Openness	Conscientiousness	Extraversion	Agreeableness	Emotional Stability
<b>LIWC Categories (continued)</b>					
Communication	-0.083*	0.170***	0.219***	0.025	0.044
Other References to People	-0.343***	-0.232***	0.328***	-0.214***	-0.006
Humans	-0.074	-0.066	0.105**	0.063	0.013
Time	-0.072	-0.179***	-0.083*	-0.054	-0.043
Past Tense Verb	-0.099**	-0.155***	0.067	-0.094*	0.118**
Present Tense Verb	-0.309***	0.096**	0.398***	-0.099**	-0.072
Future Tense Verb	0.112**	0.331***	0.079	0.222***	-0.067
Space	-0.142***	0.180***	0.265***	0.021	0.041
Up	-0.013	0.143***	0.063	0.110**	-0.010
Down	-0.010	0.236***	0.062	0.131***	0.060
Inclusive	-0.007	-0.203***	0.025	-0.130***	-0.029
Exclusive	-0.102**	0.320***	0.420***	0.064	0.050
Motion	-0.170***	0.084*	0.310***	-0.019	0.068
Achievement	-0.137***	-0.278***	-0.122**	-0.098**	-0.007
Religion	-0.041	-0.184***	-0.008	-0.014	0.111**
Sex and Sexuality	-0.075	-0.085*	-0.063	0.006	0.020
Swear Words	0.036	0.029	0.023	-0.035	-0.020
Nonfluencies	0.110**	0.099**	0.188***	0.006	0.063
Fillers	0.006	-0.055	-0.015	-0.020	0.033
Number of Periods	0.098**	-0.038	-0.327***	0.076	-0.112**
Number of Commas	-0.013	-0.109**	-0.036	0.001	0.097**
Number of Colons	-0.002	-0.010	-0.035	0.021	0.012
Number of Semicolons	0.046	0.015	-0.033	0.018	-0.020
Number of Question Marks	-0.129***	0.155***	0.263***	-0.022	0.027
Number of Exclamations	-0.057	-0.130***	-0.076	-0.095**	-0.008
Number of Dashes	0.154***	-0.020	-0.188***	0.021	-0.104**
Number of Quotation Marks	-0.095**	0.089*	0.155***	0.006	0.023
Number of Apostrophes	0.088*	-0.174***	-0.230***	-0.073	0.049
Number of Pairs of Parentheses	0.163***	0.227***	-0.107**	0.219***	0.017
Number of Other Punctuation	0.100**	0.140***	-0.104**	0.171***	0.085*
All Punctuation	0.135***	0.005	-0.250***	0.112**	0.013

Note: Pearson correlation coefficients presented. Two-tailed tests: \*\*\*  $p < 0.01$ , \*\*  $p < 0.05$ , \*  $p < 0.1$

Table A4: Correlations Between LIWC/MRC Features and Jackknifed Big Five Traits (1999)

	Openness	Conscientiousness	Extraversion	Agreeableness	Emotional Stability
<b>MRCPD Categories</b>					
Age of Acquisition	0.266***	0.342***	-0.135***	0.250***	-0.047
Concreteness	-0.135***	-0.383***	-0.248***	-0.032	0.096**
Familiarity	-0.245***	-0.014	0.319***	-0.099**	0.014
Imagability	-0.215***	-0.480***	-0.228***	-0.135***	0.035
Meaningfulness (Colorado)	-0.428***	-0.412***	0.084*	-0.283***	-0.088*
Meaningfulness (Paivo)	-0.184***	-0.363***	-0.028	-0.135***	0.147***
Number of Letters	0.313***	-0.002	-0.329***	0.109**	0.002
Number of Phonemes	0.333***	0.055	-0.328***	0.179***	0.037
Number of Syllables	0.361***	0.056	-0.328***	0.177***	0.033
<b>LIWC Categories</b>					
Word Count	0.272***	0.291***	0.501***	0.105**	0.250***
Words Per Sentence	0.087*	0.117**	0.198***	-0.014	0.059
Sentences Ending With ?	-0.101**	0.171***	0.349***	-0.058	0.130***
Unique Words	-0.078	-0.290***	-0.489***	0.025	-0.132***
Dictionary Words	-0.305***	0.059	0.402***	-0.131***	-0.061
Words > 6 Letters	0.317***	-0.015	-0.368***	0.130***	0.004
Abbreviations	0.162***	0.055	-0.279***	0.098**	-0.025
Total Pronouns	-0.369***	-0.119**	0.247***	-0.165***	-0.094*
First-Person Singular	-0.147***	-0.050	-0.022	0.003	-0.125***
First-Person Plural	-0.298***	0.014	0.299***	-0.189***	-0.192***
Total First Person	-0.297***	-0.020	0.196**	-0.131***	-0.210***
Total Second Person	-0.202***	-0.047	0.211**	-0.025	0.160***
Total Third Person	-0.240***	-0.346***	0.063	-0.163***	0.108**
Negations	-0.046	0.262***	0.281***	-0.033	-0.067
Assents	0.073	0.103**	-0.125***	0.154***	0.011
Articles	0.389***	0.355***	-0.020	0.335***	0.184***
Prepositions	0.263***	0.167***	-0.079	0.252***	0.104**
Numbers	-0.080*	-0.005	-0.119**	0.003	0.023
Affective Processes	-0.095**	-0.276***	0.004	-0.263***	-0.199***
Positive Feelings	-0.035	-0.133***	0.032	-0.094*	-0.049
Optimism and Energy	-0.178***	-0.174***	-0.031	-0.160***	-0.116**
Anxiety or Fear	0.129***	-0.020	0.046	-0.129***	-0.048
Anger	0.072	-0.066	0.115**	-0.155***	0.005
Sadness or Depression	-0.011	-0.165***	-0.024	-0.175***	-0.125***
Cognitive Processes	-0.024	0.358***	0.358***	0.006	-0.085*
Causation	-0.017	0.181***	0.153***	-0.009	0.010
Insight	0.005	0.225***	0.216***	0.041	-0.013
Discrepancy	-0.178***	0.178***	0.359***	-0.110**	-0.098**
Inhibition	0.251***	0.323***	-0.051	0.179***	-0.057
Tentative	0.079	0.345***	0.236***	0.088*	0.038
Certainty	-0.026	0.085*	0.145***	-0.053	-0.096**
Sensory and Perceptual Processes	-0.197***	0.065	0.338***	-0.091*	0.049
Seeing	-0.191***	-0.060	0.304***	-0.159***	0.012
Hearing	-0.159***	0.135***	0.288***	-0.020	0.076
Feeling	-0.037	-0.108**	0.030	-0.120**	-0.097**
Social Processes	-0.362***	-0.230***	0.281***	-0.216***	-0.018

Continued on next page

Continued from previous page					
	Openness	Conscientiousness	Extraversion	Agreeableness	Emotional Stability
<b>LIWC Categories (continued)</b>					
Communication	-0.067	0.190***	0.257***	0.028	0.066
Other References to People	-0.385***	-0.187***	0.292***	-0.232***	-0.050
Humans	-0.134***	-0.152***	0.157***	-0.075	0.001
Time	-0.073	-0.141***	-0.021	-0.097**	-0.043
Past Tense Verb	-0.047	-0.129***	0.097**	-0.033	0.246***
Present Tense Verb	-0.345***	0.047	0.376***	-0.189***	-0.134***
Future Tense Verb	0.139***	0.349***	0.053	0.103**	-0.136***
Space	-0.141***	0.102**	0.173***	-0.028	0.053
Up	-0.061	0.113**	0.001	0.028	-0.018
Down	-0.075	0.121**	0.096**	-0.025	0.010
Inclusive	-0.040	-0.194***	-0.045	-0.034	-0.036
Exclusive	-0.106**	0.308***	0.419***	-0.008	0.005
Motion	-0.125***	0.147***	0.294***	-0.034	0.090*
Achievement	-0.141***	-0.280***	-0.168***	-0.062	-0.092*
Religion	-0.005	-0.154***	0.012	0.010	0.123**
Sex and Sexuality	-0.080	-0.115**	0.033	0.047	0.168***
Swear Words	-0.016	-0.083*	0.079	-0.081*	0.090*
Nonfluencies	0.192***	0.114**	0.154***	0.022	0.091*
Fillers	-0.001	-0.026	0.023	-0.062	-0.050
Number of Periods	-0.061	-0.012	-0.201***	0.022	-0.115**
Number of Commas	-0.032	-0.030	0.042	0.046	0.144***
Number of Colons	0.029	0.040	0.041	0.001	0.019
Number of Semicolons	-0.034	0.016	0.047	-0.007	0.047
Number of Question Marks	-0.112**	0.159***	0.313***	-0.054	0.108**
Number of Exclamations	-0.021	-0.142***	-0.087*	-0.071	0.050
Number of Dashes	0.169***	-0.026	-0.205***	0.020	-0.071
Number of Quotation Marks	0.019	0.019	0.094*	-0.001	0.136***
Number of Apostrophes	0.066	-0.231***	-0.216***	0.016	0.158***
Number of Pairs of Parentheses	0.140***	0.200***	-0.038	0.138***	0.030
Number of Other Punctuation	0.096**	0.115**	-0.058	0.114**	0.097**
All Punctuation	0.054	0.027	-0.170***	0.098**	0.056

Note: Pearson correlation coefficients presented. Two-tailed tests: \*\*\*  $p < 0.01$ , \*\*  $p < 0.05$ , \*  $p < 0.1$

Table A-5: Correlations Between LIWC/MRC Features and Jackknifed Big Five Traits (2000)

	Openness	Conscientiousness	Extraversion	Agreeableness	Emotional Stability
<b>MRCPD Categories</b>					
Age of Acquisition	0.257***	0.315***	-0.151***	0.261***	-0.026
Concreteness	-0.125***	-0.381***	-0.319***	-0.081*	0.061
Familiarity	-0.169***	0.030	0.363***	-0.069	-0.018
Imagability	-0.219***	-0.500***	-0.287***	-0.207***	-0.019
Meaningfulness (Colorado)	-0.442***	-0.445***	0.055	-0.335***	-0.124***
Meaningfulness (Paivo)	-0.201***	-0.296***	0.019	-0.192***	0.030
Number of Letters	0.271***	-0.025	-0.273***	0.079	0.021
Number of Phonemes	0.293***	0.024	-0.296***	0.164***	0.053
Number of Syllables	0.303***	0.000	-0.288***	0.127***	0.026
<b>LIWC Categories</b>					
Word Count	0.245***	0.288***	0.528***	0.132***	0.237***
Words Per Sentence	0.052	0.100**	0.221***	0.016	0.102**
Sentences Ending With ?	-0.049	0.179***	0.334***	-0.053	0.064
Unique Words	-0.076	-0.356***	-0.554***	-0.065	-0.171***
Dictionary Words	-0.259***	0.092*	0.425***	-0.086*	-0.051
Words > 6 Letters	0.272***	-0.034	-0.335***	0.101**	0.008
Abbreviations	0.162***	0.008	-0.284***	0.064	-0.036
Total Pronouns	-0.360***	-0.115**	0.283***	-0.188***	-0.140***
First-Person Singular	-0.042	-0.014	-0.169***	0.065	-0.159***
First-Person Plural	-0.290***	0.046	0.370***	-0.210***	-0.223***
Total First Person	-0.236***	0.024	0.162***	-0.112**	-0.261***
Total Second Person	-0.165***	-0.004	0.264***	-0.027	0.116**
Total Third Person	-0.275***	-0.401***	0.099**	-0.185***	0.114**
Negations	-0.099**	0.166***	0.229***	-0.082*	-0.096**
Assents	0.109**	0.045	-0.267***	0.090*	-0.141***
Articles	0.398***	0.368***	-0.037	0.369***	0.202***
Prepositions	0.278***	0.100**	-0.091*	0.233***	0.152***
Numbers	-0.014	-0.010	-0.250***	0.040	0.009
Affective Processes	-0.010	-0.261***	-0.084*	-0.249***	-0.234***
Positive Feelings	-0.073	-0.066	0.020	-0.074	-0.089*
Optimism and Energy	-0.066	-0.204***	-0.125***	-0.140***	-0.152***
Anxiety or Fear	0.154***	-0.024	0.004	-0.132***	-0.111**
Anger	0.129***	-0.067	0.051	-0.194***	-0.078
Sadness or Depression	-0.072	-0.142***	-0.021	-0.150***	-0.024
Cognitive Processes	0.003	0.406***	0.389***	0.098**	-0.027
Causation	-0.041	0.252***	0.286***	0.009	0.028
Insight	-0.037	0.279***	0.242***	0.121**	0.041
Discrepancy	-0.122**	0.203***	0.363***	-0.062	-0.087*
Inhibition	0.360***	0.387***	-0.054	0.254***	-0.001
Tentative	0.000	0.304***	0.336***	0.055	0.082*
Certainty	0.014	0.147***	0.175***	-0.009	-0.098**
Sensory and Perceptual Processes	-0.135***	0.112**	0.360***	-0.039	0.125***
Seeing	-0.114**	0.018	0.213***	-0.053	0.076
Hearing	-0.104**	0.153***	0.331***	-0.000	0.125***
Feeling	-0.077	-0.095**	0.112**	-0.123**	-0.011
Social Processes	-0.353***	-0.262***	0.315***	-0.205***	0.030

Continued on next page

Continued from previous page					
	Openness	Conscientiousness	Extraversion	Agreeableness	Emotional Stability
<b>LIWC Categories (continued)</b>					
Communication	-0.003	0.206***	0.294***	0.065	0.147***
Other References to People	-0.397***	-0.212***	0.361***	-0.260***	-0.063
Humans	-0.071	-0.130***	0.105**	0.061	0.125***
Time	-0.045	-0.164***	-0.170***	-0.069	-0.018
Past Tense Verb	-0.096**	-0.148***	0.020	-0.086*	0.129***
Present Tense Verb	-0.301***	0.116**	0.436***	-0.134***	-0.105**
Future Tense Verb	0.197***	0.383***	0.064	0.218***	-0.011
Space	-0.159***	0.119**	0.175***	-0.062	-0.022
Up	-0.092*	0.012	-0.130***	0.016	-0.033
Down	-0.052	0.208***	0.119**	-0.005	-0.033
Inclusive	-0.066	-0.253***	-0.015	-0.094*	-0.053
Exclusive	-0.057	0.340***	0.456***	0.055	0.054
Motion	-0.152***	0.123**	0.309***	-0.082*	-0.011
Achievement	-0.144***	-0.272***	-0.217***	-0.038	-0.062
Religion	-0.060	-0.302***	-0.051	-0.102**	0.174***
Sex and Sexuality	-0.080*	-0.076	-0.046	0.028	0.131***
Swear Words	-0.034	-0.047	0.026	-0.053	-0.006
Nonfluencies	0.183***	0.160***	0.117**	0.092*	0.089*
Fillers	NA	NA	NA	NA	NA
Number of Periods	0.044	-0.031	-0.274***	0.025	-0.121**
Number of Commas	-0.091*	-0.069	-0.031	0.019	0.094*
Number of Colons	0.039	0.016	-0.069	-0.014	-0.033
Number of Semicolons	0.033	-0.016	-0.042	0.056	0.013
Number of Question Marks	-0.047	0.156***	0.293***	-0.055	0.044
Number of Exclamations	-0.061	-0.107**	-0.056	-0.106**	-0.088*
Number of Dashes	0.130***	-0.064	-0.212***	-0.041	-0.050
Number of Quotation Marks	0.026	-0.062	-0.004	0.023	0.063
Number of Apostrophes	0.058	-0.282***	-0.330***	-0.061	0.031
Number of Pairs of Parentheses	0.150***	0.167***	-0.131***	0.176***	0.059
Number of Other Punctuation	0.116**	0.089*	-0.199***	0.136***	0.052
All Punctuation	0.097**	-0.017	-0.304***	0.081*	-0.006

Note: Pearson correlation coefficients presented. Two-tailed tests: \*\*\*  $p < 0.01$ , \*\*  $p < 0.05$ , \*  $p < 0.1$

Table A-6: Correlations Between LIWC/MRC Features and Jackknifed Big Five Traits (2001)

	Openness	Conscientiousness	Extraversion	Agreeableness	Emotional Stability
<b>MRCPD Categories</b>					
Age of Acquisition	0.262***	0.330***	-0.052	0.237***	-0.094**
Concreteness	-0.061	-0.382***	-0.275***	-0.115**	0.057
Familiarity	-0.274***	0.014	0.309***	-0.042	0.059
Imagability	-0.136***	-0.505***	-0.278***	-0.239***	-0.033
Meaningfulness (Colorado)	-0.378***	-0.484***	-0.013	-0.385***	-0.137***
Meaningfulness (Paivo)	-0.104**	-0.316***	0.004	-0.192***	0.113**
Number of Letters	0.302***	0.017	-0.284***	0.083*	-0.064
Number of Phonemes	0.334***	0.072	-0.274***	0.180***	-0.011
Number of Syllables	0.339***	0.068	-0.257***	0.149***	-0.030
<b>LIWC Categories</b>					
Word Count	0.227***	0.220***	0.525***	0.108**	0.275***
Words Per Sentence	0.065	0.108**	0.183***	0.023	0.082*
Sentences Ending With ?	-0.016	0.210***	0.256***	0.066	0.154***
Unique Words	-0.068	-0.238***	-0.489***	0.003	-0.184***
Dictionary Words	-0.258***	0.023	0.297***	-0.101**	-0.035
Words > 6 Letters	0.292***	0.027	-0.317***	0.125***	-0.060
Abbreviations	0.177***	0.063	-0.219***	0.136***	-0.015
Total Pronouns	-0.324***	-0.178***	0.188***	-0.226***	-0.130***
First-Person Singular	-0.039	-0.009	-0.117**	0.034	-0.126***
First-Person Plural	-0.292***	-0.047	0.278***	-0.264***	-0.224***
Total First Person	-0.234***	-0.040	0.124***	-0.167***	-0.240***
Total Second Person	-0.177***	-0.061	0.211***	-0.143***	0.027
Total Third Person	-0.202***	-0.403***	-0.044	-0.140***	0.101**
Negations	-0.074	0.174***	0.234***	-0.128***	-0.138***
Assents	-0.004	0.043	-0.149***	0.048	-0.053
Articles	0.270***	0.354***	0.026	0.397***	0.263***
Prepositions	0.246***	0.145***	-0.087*	0.317***	0.164***
Numbers	0.019	0.113**	-0.034	0.053	0.003
Affective Processes	-0.033	-0.304***	-0.115**	-0.303***	-0.292***
Positive Feelings	-0.109**	-0.244***	-0.079*	-0.184***	-0.147***
Optimism and Energy	-0.046	-0.272***	-0.119**	-0.222***	-0.181***
Anxiety or Fear	0.140***	0.012	-0.033	-0.050	-0.042
Anger	0.157***	-0.019	-0.006	-0.097**	-0.069
Sadness or Depression	-0.070	-0.168***	-0.069	-0.242***	-0.191***
Cognitive Processes	-0.074	0.349***	0.274***	0.065	-0.080*
Causation	-0.090*	0.183***	0.226***	-0.085*	-0.005
Insight	-0.065	0.206***	0.130***	0.091*	-0.012
Discrepancy	-0.169***	0.133***	0.243***	-0.079*	-0.126***
Inhibition	0.274***	0.356***	-0.045	0.208***	-0.100**
Tentative	0.021	0.359***	0.273***	0.102**	0.073
Certainty	-0.163***	-0.024	0.125***	-0.100**	-0.063
Sensory and Perceptual Processes	-0.173***	0.061	0.275***	-0.031	0.140***
Seeing	-0.112**	0.034	0.148***	-0.004	0.059
Hearing	-0.135***	0.108**	0.258***	-0.002	0.145***
Feeling	-0.122**	-0.202***	0.051	-0.160***	0.017
Social Processes	-0.294***	-0.316***	0.133***	-0.219***	-0.032

Continued on next page



Continued from previous page					
	Openness	Conscientiousness	Extraversion	Agreeableness	Emotional Stability
<b>LIWC Categories (continued)</b>					
Communication	0.024	0.210***	0.166***	0.125***	0.108**
Other References to People	-0.363***	-0.292***	0.209***	-0.296***	-0.084*
Humans	-0.036	-0.072	0.037	0.090*	0.068
Time	-0.034	-0.076	-0.047	-0.043	-0.007
Past Tense Verb	-0.103**	-0.193***	0.042	-0.103**	0.157***
Present Tense Verb	-0.330***	0.051	0.337***	-0.172***	-0.097**
Future Tense Verb	0.082*	0.330***	0.111**	0.173***	-0.038
Space	-0.148***	0.133***	0.177***	-0.042	0.118**
Up	-0.015	0.137***	-0.036	0.107**	0.106**
Down	-0.077	0.110**	0.103**	-0.047	0.009
Inclusive	-0.027	-0.283***	-0.081*	-0.085*	-0.078
Exclusive	-0.097**	0.341***	0.398***	0.045	0.042
Motion	-0.152***	0.037	0.277***	-0.089*	0.112**
Achievement	-0.138***	-0.284***	-0.149***	-0.103**	-0.097**
Religion	-0.041	-0.264***	-0.049	-0.130***	0.059
Sex and Sexuality	-0.107**	-0.156***	-0.005	-0.102**	0.020
Swear Words	0.017	0.017	0.089*	-0.011	0.043
Nonfluencies	0.073	0.033	0.124***	-0.032	0.021
Fillers	-0.002	-0.050	-0.051	-0.085*	-0.092*
Number of Periods	0.030	0.048	-0.194***	0.132***	-0.048
Number of Commas	-0.052	-0.015	0.032	-0.021	0.037
Number of Colons	0.027	0.129***	0.059	0.025	-0.061
Number of Semicolons	-0.028	0.024	0.067	0.039	0.100**
Number of Question Marks	-0.011	0.191***	0.204***	0.078	0.126***
Number of Exclamations	0.027	-0.076	-0.117**	0.002	0.038
Number of Dashes	0.129***	0.017	-0.118**	0.041	-0.020
Number of Quotation Marks	-0.067	-0.050	0.029	0.018	0.134***
Number of Apostrophes	0.068	-0.222***	-0.227***	-0.049	0.008
Number of Pairs of Parentheses	0.108**	0.222***	-0.015	0.195***	0.002
Number of Other Punctuation	0.086*	0.178***	-0.028	0.159***	0.006
All Punctuation	0.065	0.103**	-0.089*	0.130***	0.002

Note: Pearson correlation coefficients presented. Two-tailed tests: \*\*\*  $p < 0.01$ , \*\*  $p < 0.05$ , \*  $p < 0.1$

Table A-7: Correlations Between LIWC/MRC Features and Jackknifed Big Five Traits (2002)

	Openness	Conscientiousness	Extraversion	Agreeableness	Emotional Stability
<b>MRCPD Categories</b>					
Age of Acquisition	0.251***	0.346***	-0.069	0.216***	-0.114**
Concreteness	-0.153***	-0.421***	-0.278***	-0.098**	0.072
Familiarity	-0.220***	-0.001	0.286***	-0.039	0.094*
Imagability	-0.207***	-0.531***	-0.253***	-0.203***	0.004
Meaningfulness (Colorado)	-0.409***	-0.483***	0.026	-0.320***	-0.044
Meaningfulness (Paivo)	-0.080*	-0.235***	0.061	-0.121**	0.156***
Number of Letters	0.310***	0.059	-0.189***	0.050	-0.136***
Number of Phonemes	0.319***	0.085*	-0.206***	0.131***	-0.094**
Number of Syllables	0.312***	0.062	-0.193***	0.085*	-0.123**
<b>LIWC Categories</b>					
Word Count	0.173***	0.269***	0.501***	0.103**	0.197***
Words Per Sentence	0.076	0.164***	0.233***	-0.031	-0.049
Sentences Ending With ?	-0.047	0.170***	0.299***	0.026	0.107**
Unique Words	-0.047	-0.276***	-0.450***	0.021	-0.140***
Dictionary Words	-0.250***	0.063	0.336***	-0.071	0.015
Words > 6 Letters	0.309***	0.049	-0.244***	0.086*	-0.121**
Abbreviations	0.154***	0.016	-0.246***	0.079*	-0.034
Total Pronouns	-0.356***	-0.201***	0.211***	-0.196***	0.009
First-Person Singular	-0.104**	-0.077	-0.159***	0.083*	0.031
First-Person Plural	-0.266***	0.006	0.281***	-0.206***	-0.138***
Total First Person	-0.252***	-0.047	0.090*	-0.088*	-0.075
Total Second Person	-0.220***	-0.056	0.214***	-0.062	0.131***
Total Third Person	-0.225***	-0.430***	0.067	-0.231***	0.095**
Negations	-0.070	0.138***	0.142***	-0.064	-0.035
Assents	-0.015	-0.053	-0.210***	-0.009	-0.065
Articles	0.305***	0.350***	0.021	0.313***	0.139***
Prepositions	0.252***	0.154***	-0.080*	0.246***	0.068
Numbers	-0.041	-0.045	-0.254***	0.016	0.004
Affective Processes	-0.006	-0.258***	0.012	-0.312***	-0.287***
Positive Feelings	-0.106**	-0.218***	-0.011	-0.156***	-0.146***
Optimism and Energy	-0.017	-0.180***	-0.019	-0.170***	-0.235***
Anxiety or Fear	0.141***	0.034	-0.015	-0.117**	-0.162***
Anger	0.133***	0.001	0.049	-0.175***	-0.106**
Sadness or Depression	0.024	-0.178***	-0.082*	-0.203***	-0.089*
Cognitive Processes	-0.014	0.386***	0.236***	0.103**	-0.005
Causation	-0.025	0.226***	0.233***	-0.060	-0.047
Insight	0.018	0.277***	0.111**	0.174***	0.042
Discrepancy	-0.148***	0.168***	0.204***	-0.028	-0.011
Inhibition	0.203***	0.314***	-0.068	0.169***	-0.047
Tentative	0.026	0.358***	0.253***	0.124***	0.085*
Certainty	-0.107**	0.009	0.106**	-0.036	0.003
Sensory and Perceptual Processes	-0.150***	0.023	0.214***	-0.034	0.155***
Seeing	-0.098**	0.020	0.162***	-0.039	0.080*
Hearing	-0.128***	0.060	0.186***	-0.000	0.150***
Feeling	-0.071	-0.140***	0.070	-0.099**	0.050
Social Processes	-0.327***	-0.300***	0.265***	-0.245***	0.052

Continued on next page

Continued from previous page					
	Openness	Conscientiousness	Extraversion	Agreeableness	Emotional Stability
<b>LIWC Categories (continued)</b>					
Communication	-0.010	0.172**	0.212***	0.111**	0.167***
Other References to People	-0.374***	-0.288***	0.283***	-0.306***	-0.005
Humans	-0.054	-0.063	0.067	0.066	0.071
Time	-0.065	-0.078	-0.075	-0.028	-0.001
Past Tense Verb	-0.105**	-0.183***	0.019	-0.125***	0.144***
Present Tense Verb	-0.292***	0.097**	0.318***	-0.090*	-0.038
Future Tense Verb	0.145***	0.380***	0.006	0.225***	-0.071
Space	-0.171***	0.042	0.018	-0.050	-0.001
Up	-0.139***	-0.083*	-0.172***	-0.028	-0.037
Down	-0.012	0.208***	0.076	-0.021	-0.103**
Inclusive	-0.010	-0.168***	0.079	-0.094**	-0.119**
Exclusive	-0.080*	0.348***	0.364***	0.073	0.072
Motion	-0.080*	0.170***	0.234***	0.022	0.086*
Achievement	-0.051	-0.230***	-0.105**	-0.092*	-0.121**
Religion	-0.046	-0.275***	-0.008	-0.098**	0.089*
Sex and Sexuality	-0.069	-0.164***	-0.029	-0.066	0.084*
Swear Words	-0.097**	-0.104**	0.064	-0.095**	-0.031
Nonfluencies	0.009	-0.041	0.078	-0.151***	-0.067
Fillers	0.049	0.067	0.008	0.020	0.058
Number of Periods	0.006	-0.039	-0.206***	0.074	-0.026
Number of Commas	-0.000	-0.068	-0.130***	0.119**	0.155***
Number of Colons	0.041	0.054	0.049	-0.009	0.012
Number of Semicolons	-0.054	-0.039	-0.029	-0.031	-0.001
Number of Question Marks	-0.041	0.158***	0.247***	0.048	0.102**
Number of Exclamations	-0.002	-0.204***	0.039	-0.192***	-0.045
Number of Dashes	0.088*	-0.001	-0.095**	-0.092*	-0.194***
Number of Quotation Marks	0.023	0.047	0.105**	0.036	0.101**
Number of Apostrophes	0.007	-0.297***	-0.177***	-0.115**	0.039
Number of Pairs of Parentheses	0.125***	0.207***	-0.041	0.135***	-0.075
Number of Other Punctuation	0.078	0.111**	-0.064	0.081*	-0.037
All Punctuation	0.062	0.004	-0.186***	0.079*	-0.025

Note: Pearson correlation coefficients presented. Two-tailed tests: \*\*\*  $p < 0.01$ , \*\*  $p < 0.05$ , \*  $p < 0.1$

Table A-8: Correlations Between LIWC/MRC Features and Jackknifed Big Five Traits (2003)

	Openness	Conscientiousness	Extraversion	Agreeableness	Emotional Stability
<b>MRCPD Categories</b>					
Age of Acquisition	0.281***	0.367***	-0.102**	0.257***	-0.121**
Concreteness	-0.136***	-0.434***	-0.299***	-0.081*	0.093*
Familiarity	-0.262***	-0.005	0.322***	-0.066	0.078
Imagability	-0.186***	-0.519***	-0.251***	-0.201***	0.037
Meaningfulness (Colorado)	-0.412***	-0.459***	0.070	-0.339***	-0.015
Meaningfulness (Paivo)	-0.147***	-0.368***	-0.118**	-0.169***	0.112**
Number of Letters	0.307***	0.008	-0.309***	0.069	-0.093*
Number of Phonemes	0.310***	0.023	-0.347***	0.135***	-0.058
Number of Syllables	0.311***	0.022	-0.320***	0.102**	-0.078
<b>LIWC Categories</b>					
Word Count	0.261***	0.304***	0.509***	0.127***	0.214***
Words Per Sentence	0.154***	0.199***	0.157***	0.152***	0.126***
Sentences Ending With ?	-0.090*	0.179***	0.313***	0.002	0.047
Unique Words	-0.075	-0.300***	-0.467***	0.006	-0.148***
Dictionary Words	-0.161***	0.087*	0.345***	-0.027	0.023
Words > 6 Letters	0.308***	0.010	-0.332***	0.096**	-0.093*
Abbreviations	0.094**	0.025	-0.225***	0.067	-0.002
Total Pronouns	-0.324***	-0.174***	0.269***	-0.213***	-0.015
First-Person Singular	-0.095**	-0.023	-0.098**	0.020	-0.098**
First-Person Plural	-0.235***	0.051	0.386***	-0.237***	-0.149***
Total First Person	-0.223***	0.023	0.222***	-0.160***	-0.163***
Total Second Person	-0.209***	0.014	0.286***	-0.033	0.186***
Total Third Person	-0.210***	-0.474***	0.009	-0.186***	0.148***
Negations	-0.059	0.209***	0.277***	-0.019	-0.046
Assents	0.053	0.190***	-0.029	0.126***	-0.071
Articles	0.356***	0.336***	-0.114**	0.385***	0.119**
Prepositions	0.281***	0.156***	-0.046	0.239***	0.113**
Numbers	-0.177***	-0.025	-0.058	-0.061	-0.049
Affective Processes	0.137***	-0.182***	0.028	-0.260***	-0.236***
Positive Feelings	-0.029	-0.126**	-0.014	-0.027	0.028
Optimism and Energy	0.117**	-0.159***	-0.086*	-0.079	-0.100**
Anxiety or Fear	0.141***	0.012	0.070	-0.195***	-0.227***
Anger	0.147***	-0.028	0.127***	-0.216***	-0.165***
Sadness or Depression	0.017	-0.100**	0.089*	-0.226***	-0.137***
Cognitive Processes	0.011	0.420***	0.299***	0.105**	-0.125***
Causation	-0.065	0.272***	0.172***	0.061	-0.059
Insight	0.008	0.324***	0.150***	0.112**	-0.119**
Discrepancy	-0.097**	0.217***	0.283***	-0.004	-0.057
Inhibition	0.217***	0.264***	-0.030	0.097**	-0.154***
Tentative	0.028	0.339***	0.212***	0.141**	0.038
Certainty	0.030	0.169***	0.230***	-0.005	-0.076
Sensory and Perceptual Processes	-0.148***	0.080*	0.296***	-0.018	0.141***
Seeing	-0.101**	0.079	0.222***	-0.027	0.089*
Hearing	-0.125***	0.112**	0.287***	0.016	0.120**
Feeling	-0.086*	-0.200***	-0.002	-0.151***	0.079
Social Processes	-0.269***	-0.298***	0.266***	-0.208***	0.076

Continued on next page

Continued from previous page

	Openness	Conscientiousness	Extraversion	Agreeableness	Emotional Stability
<b>LIWC Categories (continued)</b>					
Communication	0.025	0.202***	0.247***	0.100**	0.077
Other References to People	-0.346***	-0.287***	0.325***	-0.295***	0.036
Humans	-0.051	-0.079*	0.107**	0.048	0.073
Time	-0.179***	-0.189***	-0.073	-0.134***	-0.067
Past Tense Verb	-0.119**	-0.238***	0.048	-0.079*	0.203***
Present Tense Verb	-0.235***	0.178***	0.363***	-0.085*	-0.060
Future Tense Verb	0.088*	0.350***	0.140***	0.187***	-0.048
Space	-0.158***	0.139***	0.160***	-0.018	0.051
Up	-0.111**	0.029	-0.032	-0.000	-0.009
Down	-0.071	0.232***	0.154***	0.073	0.030
Inclusive	0.038	-0.218***	-0.026	-0.061	0.057
Exclusive	-0.030	0.399***	0.403***	0.129***	0.073
Motion	-0.180***	0.118**	0.344***	-0.058	0.162***
Achievement	-0.056	-0.254***	-0.168***	-0.038	-0.047
Religion	0.001	-0.314***	-0.005	-0.068	0.183***
Sex and Sexuality	-0.080*	-0.134***	0.002	-0.062	0.104**
Swear Words	-0.042	-0.002	0.028	-0.068	0.082*
Nonfluencies	0.120**	0.068	0.137***	0.024	0.053
Fillers	0.123**	0.042	-0.070	0.033	-0.010
Number of Periods	-0.102**	-0.089*	-0.205***	-0.040	-0.129***
Number of Commas	-0.153***	-0.099**	0.006	-0.069	0.028
Number of Colons	-0.113**	0.023	0.016	-0.043	-0.094*
Number of Semicolons	-0.038	0.113**	0.073	0.127***	0.102**
Number of Question Marks	-0.114**	0.148***	0.277***	-0.010	0.020
Number of Exclamations	-0.080*	-0.152***	0.024	-0.160***	-0.038
Number of Dashes	0.117**	0.029	-0.125***	0.062	-0.009
Number of Quotation Marks	0.062	0.015	-0.042	0.025	-0.000
Number of Apostrophes	0.079*	-0.137***	-0.199***	-0.028	-0.024
Number of Pairs of Parentheses	0.050	0.173***	-0.041	0.174***	-0.010
Number of Other Punctuation	0.029	0.149***	-0.040	0.133***	-0.008
All Punctuation	-0.032	0.042	-0.104**	0.069	-0.032

Note: Pearson correlation coefficients presented. Two-tailed tests: \*\*\*  $p < 0.01$ , \*\*  $p < 0.05$ , \*  $p < 0.1$

Table A-9: Correlations Between LIWC/MRC Features and Jackknifed Big Five Traits (2004)

	Openness	Conscientiousness	Extraversion	Agreeableness	Emotional Stability
<b>MRCPD Categories</b>					
Age of Acquisition	0.241***	0.250***	-0.082*	0.147***	-0.128***
Concreteness	-0.150***	-0.385***	-0.207***	-0.091*	0.089*
Familiarity	-0.206***	0.072	0.252***	-0.004	0.063
Imagability	-0.199***	-0.484***	-0.203***	-0.194***	0.024
Meaningfulness (Colorado)	-0.407***	-0.424***	0.088*	-0.303***	-0.018
Meaningfulness (Paivo)	-0.182***	-0.236***	-0.029	-0.098**	0.112**
Number of Letters	0.258***	-0.043	-0.283***	0.026	-0.114**
Number of Phonemes	0.280***	-0.013	-0.286***	0.090*	-0.053
Number of Syllables	0.286***	-0.013	-0.293***	0.087*	-0.064
<b>LIWC Categories</b>					
Word Count	0.202***	0.257***	0.513***	0.096**	0.233***
Words Per Sentence	0.076	0.133***	0.214***	0.052	0.136***
Sentences Ending With ?	0.005	0.250***	0.289***	0.080*	0.153***
Unique Words	-0.045	-0.282***	-0.481***	0.010	-0.156***
Dictionary Words	-0.201***	0.129***	0.345***	-0.017	0.009
Words > 6 Letters	0.248***	-0.058	-0.336***	0.052	-0.122**
Abbreviations	0.183***	0.041	-0.282***	0.058	-0.026
Total Pronouns	-0.299***	-0.097**	0.216***	-0.116**	-0.019
First-Person Singular	-0.084*	0.015	-0.099**	0.065	-0.086*
First-Person Plural	-0.216***	0.064	0.313***	-0.183***	-0.146***
Total First Person	-0.206***	0.054	0.148***	-0.082*	-0.159***
Total Second Person	-0.158***	-0.057	0.117**	0.002	0.152***
Total Third Person	-0.164***	-0.342***	0.036	-0.109**	0.130***
Negations	-0.027	0.226***	0.314***	-0.055	-0.034
Assents	0.048	0.001	-0.091*	0.049	-0.049
Articles	0.268***	0.309***	-0.100**	0.291***	0.103**
Prepositions	0.302***	0.106**	-0.147***	0.261***	0.127***
Numbers	-0.111**	-0.011	-0.070	-0.028	0.031
Affective Processes	0.051	-0.234***	-0.055	-0.240***	-0.309***
Positive Feelings	-0.084*	-0.261***	-0.040	-0.127***	-0.023
Optimism and Energy	-0.006	-0.215***	-0.115**	-0.121**	-0.178***
Anxiety or Fear	0.122**	0.023	0.069	-0.165***	-0.160***
Anger	0.071	-0.036	0.059	-0.205***	-0.186***
Sadness or Depression	0.012	-0.097**	0.052	-0.184***	-0.148***
Cognitive Processes	0.026	0.417***	0.265***	0.108**	-0.082*
Causation	-0.039	0.259***	0.204***	0.026	0.002
Insight	0.043	0.296***	0.127***	0.145***	-0.057
Discrepancy	-0.103**	0.274***	0.318***	-0.000	-0.038
Inhibition	0.209***	0.266***	-0.111**	0.127***	-0.168***
Tentative	0.051	0.342***	0.246***	0.108**	0.067
Certainty	-0.071	0.068	0.107**	-0.040	-0.082*
Sensory and Perceptual Processes	-0.138***	0.119**	0.277***	-0.004	0.151***
Seeing	-0.129***	0.038	0.132***	-0.019	0.088*
Hearing	-0.112**	0.162***	0.298***	0.024	0.145***
Feeling	-0.023	-0.085*	0.029	-0.074	0.056
Social Processes	-0.255***	-0.210***	0.242***	-0.143***	0.055

Continued on next page

Continued from previous page

	Openness	Conscientiousness	Extraversion	Agreeableness	Emotional Stability
<b>LIWC Categories (continued)</b>					
Communication	0.006	0.197***	0.236***	0.083*	0.107**
Other References to People	-0.306***	-0.202***	0.275***	-0.195***	0.039
Humans	-0.014	-0.042	0.078	0.037	0.032
Time	-0.080*	-0.128***	0.016	-0.133***	-0.036
Past Tense Verb	-0.095**	-0.108**	0.103**	-0.080*	0.184***
Present Tense Verb	-0.245***	0.153***	0.359***	-0.070	-0.032
Future Tense Verb	0.076	0.296***	0.016	0.161***	-0.075
Space	-0.110**	0.153***	0.129***	-0.007	0.087*
Up	-0.010	0.068	-0.025	0.011	0.017
Down	-0.076	0.233***	0.081*	0.003	-0.027
Inclusive	0.048	-0.243***	-0.074	-0.026	0.024
Exclusive	-0.042	0.393***	0.407***	0.108**	0.090*
Motion	-0.147***	0.131***	0.299***	0.002	0.209***
Achievement	-0.098**	-0.251***	-0.122**	-0.021	-0.080*
Religion	0.002	-0.280***	-0.026	-0.059	0.138***
Sex and Sexuality	-0.090*	-0.113**	-0.051	0.006	0.089*
Swear Words	0.023	-0.048	0.000	-0.005	0.090*
Nonfluencies	0.076	0.035	0.054	0.045	0.036
Fillers	0.022	-0.003	0.059	-0.030	0.032
Number of Periods	-0.024	-0.162***	-0.328***	-0.026	-0.127***
Number of Commas	-0.116**	-0.130***	0.021	-0.052	0.082*
Number of Colons	-0.013	0.084*	0.003	0.055	0.024
Number of Semicolons	-0.007	0.053	0.084*	0.070	0.116**
Number of Question Marks	-0.010	0.232***	0.253***	0.072	0.132***
Number of Exclamations	-0.049	-0.210***	-0.056	-0.133***	0.010
Number of Dashes	0.055	-0.060	-0.057	-0.025	-0.038
Number of Quotation Marks	-0.035	0.013	0.022	0.051	0.108**
Number of Apostrophes	0.084*	-0.206***	-0.181***	-0.061	0.033
Number of Pairs of Parentheses	0.081*	0.103**	-0.096**	0.102**	-0.011
Number of Other Punctuation	0.066	0.062	-0.080*	0.069	0.033
All Punctuation	0.015	-0.082*	-0.158***	0.013	0.018

Note: Pearson correlation coefficients presented. Two-tailed tests: \*\*\*  $p < 0.01$ , \*\*  $p < 0.05$ , \*  $p < 0.1$

Table A-10: Correlations Between LIWC/MRC Features and Jackknifed Big Five Traits (2005)

	Openness	Conscientiousness	Extraversion	Agreeableness	Emotional Stability
<b>MRCPD Categories</b>					
Age of Acquisition	0.259***	0.264***	-0.177***	0.191***	-0.206***
Concreteness	-0.168***	-0.436***	-0.261***	-0.106**	0.115**
Familiarity	-0.181***	0.145***	0.313***	0.087*	0.136***
Imagability	-0.209***	-0.526***	-0.231***	-0.199***	0.052
Meaningfulness (Colorado)	-0.384***	-0.398***	0.068	-0.269***	-0.011
Meaningfulness (Paivo)	-0.218***	-0.337***	-0.013	-0.198***	0.130***
Number of Letters	0.254***	-0.074	-0.325***	-0.044	-0.174***
Number of Phonemes	0.264***	-0.058	-0.328***	0.024	-0.125***
Number of Syllables	0.272***	-0.075	-0.325***	-0.000	-0.138***
<b>LIWC Categories</b>					
Word Count	0.147***	0.243***	0.535***	0.095**	0.299***
Words Per Sentence	0.153***	0.144***	0.169***	0.073	0.030
Sentences Ending With ?	0.025	0.248***	0.320***	0.032	0.130***
Unique Words	-0.026	-0.302***	-0.566***	-0.003	-0.229***
Dictionary Words	-0.137***	0.200***	0.400***	0.050	0.050
Words > 6 Letters	0.242***	-0.095**	-0.364***	-0.032	-0.176***
Abbreviations	0.129***	-0.004	-0.214***	-0.023	-0.084*
Total Pronouns	-0.273***	-0.045	0.259***	-0.058	0.017
First-Person Singular	-0.068	0.032	-0.213***	0.092*	-0.101**
First-Person Plural	-0.214***	0.140***	0.371***	-0.165***	-0.138***
Total First Person	-0.208***	0.128***	0.149***	-0.067	-0.171***
Total Second Person	-0.229***	0.096**	0.331***	-0.026	0.178***
Total Third Person	-0.111**	-0.386***	0.015	-0.016	0.183***
Negations	0.012	0.234***	0.237***	-0.040	-0.110**
Assents	0.012	0.058	-0.125***	0.097**	-0.032
Articles	0.275***	0.221***	-0.049	0.301***	0.183***
Prepositions	0.264***	0.061	-0.122**	0.273***	0.143***
Numbers	-0.044	0.021	-0.136***	0.040	0.010
Affective Processes	0.004	-0.205***	0.027	-0.290***	-0.309***
Positive Feelings	-0.077	-0.104**	0.048	-0.082*	-0.083*
Optimism and Energy	-0.045	-0.199***	-0.076	-0.119**	-0.130***
Anxiety or Fear	0.108**	0.043	0.054	-0.173***	-0.184***
Anger	0.114**	-0.039	0.087*	-0.200***	-0.196***
Sadness or Depression	0.060	-0.041	0.042	-0.189***	-0.215***
Cognitive Processes	0.047	0.442***	0.305***	0.080*	-0.080*
Causation	0.061	0.321***	0.232***	0.062	-0.020
Insight	0.080*	0.343***	0.103**	0.199***	-0.028
Discrepancy	-0.077	0.277***	0.329***	-0.044	-0.082*
Inhibition	0.148***	0.188***	-0.056	0.019	-0.127***
Tentative	0.079	0.358***	0.336***	0.097**	0.078
Certainty	-0.063	0.122**	0.176***	0.021	-0.029
Sensory and Perceptual Processes	-0.087*	0.123**	0.288***	0.033	0.160***
Seeing	-0.023	0.114**	0.160***	0.065	0.157***
Hearing	-0.101**	0.151***	0.285***	0.053	0.139***
Feeling	0.000	-0.166***	0.086*	-0.162***	-0.019
Social Processes	-0.194***	-0.185***	0.290***	-0.075	0.095**

Continued on next page



Continued from previous page

	Openness	Conscientiousness	Extraversion	Agreeableness	Emotional Stability
<b>LIWC Categories (continued)</b>					
Communication	0.013	0.188***	0.240***	0.114**	0.103**
Other References to People	-0.273***	-0.162***	0.336***	-0.125***	0.079
Humans	0.081*	-0.006	0.071	0.122**	0.063
Time	-0.115**	-0.173***	0.004	-0.121**	-0.004
Past Tense Verb	-0.041	-0.177***	0.119**	-0.012	0.174***
Present Tense Verb	-0.212***	0.252***	0.361***	-0.032	-0.038
Future Tense Verb	0.088*	0.378***	0.148***	0.128***	-0.013
Space	-0.173***	0.148***	0.188***	0.061	0.195***
Up	-0.110**	0.055	-0.099**	0.114**	0.086*
Down	-0.006	0.154***	0.127***	0.064	0.107**
Inclusive	-0.052	-0.179***	0.013	-0.079*	-0.021
Exclusive	0.018	0.438***	0.403***	0.158***	0.078
Motion	-0.208***	0.117**	0.275***	-0.017	0.169***
Achievement	-0.149***	-0.245***	-0.157***	-0.093*	-0.092*
Religion	-0.006	-0.280***	0.010	-0.031	0.136***
Sex and Sexuality	-0.080*	-0.154***	-0.052	-0.058	0.021
Swear Words	0.035	-0.023	-0.036	0.003	0.073
Nonfluencies	0.004	0.008	-0.016	-0.057	-0.084*
Fillers	-0.027	-0.073	0.045	-0.052	0.000
Number of Periods	0.011	-0.003	-0.195***	0.047	-0.027
Number of Commas	0.038	-0.017	-0.085*	0.079	0.106**
Number of Colons	0.078	0.165***	-0.041	0.103**	-0.033
Number of Semicolons	-0.083*	-0.069	-0.002	-0.066	-0.007
Number of Question Marks	0.020	0.223***	0.276***	0.042	0.116**
Number of Exclamations	-0.023	-0.133***	-0.017	-0.118**	0.024
Number of Dashes	0.078	-0.041	-0.139***	-0.049	-0.090*
Number of Quotation Marks	-0.042	-0.018	0.101**	-0.026	0.142***
Number of Apostrophes	0.014	-0.244***	-0.220***	-0.100**	0.000
Number of Pairs of Parentheses	0.035	0.131***	-0.055	0.105**	0.007
Number of Other Punctuation	0.028	0.100**	-0.089*	0.080*	0.024
All Punctuation	0.039	0.012	-0.182***	0.060	0.017

Note: Pearson correlation coefficients presented. Two-tailed tests: \*\*\*  $p < 0.01$ , \*\*  $p < 0.05$ , \*  $p < 0.1$

Table A-11: Correlations Between LIWC/MRC Features and Jackknifed Big Five Traits (2006)

	Openness	Conscientiousness	Extraversion	Agreeableness	Emotional Stability
<b>MRCPD Categories</b>					
Age of Acquisition	0.226***	0.227***	-0.168***	0.173***	-0.134***
Concreteness	-0.168***	-0.380***	-0.232***	-0.110**	0.078
Familiarity	-0.172***	0.120**	0.347***	0.014	0.109**
Imagability	-0.196***	-0.475***	-0.212***	-0.196***	0.021
Meaningfulness (Colorado)	-0.326***	-0.384***	0.095**	-0.269***	-0.028
Meaningfulness (Paivo)	-0.227***	-0.249***	0.050	-0.179***	0.063
Number of Letters	0.212***	-0.047	-0.296***	0.013	-0.110**
Number of Phonemes	0.214***	-0.063	-0.307***	0.037	-0.070
Number of Syllables	0.213***	-0.064	-0.316***	0.041	-0.074
<b>LIWC Categories</b>					
Word Count	0.138***	0.186***	0.513***	0.125***	0.325***
Words Per Sentence	0.098**	0.070	0.139***	-0.010	0.027
Sentences Ending With ?	-0.051	0.095**	0.277***	-0.060	0.073
Unique Words	0.011	-0.181***	-0.462***	0.033	-0.202***
Dictionary Words	-0.139***	0.158***	0.382***	0.020	0.033
Words > 6 Letters	0.197***	-0.055	-0.333***	0.019	-0.119**
Abbreviations	0.093*	-0.072	-0.227***	-0.038	-0.042
Total Pronouns	-0.229***	-0.028	0.250***	-0.091*	-0.034
First-Person Singular	-0.017	0.073	-0.119**	0.072	-0.092*
First-Person Plural	-0.185***	0.035	0.262***	-0.186***	-0.146***
Total First Person	-0.144***	0.075	0.107**	-0.084*	-0.167***
Total Second Person	-0.130***	0.049	0.229***	0.002	0.087*
Total Third Person	-0.161***	-0.315***	0.085*	-0.087*	0.129***
Negations	0.064	0.226***	0.164***	0.020	-0.086*
Assents	0.077	0.156***	-0.098**	0.094*	-0.022
Articles	0.207***	0.273***	-0.035	0.298***	0.200***
Prepositions	0.153***	0.018	-0.111**	0.217***	0.070
Numbers	-0.085*	-0.074	-0.123**	-0.063	0.027
Affective Processes	0.112**	-0.132***	-0.062	-0.178***	-0.293***
Positive Feelings	-0.007	-0.115**	-0.064	-0.064	-0.096**
Optimism and Energy	0.043	-0.149***	-0.128***	-0.056	-0.131***
Anxiety or Fear	0.179***	0.032	-0.042	-0.079	-0.137***
Anger	0.144***	0.014	-0.044	-0.106**	-0.137***
Sadness or Depression	0.066	-0.012	0.112**	-0.116**	-0.114**
Cognitive Processes	0.058	0.381***	0.254***	0.095**	-0.062
Causation	-0.034	0.176***	0.059	0.027	-0.016
Insight	0.029	0.290***	0.173***	0.169***	0.020
Discrepancy	0.040	0.281***	0.278***	0.045	-0.054
Inhibition	0.173***	0.230***	-0.044	0.040	-0.158***
Tentative	0.034	0.315***	0.300***	0.061	0.061
Certainty	-0.040	0.067	0.133***	-0.030	-0.008
Sensory and Perceptual Processes	-0.147***	0.063	0.291***	-0.032	0.165***
Seeing	-0.160***	0.068	0.232***	-0.031	0.119**
Hearing	-0.108**	0.101**	0.284***	0.004	0.154***
Feeling	-0.045	-0.198***	-0.025	-0.139***	0.017
Social Processes	-0.227***	-0.208***	0.271***	-0.147***	0.053

Continued on next page

Continued from previous page

	Openness	Conscientiousness	Extraversion	Agreeableness	Emotional Stability
<b>LIWC Categories (continued)</b>					
Communication	-0.007	0.138***	0.222***	0.031	0.079
Other References to People	-0.272***	-0.165***	0.307***	-0.182***	0.014
Humans	0.009	-0.132***	0.093*	0.061	0.115**
Time	-0.075	-0.152***	-0.036	-0.124**	-0.041
Past Tense Verb	-0.096**	-0.133***	0.167***	-0.077	0.192***
Present Tense Verb	-0.170***	0.168***	0.311***	-0.038	-0.047
Future Tense Verb	0.150***	0.245***	0.130***	0.112**	0.041
Space	-0.182***	0.080*	0.200***	-0.037	0.126***
Up	-0.113**	-0.010	-0.079	-0.033	-0.022
Down	-0.047	0.139***	0.106**	0.043	0.065
Inclusive	0.013	-0.171***	-0.017	-0.027	-0.022
Exclusive	0.007	0.364***	0.386***	0.104**	0.083*
Motion	-0.048	0.152***	0.283***	0.005	0.114**
Achievement	-0.111**	-0.283***	-0.164***	-0.056	-0.073
Religion	-0.044	-0.186***	0.051	-0.011	0.084*
Sex and Sexuality	-0.113**	-0.211***	-0.064	-0.061	0.167***
Swear Words	-0.029	-0.114**	0.061	-0.046	0.043
Nonfluencies	0.079	0.047	0.012	-0.018	-0.004
Fillers	0.050	0.003	-0.012	-0.018	-0.086*
Number of Periods	-0.059	-0.092*	-0.238***	0.005	-0.040
Number of Commas	-0.106**	-0.097**	-0.003	-0.018	0.050
Number of Colons	0.030	0.017	-0.036	0.048	0.019
Number of Semicolons	-0.021	-0.021	-0.004	-0.030	-0.002
Number of Question Marks	-0.064	0.086*	0.251***	-0.053	0.065
Number of Exclamations	0.021	-0.145***	-0.011	-0.155***	-0.032
Number of Dashes	0.066	-0.096**	-0.214***	-0.058	-0.059
Number of Quotation Marks	-0.135***	-0.059	0.120**	-0.122**	0.042
Number of Apostrophes	0.052	-0.184***	-0.189***	-0.074	-0.001
Number of Pairs of Parentheses	0.006	0.048	-0.059	0.019	0.011
Number of Other Punctuation	0.019	0.035	-0.069	0.006	0.033
All Punctuation	-0.029	-0.089*	-0.165***	-0.029	0.012

Note: Pearson correlation coefficients presented. Two-tailed tests: \*\*\*  $p < 0.01$ , \*\*  $p < 0.05$ , \*  $p < 0.1$

Table A-12: Correlations Between LIWC/MRC Features and Jackknifed Big Five Traits (2007)

	Openness	Conscientiousness	Extraversion	Agreeableness	Emotional Stability
<b>MRCPD Categories</b>					
Age of Acquisition	0.255***	0.328***	-0.174***	0.114**	-0.235***
Concreteness	-0.109**	-0.377***	-0.191***	-0.021	0.081*
Familiarity	-0.278***	0.054	0.280***	-0.003	0.103**
Imagability	-0.143***	-0.462***	-0.172***	-0.080*	0.037
Meaningfulness (Colorado)	-0.370***	-0.385***	0.084*	-0.184***	-0.015
Meaningfulness (Paivo)	-0.184***	-0.342***	0.051	-0.097**	0.158***
Number of Letters	0.293***	-0.046	-0.274***	0.024	-0.103**
Number of Phonemes	0.309***	-0.047	-0.287***	0.063	-0.074
Number of Syllables	0.305***	-0.061	-0.282***	0.044	-0.084*
<b>LIWC Categories</b>					
Word Count	0.180***	0.340***	0.543***	0.207***	0.290***
Words Per Sentence	0.128***	0.122**	0.193***	0.056	0.143***
Sentences Ending With ?	-0.070	0.259***	0.271***	0.082*	0.183***
Unique Words	-0.019	-0.293***	-0.447***	-0.062	-0.199***
Dictionary Words	-0.190***	0.164***	0.337***	0.028	0.033
Words > 6 Letters	0.296***	-0.034	-0.317***	0.046	-0.127***
Abbreviations	0.110**	-0.068	-0.210***	0.030	-0.082*
Total Pronouns	-0.320***	-0.071	0.214***	-0.103**	-0.028
First-Person Singular	-0.074	0.049	-0.145***	0.026	-0.152***
First-Person Plural	-0.225***	0.028	0.297***	-0.258***	-0.188***
Total First Person	-0.217***	0.057	0.101**	-0.163***	-0.250***
Total Second Person	-0.259***	-0.024	0.224***	-0.083*	0.126***
Total Third Person	-0.123**	-0.312***	0.039	0.086*	0.215***
Negations	-0.028	0.179***	0.084*	-0.014	-0.106**
Assents	-0.019	0.065	-0.027	0.004	-0.116**
Articles	0.319***	0.296***	-0.030	0.291***	0.184***
Prepositions	0.237***	0.049	-0.031	0.178***	0.118**
Numbers	-0.071	-0.124***	-0.180***	-0.037	-0.018
Affective Processes	0.026	-0.189***	-0.091*	-0.116**	-0.190***
Positive Feelings	-0.078	-0.099**	0.064	-0.027	-0.070
Optimism and Energy	0.078	-0.198***	-0.160***	0.018	-0.035
Anxiety or Fear	0.118**	0.034	0.024	-0.099**	-0.064
Anger	0.140***	-0.007	0.036	-0.125***	-0.052
Sadness or Depression	0.067	0.010	-0.021	-0.095**	-0.151***
Cognitive Processes	0.011	0.440***	0.237***	0.057	-0.062
Causation	-0.021	0.281***	0.221***	-0.028	-0.010
Insight	0.008	0.344***	0.108**	0.116**	-0.048
Discrepancy	-0.088*	0.303***	0.217***	0.006	-0.037
Inhibition	0.283***	0.283***	-0.043	0.035	-0.188***
Tentative	-0.003	0.351***	0.346***	0.083*	0.115**
Certainty	-0.169***	0.071	0.125***	-0.014	0.046
Sensory and Perceptual Processes	-0.202***	0.046	0.270***	-0.078	0.146***
Seeing	-0.192***	0.026	0.189***	-0.095**	0.108**
Hearing	-0.156***	0.086*	0.264***	-0.049	0.120**
Feeling	-0.102**	-0.176***	0.012	-0.060	0.097**
Social Processes	-0.252***	-0.203***	0.283***	-0.068	0.114**

Continued on next page

Continued from previous page

	Openness	Conscientiousness	Extraversion	Agreeableness	Emotional Stability
<b>LIWC Categories (continued)</b>					
Communication	-0.033	0.159***	0.225***	0.030	0.108**
Other References to People	-0.319***	-0.194***	0.305***	-0.138***	0.067
Humans	-0.007	-0.029	0.132***	0.067	0.051
Time	-0.058	-0.110**	0.006	-0.039	0.004
Past Tense Verb	-0.093*	-0.032	0.091*	0.044	0.147***
Present Tense Verb	-0.257***	0.130***	0.301***	-0.139***	-0.082*
Future Tense Verb	0.021	0.306***	0.181***	0.071	-0.041
Space	-0.206***	0.098**	0.148***	-0.034	0.099**
Up	-0.112**	-0.001	-0.060	-0.023	-0.019
Down	0.021	0.225***	0.148***	-0.049	-0.006
Inclusive	-0.005	-0.167***	-0.008	0.010	0.013
Exclusive	-0.031	0.373***	0.382***	0.057	0.054
Motion	-0.165***	0.070	0.291***	-0.078	0.114**
Achievement	-0.099**	-0.256***	-0.229***	-0.001	-0.054
Religion	-0.079*	-0.245***	0.034	-0.020	0.152***
Sex and Sexuality	-0.083*	-0.101**	-0.001	-0.037	0.069
Swear Words	0.025	0.018	0.032	0.041	0.117**
Nonfluencies	0.024	0.068	0.026	-0.003	0.002
Fillers	0.023	-0.038	-0.097**	-0.027	-0.064
Number of Periods	-0.067	-0.121**	-0.234***	-0.047	-0.165***
Number of Commas	-0.068	-0.143***	-0.035	-0.023	0.071
Number of Colons	-0.075	-0.053	-0.025	-0.011	0.004
Number of Semicolons	-0.079*	-0.072	0.016	-0.028	0.012
Number of Question Marks	-0.082*	0.246***	0.240***	0.076	0.160***
Number of Exclamations	0.046	-0.125***	-0.082*	0.065	0.023
Number of Dashes	0.127***	-0.056	-0.178***	0.036	0.002
Number of Quotation Marks	-0.028	0.024	0.034	0.011	0.053
Number of Apostrophes	-0.073	-0.011	0.032	0.006	0.034
Number of Pairs of Parentheses	0.040	0.017	-0.055	-0.009	-0.049
Number of Other Punctuation	0.021	0.034	-0.052	-0.011	-0.056
All Punctuation	-0.045	-0.084*	-0.112**	-0.019	-0.033

Note: Pearson correlation coefficients presented. Two-tailed tests: \*\*\*  $p < 0.01$ , \*\*  $p < 0.05$ , \*  $p < 0.1$

Table A-13: Correlations Between LIWC/MRC Features and Jackknifed Big Five Traits (2008)

	Openness	Conscientiousness	Extraversion	Agreeableness	Emotional Stability
<b>MRCPD Categories</b>					
Age of Acquisition	0.306***	0.171***	-0.157***	0.098**	-0.189***
Concreteness	-0.054	-0.351***	-0.282***	-0.030	0.036
Familiarity	-0.222***	0.147***	0.319***	-0.016	0.093*
Imagability	-0.133***	-0.469***	-0.285***	-0.121**	-0.003
Meaningfulness (Colorado)	-0.399***	-0.329***	0.090*	-0.278***	-0.024
Meaningfulness (Paivo)	-0.059	-0.243***	-0.027	-0.053	0.110**
Number of Letters	0.298***	-0.111**	-0.314***	0.095**	-0.116**
Number of Phonemes	0.314***	-0.114**	-0.311***	0.126***	-0.083*
Number of Syllables	0.277***	-0.140***	-0.310***	0.116**	-0.087*
<b>LIWC Categories</b>					
Word Count	0.194***	0.278***	0.494***	0.175***	0.280***
Words Per Sentence	0.082*	0.071	0.148***	0.115**	0.152***
Sentences Ending With ?	-0.117**	0.205***	0.294***	0.013	0.152***
Unique Words	-0.006	-0.197***	-0.444***	0.113**	-0.177***
Dictionary Words	-0.173***	0.208***	0.381***	-0.055	0.037
Words > 6 Letters	0.295***	-0.106**	-0.333***	0.078	-0.140***
Abbreviations	0.101**	-0.116**	-0.260***	-0.100**	-0.134***
Total Pronouns	-0.329***	0.006	0.265***	-0.127***	-0.006
First-Person Singular	-0.002	0.075	-0.120**	0.006	-0.183***
First-Person Plural	-0.311***	0.100**	0.272***	-0.138***	-0.050
Total First Person	-0.219***	0.123**	0.105**	-0.092*	-0.165***
Total Second Person	-0.260***	0.053	0.313***	-0.032	0.154***
Total Third Person	-0.196***	-0.383***	0.059	-0.115**	0.144***
Negations	-0.038	0.314***	0.184***	-0.009	-0.025
Assents	-0.026	0.075	-0.030	-0.034	-0.156***
Articles	0.155***	0.215***	-0.043	0.168***	0.124***
Prepositions	0.187***	-0.045	-0.068	0.133***	0.089*
Numbers	-0.059	-0.168***	-0.256***	0.053	0.038
Affective Processes	0.028	-0.237***	-0.068	-0.162***	-0.246***
Positive Feelings	-0.025	-0.169***	0.058	-0.187***	-0.149***
Optimism and Energy	0.016	-0.236***	-0.170***	-0.032	-0.107**
Anxiety or Fear	0.187***	0.106**	0.081*	-0.120**	-0.109**
Anger	0.103**	-0.043	0.036	-0.060	-0.019
Sadness or Depression	0.020	-0.002	0.075	-0.125***	-0.080*
Cognitive Processes	0.027	0.450***	0.336***	0.065	0.007
Causation	0.026	0.302***	0.290***	0.112**	0.076
Insight	0.019	0.297***	0.195***	0.091*	-0.025
Discrepancy	-0.086*	0.347***	0.298***	0.006	0.002
Inhibition	0.301***	0.275***	-0.034	0.102**	-0.071
Tentative	0.019	0.373***	0.380***	0.035	0.085*
Certainty	-0.070	0.194***	0.185***	-0.001	-0.006
Sensory and Perceptual Processes	-0.116**	0.172***	0.346***	0.017	0.182***
Seeing	-0.091*	0.065	0.247***	0.043	0.192***
Hearing	-0.094*	0.225***	0.322***	0.038	0.147***
Feeling	-0.111**	-0.123**	0.111**	-0.191***	0.008
Social Processes	-0.340***	-0.174***	0.272***	-0.157***	0.076

Continued on next page

Continued from previous page					
	Openness	Conscientiousness	Extraversion	Agreeableness	Emotional Stability
<b>LIWC Categories (continued)</b>					
Communication	0.018	0.231***	0.182***	0.116**	0.088*
Other References to People	-0.422***	-0.148***	0.330***	-0.185***	0.094*
Humans	0.019	-0.063	0.081*	-0.013	0.012
Time	-0.171***	-0.126***	-0.013	-0.186***	-0.008
Past Tense Verb	-0.124**	-0.036	0.160***	0.012	0.245***
Present Tense Verb	-0.183***	0.228***	0.365***	-0.142***	-0.077
Future Tense Verb	0.091*	0.331***	0.146***	0.094*	-0.042
Space	-0.223***	0.144***	0.244***	-0.057	0.162***
Up	-0.187***	0.061	-0.033	0.008	0.066
Down	-0.033	0.223***	0.249***	-0.026	0.043
Inclusive	-0.007	-0.205***	0.017	-0.089*	-0.043
Exclusive	-0.026	0.427***	0.456***	0.076	0.151***
Motion	-0.240***	0.053	0.230***	-0.002	0.183***
Achievement	-0.045	-0.251***	-0.201***	0.060	-0.059
Religion	-0.011	-0.247***	0.015	0.013	0.165***
Sex and Sexuality	-0.034	-0.102**	-0.037	-0.018	0.120**
Swear Words	-0.021	0.033	-0.026	0.041	0.055
Nonfluencies	-0.001	-0.060	-0.035	0.000	0.010
Fillers	-0.064	-0.073	-0.026	-0.065	-0.033
Number of Periods	-0.042	-0.089*	-0.245***	-0.112**	-0.184***
Number of Commas	-0.079	-0.112**	-0.092*	0.147***	0.177***
Number of Colons	0.011	-0.052	-0.158***	0.074	-0.004
Number of Semicolons	0.026	-0.027	0.037	0.022	0.046
Number of Question Marks	-0.107**	0.189***	0.219***	0.022	0.113**
Number of Exclamations	-0.024	-0.094*	-0.056	-0.038	-0.023
Number of Dashes	0.132***	-0.053	-0.142***	0.187***	0.025
Number of Quotation Marks	0.055	0.046	0.077	0.028	0.099**
Number of Apostrophes	-0.135***	0.009	0.139***	-0.024	0.096**
Number of Pairs of Parentheses	0.136***	0.004	-0.093*	0.154***	-0.062
Number of Other Punctuation	0.078	-0.037	-0.127***	0.123**	-0.011
All Punctuation	-0.007	-0.082*	-0.155***	0.118**	0.031

Note: Pearson correlation coefficients presented. Two-tailed tests: \*\*\*  $p < 0.01$ , \*\*  $p < 0.05$ , \*  $p < 0.1$

Table A-14: Correlations Between LIWC/MRC Features and Jackknifed Big Five Traits (2009)

	Openness	Conscientiousness	Extraversion	Agreeableness	Emotional Stability
<b>MRCPD Categories</b>					
Age of Acquisition	0.235***	0.196***	-0.085*	0.068	-0.143***
Concreteness	-0.083*	-0.320***	-0.332***	-0.007	-0.033
Familiarity	-0.201***	0.051	0.356***	-0.056	0.146***
Imagability	-0.118**	-0.387***	-0.292***	-0.083*	-0.064
Meaningfulness (Colorado)	-0.369***	-0.320***	0.107**	-0.234***	0.002
Meaningfulness (Paivo)	-0.113**	-0.216***	-0.154***	-0.064	-0.006
Number of Letters	0.190***	-0.066	-0.331***	0.009	-0.193***
Number of Phonemes	0.206***	-0.087*	-0.353***	0.053	-0.130***
Number of Syllables	0.188***	-0.072	-0.328***	0.027	-0.174***
<b>LIWC Categories</b>					
Word Count	0.157***	0.285***	0.491***	0.261***	0.391***
Words Per Sentence	0.062	0.061	0.163***	0.055	0.135***
Sentences Ending With ?	-0.179***	0.229***	0.169***	0.131***	0.295***
Unique Words	-0.087*	-0.194***	-0.421***	-0.029	-0.206***
Dictionary Words	-0.161***	0.104**	0.410***	-0.051	0.131***
Words > 6 Letters	0.194***	-0.078	-0.343***	0.000	-0.234***
Abbreviations	0.103**	-0.109**	-0.292***	-0.052	-0.197***
Total Pronouns	-0.258***	-0.017	0.329***	-0.089*	0.100**
First-Person Singular	-0.076	-0.044	-0.065	-0.060	-0.200***
First-Person Plural	-0.258***	0.103**	0.324***	-0.156***	-0.014
Total First Person	-0.228***	0.041	0.179***	-0.147***	-0.144***
Total Second Person	-0.192***	0.041	0.327***	-0.017	0.268***
Total Third Person	-0.138***	-0.260***	0.133***	-0.018	0.222***
Negations	0.034	0.262***	0.158***	0.012	-0.040
Assents	0.028	0.073	-0.112**	0.024	-0.174***
Articles	0.228***	0.157***	0.014	0.167***	0.078
Prepositions	0.137***	-0.012	0.042	0.037	-0.011
Numbers	0.008	-0.056	-0.283***	0.042	-0.073
Affective Processes	-0.079*	-0.116**	0.102**	-0.266***	-0.314***
Positive Feelings	-0.108**	-0.038	0.207***	-0.150***	-0.092*
Optimism and Energy	-0.075	-0.123**	-0.104**	-0.107**	-0.179***
Anxiety or Fear	0.180***	0.013	0.132***	-0.179***	-0.090*
Anger	0.217***	0.012	0.083*	-0.127***	-0.140***
Sadness or Depression	-0.144***	-0.007	-0.063	-0.052	-0.015
Cognitive Processes	0.038	0.317***	0.205***	0.125***	0.117**
Causation	0.027	0.253***	0.235***	0.035	0.102**
Insight	-0.059	0.167***	0.215***	0.009	-0.011
Discrepancy	0.019	0.249***	0.180***	0.107**	0.119**
Inhibition	0.196***	0.211***	-0.155***	0.179***	0.007
Tentative	0.074	0.283***	0.206***	0.206***	0.290***
Certainty	-0.099**	0.122**	0.332***	-0.071	0.029
Sensory and Perceptual Processes	-0.179***	0.063	0.334***	-0.018	0.284***
Seeing	-0.206***	0.008	0.276***	-0.084*	0.192***
Hearing	-0.128***	0.099**	0.314***	0.024	0.277***
Feeling	-0.130***	-0.089*	0.080*	-0.052	0.123**
Social Processes	-0.264***	-0.116**	0.301***	-0.043	0.242***

Continued on next page



Continued from previous page

	Openness	Conscientiousness	Extraversion	Agreeableness	Emotional Stability
<b>LIWC Categories (continued)</b>					
Communication	-0.018	0.104**	0.009	0.121**	0.157***
Other References to People	-0.294***	-0.081*	0.378***	-0.107**	0.212***
Humans	0.052	0.003	-0.063	0.158***	0.136***
Time	-0.073	-0.077	-0.007	-0.009	0.034
Past Tense Verb	-0.070	-0.076	0.228***	-0.004	0.243***
Present Tense Verb	-0.199***	0.159***	0.393***	-0.107**	0.029
Future Tense Verb	0.086*	0.219***	-0.004	0.176***	0.109**
Space	-0.196***	0.100**	0.146***	-0.047	0.128***
Up	-0.131***	-0.097**	-0.051	-0.110**	-0.040
Down	-0.084*	0.299***	0.127***	0.077	0.094*
Inclusive	-0.026	-0.180***	0.188***	-0.180***	-0.109**
Exclusive	-0.058	0.336***	0.410***	0.107**	0.242***
Motion	-0.172***	0.001	0.249***	-0.085*	0.230***
Achievement	-0.197***	-0.239***	0.014	-0.208***	-0.178***
Religion	-0.012	-0.151***	-0.018	0.063	0.122**
Sex and Sexuality	-0.053	-0.146***	0.080*	-0.067	0.034
Swear Words	-0.022	-0.050	0.016	-0.028	0.011
Nonfluencies	0.081*	0.073	0.023	0.057	0.018
Fillers	-0.035	0.004	-0.014	0.034	-0.004
Number of Periods	0.001	-0.046	-0.243***	-0.010	-0.172***
Number of Commas	-0.022	-0.065	-0.141***	0.191***	0.122**
Number of Colons	0.036	-0.007	-0.207***	0.090*	-0.017
Number of Semicolons	-0.027	-0.070	0.022	0.024	0.034
Number of Question Marks	-0.178***	0.223***	0.155***	0.124***	0.271***
Number of Exclamations	-0.030	-0.108**	0.012	-0.117**	0.032
Number of Dashes	-0.094*	0.033	-0.164***	0.096**	0.083*
Number of Quotation Marks	-0.029	0.010	-0.056	-0.022	-0.074
Number of Apostrophes	-0.157***	-0.017	0.197***	-0.075	0.166***
Number of Pairs of Parentheses	0.125***	0.061	-0.252***	0.199***	0.045
Number of Other Punctuation	0.069	-0.014	-0.315***	0.127***	0.014
All Punctuation	-0.024	-0.031	-0.301***	0.157***	0.045

Note: Pearson correlation coefficients presented. Two-tailed tests: \*\*\*  $p < 0.01$ , \*\*  $p < 0.05$ , \*  $p < 0.1$

Table A-15: Correlations Between LIWC/MRC Features and Jackknifed Big Five Traits (2010)

	Openness	Conscientiousness	Extraversion	Agreeableness	Emotional Stability
<b>MRCPD Categories</b>					
Age of Acquisition	0.195***	0.163***	-0.236***	0.090*	-0.194***
Concreteness	-0.150***	-0.373***	-0.232***	-0.060	0.040
Familiarity	-0.129***	0.123**	0.292***	0.057	0.207***
Imagability	-0.162***	-0.427***	-0.207***	-0.119**	-0.015
Meaningfulness (Colorado)	-0.420***	-0.273***	-0.000	-0.133***	0.055
Meaningfulness (Paivo)	-0.097**	-0.242***	-0.025	-0.146***	0.014
Number of Letters	0.208***	-0.105**	-0.236***	-0.072	-0.248***
Number of Phonemes	0.184***	-0.127***	-0.263***	-0.032	-0.176***
Number of Syllables	0.203***	-0.116**	-0.240***	-0.048	-0.210***
<b>LIWC Categories</b>					
Word Count	0.143***	0.235***	0.463***	0.259***	0.394***
Words Per Sentence	0.061	0.087*	0.108**	0.040	0.079
Sentences Ending With ?	-0.143***	0.202***	0.238***	0.091*	0.264***
Unique Words	0.042	-0.183***	-0.269***	-0.116**	-0.301***
Dictionary Words	-0.199***	0.179***	0.281***	0.051	0.143***
Words > 6 Letters	0.226***	-0.089*	-0.253***	-0.056	-0.281***
Abbreviations	0.026	-0.070	-0.221***	-0.049	-0.136***
Total Pronouns	-0.321***	0.008	0.207***	-0.024	0.092*
First-Person Singular	-0.065	0.031	-0.132***	0.031	-0.168***
First-Person Plural	-0.195***	0.162***	0.283***	-0.111**	-0.058
Total First Person	-0.172***	0.127***	0.095**	-0.052	-0.152***
Total Second Person	-0.228***	0.089*	0.234***	0.081*	0.286***
Total Third Person	-0.227***	-0.359***	0.039	-0.075	0.164***
Negations	-0.032	0.305***	0.020	0.106**	0.025
Assents	-0.017	0.116**	-0.143***	0.068	-0.103**
Articles	0.287***	0.139***	0.006	0.163***	0.156***
Prepositions	0.296***	0.030	0.049	0.066	-0.020
Numbers	0.004	-0.112**	-0.181***	-0.004	0.021
Affective Processes	-0.081*	-0.171***	-0.047	-0.232***	-0.333***
Positive Feelings	-0.064	-0.001	-0.029	-0.007	-0.080*
Optimism and Energy	-0.042	-0.232***	-0.129***	-0.170***	-0.194***
Anxiety or Fear	0.123**	0.064	0.027	-0.081*	-0.104**
Anger	0.116**	0.012	0.043	-0.123**	-0.140***
Sadness or Depression	-0.139***	0.003	0.065	-0.102**	-0.073
Cognitive Processes	-0.058	0.381***	0.176***	0.124***	0.092*
Causation	-0.086*	0.288***	0.182***	0.064	0.107**
Insight	0.062	0.192***	0.161***	0.091*	0.048
Discrepancy	-0.087*	0.326***	0.179***	0.101**	0.115**
Inhibition	0.120**	0.235***	-0.123**	0.064	-0.134***
Tentative	0.074	0.321***	0.304***	0.103**	0.162***
Certainty	-0.106**	0.143***	0.151***	0.036	0.020
Sensory and Perceptual Processes	-0.183***	0.157***	0.215***	0.094*	0.294***
Seeing	-0.194***	0.097**	0.189***	0.039	0.205***
Hearing	-0.151***	0.182***	0.194***	0.113**	0.272***
Feeling	-0.033	-0.096**	0.030	-0.034	0.141***
Social Processes	-0.275***	-0.137***	0.271***	-0.049	0.199***

Continued on next page

Continued from previous page

	Openness	Conscientiousness	Extraversion	Agreeableness	Emotional Stability
<b>LIWC Categories (continued)</b>					
Communication	-0.001	0.243***	0.177***	0.148***	0.230***
Other References to People	-0.355***	-0.114**	0.285***	-0.093*	0.172***
Humans	0.124**	-0.030	0.203***	0.082*	0.154***
Time	0.040	-0.093*	-0.048	-0.070	-0.014
Past Tense Verb	-0.144***	-0.074	0.084*	0.070	0.366***
Present Tense Verb	-0.160***	0.228***	0.280***	-0.001	0.030
Future Tense Verb	0.181***	0.270***	0.200***	0.016	-0.087*
Space	-0.005	0.083*	0.216***	-0.025	0.061
Up	0.025	-0.015	-0.036	-0.023	-0.040
Down	0.294***	0.150***	0.319***	-0.005	0.056
Inclusive	-0.140***	-0.161***	-0.084*	-0.059	-0.049
Exclusive	-0.040	0.393***	0.324***	0.187***	0.235***
Motion	-0.225***	0.048	0.207***	-0.009	0.255***
Achievement	-0.168***	-0.210***	-0.117**	-0.091*	-0.119**
Religion	0.024	-0.211***	-0.038	0.026	0.136***
Sex and Sexuality	-0.097**	-0.162***	0.010	-0.020	0.077
Swear Words	0.006	-0.001	0.035	0.028	0.146***
Nonfluencies	-0.018	-0.024	0.002	0.022	0.072
Fillers	NA	NA	NA	NA	NA
Number of Periods	-0.069	-0.086*	-0.164***	-0.040	-0.087*
Number of Commas	-0.063	-0.099**	-0.099**	0.085*	0.147***
Number of Colons	-0.020	0.028	-0.078	0.040	0.003
Number of Semicolons	0.005	0.069	-0.005	0.100**	0.045
Number of Question Marks	-0.155***	0.185***	0.221***	0.083*	0.247***
Number of Exclamations	-0.023	-0.095**	-0.083*	-0.004	0.058
Number of Dashes	-0.002	0.033	-0.009	0.044	0.043
Number of Quotation Marks	-0.096**	-0.070	0.005	0.000	0.060
Number of Apostrophes	-0.121**	0.074	0.001	0.084*	0.185***
Number of Pairs of Parentheses	0.137***	0.038	-0.007	0.008	-0.092*
Number of Other Punctuation	0.037	0.038	-0.170***	0.039	-0.041
All Punctuation	-0.054	-0.008	-0.136***	0.067	0.058

Note: Pearson correlation coefficients presented. Two-tailed tests: \*\*\*  $p < 0.01$ , \*\*  $p < 0.05$ , \*  $p < 0.1$

Table A-16: Correlations Between LIWC/MRC Features and Jackknifed Big Five Traits (2011)

	Openness	Conscientiousness	Extraversion	Agreeableness	Emotional Stability
<b>MRCPD Categories</b>					
Age of Acquisition	0.142***	0.195***	-0.253***	0.116**	-0.205***
Concreteness	0.011	-0.355***	-0.166***	-0.042	0.040
Familiarity	-0.229***	0.077	0.289***	-0.057	0.183***
Imagability	-0.009	-0.424***	-0.121**	-0.099**	0.011
Meaningfulness (Colorado)	-0.272***	-0.324***	0.188***	-0.282***	0.049
Meaningfulness (Paivo)	-0.066	-0.239***	0.017	-0.136***	0.104**
Number of Letters	0.294***	-0.082*	-0.317***	0.058	-0.248***
Number of Phonemes	0.259***	-0.112**	-0.311***	0.052	-0.157***
Number of Syllables	0.280***	-0.104**	-0.321***	0.063	-0.197***
<b>LIWC Categories</b>					
Word Count	0.129***	0.237***	0.528***	0.155***	0.352***
Words Per Sentence	0.125***	0.065	0.133***	0.086*	0.054
Sentences Ending With ?	-0.038	0.245***	0.297***	0.029	0.174***
Unique Words	0.071	-0.153***	-0.419***	0.053	-0.293***
Dictionary Words	-0.279***	0.100**	0.327***	-0.186***	0.103**
Words > 6 Letters	0.277***	-0.078	-0.338***	0.065	-0.255***
Abbreviations	0.152***	-0.018	-0.188***	0.023	-0.125***
Total Pronouns	-0.275***	0.025	0.269***	-0.086*	0.077
First-Person Singular	-0.098**	0.041	-0.122**	-0.023	-0.178***
First-Person Plural	-0.219***	0.131***	0.268***	-0.150***	-0.074
Total First Person	-0.215***	0.118**	0.127***	-0.122**	-0.153***
Total Second Person	-0.171***	0.143***	0.269***	0.062	0.255***
Total Third Person	-0.066	-0.284***	0.090*	-0.023	0.177***
Negations	-0.032	0.181***	0.102**	-0.065	-0.063
Assents	0.053	0.098**	-0.142***	0.078	-0.122**
Articles	0.036	0.129***	-0.107**	0.068	0.149***
Prepositions	-0.046	-0.042	0.024	-0.119**	0.046
Numbers	-0.137***	-0.164***	-0.138***	-0.123**	0.032
Affective Processes	0.086*	-0.150***	0.036	-0.205***	-0.308***
Positive Feelings	-0.011	-0.025	0.061	-0.068	-0.167***
Optimism and Energy	0.087*	-0.147***	-0.073	-0.085*	-0.179***
Anxiety or Fear	0.104**	0.030	-0.025	-0.092*	-0.129***
Anger	0.157***	0.045	0.060	-0.136***	-0.152***
Sadness or Depression	-0.071	-0.078	0.065	-0.143***	-0.082*
Cognitive Processes	-0.154***	0.283***	0.190***	-0.026	-0.006
Causation	-0.186***	0.199***	0.165***	-0.058	0.061
Insight	-0.215***	0.113**	0.146***	-0.028	0.054
Discrepancy	-0.137***	0.231***	0.239***	-0.009	0.034
Inhibition	0.093*	0.202***	-0.113**	-0.014	-0.234***
Tentative	-0.035	0.263***	0.215***	0.079	0.169***
Certainty	-0.092*	0.137***	0.184***	-0.059	0.046
Sensory and Perceptual Processes	-0.187***	0.095**	0.237***	-0.020	0.238***
Seeing	-0.120**	0.080*	0.245***	-0.024	0.228***
Hearing	-0.192***	0.115**	0.185***	-0.010	0.192***
Feeling	0.007	-0.119**	0.081*	-0.035	0.095**
Social Processes	-0.186***	-0.111**	0.298***	-0.086*	0.155***

Continued on next page

Continued from previous page					
	Openness	Conscientiousness	Extraversion	Agreeableness	Emotional Stability
<b>LIWC Categories (continued)</b>					
Communication	-0.098**	0.144***	0.136***	0.040	0.150***
Other References to People	-0.261***	-0.069	0.344***	-0.109**	0.153***
Humans	0.115**	0.026	0.082*	0.107**	0.037
Time	-0.138***	-0.192***	0.011	-0.184***	0.039
Past Tense Verb	-0.101**	-0.164***	0.109**	-0.019	0.340***
Present Tense Verb	-0.229***	0.181***	0.258***	-0.088*	0.001
Future Tense Verb	-0.041	0.245***	0.089*	-0.031	-0.147***
Space	-0.246***	0.044	0.181***	-0.157***	0.112**
Up	-0.185***	-0.076	-0.059	-0.189***	-0.054
Down	-0.107**	0.169***	0.174***	-0.012	0.122**
Inclusive	-0.032	-0.198***	0.045	-0.148***	0.060
Exclusive	-0.101**	0.360***	0.314***	0.085*	0.170***
Motion	-0.168***	0.102**	0.254***	-0.054	0.241***
Achievement	-0.037	-0.173***	-0.037	-0.079	-0.152***
Religion	0.078	-0.232***	0.041	-0.029	0.132***
Sex and Sexuality	-0.066	-0.150***	0.033	-0.039	0.107**
Swear Words	0.000	-0.027	0.004	0.053	0.155***
Nonfluencies	-0.012	-0.115**	-0.046	-0.084*	0.011
Fillers	-0.047	-0.041	-0.052	-0.038	-0.003
Number of Periods	-0.088*	-0.126***	-0.234***	-0.087*	-0.107**
Number of Commas	0.204***	0.023	-0.135***	0.374***	0.068
Number of Colons	-0.081*	-0.042	-0.002	-0.047	-0.015
Number of Semicolons	-0.004	0.018	-0.034	-0.004	-0.004
Number of Question Marks	-0.058	0.223***	0.251***	0.017	0.142***
Number of Exclamations	-0.016	-0.159***	-0.029	-0.108**	-0.008
Number of Dashes	0.062	0.037	0.037	0.008	-0.017
Number of Quotation Marks	0.014	-0.015	-0.043	0.035	0.061
Number of Apostrophes	-0.214***	-0.008	0.178***	-0.098**	0.214***
Number of Pairs of Parentheses	-0.035	0.052	-0.126***	-0.004	-0.080*
Number of Other Punctuation	-0.059	-0.051	-0.161***	-0.080*	-0.010
All Punctuation	0.052	-0.015	-0.164***	0.202***	0.065

Note: Pearson correlation coefficients presented. Two-tailed tests: \*\*\*  $p < 0.01$ , \*\*  $p < 0.05$ , \*  $p < 0.1$

Table A-17: Correlations Between LIWC/MRC Features and Jackknifed Big Five Traits (2012)

	Openness	Conscientiousness	Extraversion	Agreeableness	Emotional Stability
<b>MRCPD Categories</b>					
Age of Acquisition	0.119**	0.094*	-0.153***	0.076	-0.173***
Concreteness	-0.095*	-0.303***	-0.249***	-0.059	0.005
Familiarity	-0.155***	0.052	0.302***	-0.017	0.176***
Imagability	-0.109**	-0.371***	-0.221***	-0.111**	0.003
Meaningfulness (Colorado)	-0.308***	-0.315***	0.064	-0.168***	0.052
Meaningfulness (Paivo)	-0.073	-0.179***	-0.043	-0.088*	0.029
Number of Letters	0.209***	-0.080	-0.316***	-0.016	-0.220***
Number of Phonemes	0.189***	-0.107**	-0.295***	0.008	-0.116**
Number of Syllables	0.197***	-0.093*	-0.285***	0.001	-0.153***
<b>LIWC Categories</b>					
Word Count	0.130***	0.109**	0.457***	0.142***	0.364***
Words Per Sentence	0.080	0.037	0.101**	0.008	0.052
Sentences Ending With ?	0.002	0.223***	0.321***	0.084*	0.207***
Unique Words	-0.013	-0.032	-0.384***	-0.007	-0.307***
Dictionary Words	-0.141***	0.077	0.315***	-0.043	0.102**
Words > 6 Letters	0.206***	-0.088*	-0.328***	-0.005	-0.230***
Abbreviations	0.130***	0.114**	-0.106**	0.040	-0.136***
Total Pronouns	-0.248***	-0.055	0.238***	-0.047	0.095*
First-Person Singular	-0.084*	-0.017	-0.121**	-0.089*	-0.191***
First-Person Plural	-0.185***	0.128***	0.280***	-0.108**	-0.025
Total First Person	-0.189***	0.078	0.112**	-0.138***	-0.151***
Total Second Person	-0.171***	0.109**	0.268***	0.100**	0.245***
Total Third Person	-0.065	-0.322***	0.010	0.031	0.188***
Negations	0.003	0.196***	0.210***	-0.008	-0.001
Assents	-0.019	0.041	-0.023	-0.027	-0.094*
Articles	0.191***	0.157***	0.024	0.145***	0.158***
Prepositions	0.090*	-0.074	-0.020	-0.025	0.035
Numbers	-0.102**	-0.106**	-0.148***	-0.032	0.057
Affective Processes	0.017	-0.130***	-0.037	-0.208***	-0.293***
Positive Feelings	0.057	-0.036	0.029	0.002	-0.104**
Optimism and Energy	0.017	-0.160***	-0.106**	-0.111**	-0.145***
Anxiety or Fear	0.140***	-0.005	0.128***	-0.088*	-0.057
Anger	0.156***	0.071	0.063	-0.109**	-0.116**
Sadness or Depression	-0.081*	-0.059	0.021	-0.127***	-0.093*
Cognitive Processes	-0.015	0.311***	0.277***	0.022	-0.030
Causation	-0.057	0.233***	0.234***	0.050	0.117**
Insight	-0.054	0.171***	0.191***	0.034	0.014
Discrepancy	-0.039	0.236***	0.265***	-0.017	-0.033
Inhibition	0.140***	0.214***	-0.069	0.012	-0.218***
Tentative	0.000	0.301***	0.277***	0.081*	0.082*
Certainty	-0.049	0.134***	0.211***	-0.020	0.024
Sensory and Perceptual Processes	-0.140***	0.073	0.279***	0.017	0.221***
Seeing	-0.127***	0.077	0.247***	-0.026	0.170***
Hearing	-0.151***	0.103**	0.254***	0.064	0.223***
Feeling	0.039	-0.105**	0.047	-0.081*	0.016
Social Processes	-0.164***	-0.186***	0.236***	-0.006	0.207***

Continued on next page

Continued from previous page					
	Openness	Conscientiousness	Extraversion	Agreeableness	Emotional Stability
<b>LIWC Categories (continued)</b>					
Communication	-0.055	0.109**	0.234***	0.109**	0.191***
Other References to People	-0.230***	-0.141***	0.284***	-0.018	0.211***
Humans	0.036	-0.056	0.090*	0.002	0.043
Time	-0.042	-0.131***	-0.020	-0.050	0.021
Past Tense Verb	0.011	-0.137***	0.079	0.073	0.293***
Present Tense Verb	-0.170***	0.202***	0.284***	-0.053	-0.030
Future Tense Verb	0.005	0.309***	0.211***	0.032	-0.083*
Space	-0.206***	0.109**	0.123**	-0.018	0.121**
Up	-0.128***	-0.084*	-0.102**	-0.045	0.032
Down	-0.075	0.281***	0.154***	0.034	0.035
Inclusive	0.025	-0.132***	-0.088*	0.016	0.018
Exclusive	-0.054	0.351***	0.360***	0.090*	0.131***
Motion	-0.147***	0.075	0.237***	-0.007	0.224***
Achievement	-0.102**	-0.244***	-0.168***	-0.073	-0.071
Religion	-0.025	-0.264***	0.014	-0.079	0.168***
Sex and Sexuality	-0.060	-0.103**	-0.022	-0.108**	0.020
Swear Words	0.038	0.033	0.139***	0.014	0.101**
Nonfluencies	-0.003	-0.021	0.082*	0.017	0.144***
Fillers	0.095*	0.155***	-0.068	0.166***	0.017
Number of Periods	-0.008	-0.010	-0.136***	0.008	-0.103**
Number of Commas	-0.099**	-0.135***	-0.078	-0.002	0.078
Number of Colons	-0.051	-0.027	0.025	-0.092*	-0.044
Number of Semicolons	0.009	-0.007	-0.055	0.070	0.019
Number of Question Marks	-0.014	0.211***	0.289***	0.080	0.189***
Number of Exclamations	-0.098**	-0.142***	-0.039	-0.128***	-0.001
Number of Dashes	0.015	0.056	0.048	0.047	0.055
Number of Quotation Marks	-0.087*	-0.072	0.023	-0.042	0.022
Number of Apostrophes	-0.111**	0.049	0.107**	0.035	0.142***
Number of Pairs of Parentheses	-0.033	0.027	-0.115**	0.011	-0.046
Number of Other Punctuation	-0.023	-0.036	-0.169***	-0.019	-0.007
All Punctuation	-0.079	-0.043	-0.099**	0.025	0.039

Note: Pearson correlation coefficients presented. Two-tailed tests: \*\*\*  $p < 0.01$ , \*\*  $p < 0.05$ , \*  $p < 0.1$

Table A-18: Correlations Between LIWC/MRC Features and Jackknifed Big Five Traits (2013)

	Openness	Conscientiousness	Extraversion	Agreeableness	Emotional Stability
<b>MRCPD Categories</b>					
Age of Acquisition	0.126***	0.179***	-0.155***	0.109**	-0.107**
Concreteness	-0.012	-0.260***	-0.232***	-0.063	0.004
Familiarity	-0.165***	0.075	0.248***	0.030	0.118**
Imagability	-0.024	-0.357***	-0.222***	-0.118**	-0.038
Meaningfulness (Colorado)	-0.182***	-0.269***	0.066	-0.147***	-0.036
Meaningfulness (Paivo)	-0.145***	-0.204***	-0.090*	-0.147***	-0.012
Number of Letters	0.154***	-0.062	-0.223***	0.003	-0.164***
Number of Phonemes	0.146***	-0.050	-0.242***	0.061	-0.077
Number of Syllables	0.141***	-0.063	-0.238***	0.024	-0.114**
<b>LIWC Categories</b>					
Word Count	0.098**	0.146***	0.478***	0.159***	0.310***
Words Per Sentence	0.104**	0.059	0.165***	0.089*	0.105**
Sentences Ending With ?	0.062	0.199***	0.297***	0.031	0.162***
Unique Words	0.052	-0.099**	-0.348***	-0.000	-0.112**
Dictionary Words	-0.138***	0.120**	0.304***	0.009	0.093*
Words > 6 Letters	0.151***	-0.065	-0.247***	0.003	-0.199***
Abbreviations	0.125**	0.044	-0.148***	-0.018	-0.079
Total Pronouns	-0.204***	-0.018	0.192***	-0.042	0.061
First-Person Singular	0.035	0.059	-0.118**	0.048	-0.077
First-Person Plural	-0.196***	-0.003	0.263***	-0.136***	-0.004
Total First Person	-0.105**	0.040	0.090*	-0.056	-0.058
Total Second Person	-0.154***	0.087*	0.190***	0.080	0.228***
Total Third Person	-0.071	-0.223***	-0.013	-0.001	0.088*
Negations	0.022	0.083*	0.076	-0.071	-0.053
Assents	0.050	0.108**	0.033	-0.018	-0.022
Articles	0.169***	0.225***	-0.048	0.183***	0.150***
Prepositions	0.160***	0.016	-0.121**	0.117**	0.034
Numbers	-0.111**	-0.186***	-0.251***	-0.044	0.022
Affective Processes	0.092*	-0.124**	-0.043	-0.087*	-0.284***
Positive Feelings	0.017	-0.036	0.019	0.035	-0.081*
Optimism and Energy	0.055	-0.113**	-0.032	-0.065	-0.178***
Anxiety or Fear	0.096*	-0.002	0.002	-0.071	-0.076
Anger	0.095*	-0.020	0.057	-0.123**	-0.168***
Sadness or Depression	0.040	-0.017	0.045	-0.061	-0.064
Cognitive Processes	-0.064	0.305***	0.256***	0.042	0.010
Causation	-0.159***	0.113**	0.182***	0.064	0.098**
Insight	-0.094*	0.132***	0.168***	0.041	0.011
Discrepancy	-0.060	0.219***	0.215***	0.020	0.023
Inhibition	0.106**	0.218***	-0.077	0.017	-0.195***
Tentative	0.019	0.306***	0.261***	0.108**	0.121**
Certainty	-0.059	0.130***	0.169***	0.015	0.004
Sensory and Perceptual Processes	-0.151***	0.101**	0.249***	-0.058	0.161***
Seeing	-0.167***	0.044	0.224***	-0.040	0.171***
Hearing	-0.127***	0.120**	0.207***	-0.034	0.121**
Feeling	0.076	-0.010	0.028	-0.072	0.003
Social Processes	-0.161***	-0.121**	0.218***	-0.057	0.092*

Continued on next page



Continued from previous page					
	Openness	Conscientiousness	Extraversion	Agreeableness	Emotional Stability
<b>LIWC Categories (continued)</b>					
Communication	-0.048	0.132***	0.169***	-0.006	0.101**
Other References to People	-0.241***	-0.128***	0.249***	-0.081*	0.126***
Humans	0.116**	0.048	0.068	0.087*	0.016
Time	0.001	-0.103**	-0.058	-0.116**	-0.071
Past Tense Verb	0.013	-0.063	0.033	0.019	0.202***
Present Tense Verb	-0.174***	0.176***	0.323***	-0.072	-0.016
Future Tense Verb	0.028	0.153***	0.073	0.079	0.002
Space	-0.051	0.060	0.153***	-0.029	0.100**
Up	-0.091*	-0.040	-0.062	-0.055	0.014
Down	-0.027	0.198***	0.154***	0.061	0.069
Inclusive	0.038	-0.160***	-0.007	0.013	0.035
Exclusive	-0.102**	0.278***	0.364***	0.068	0.151***
Motion	-0.067	0.081*	0.258***	-0.022	0.126**
Achievement	-0.082*	-0.146***	-0.049	-0.086*	-0.137***
Religion	0.065	-0.183***	-0.021	0.025	0.165***
Sex and Sexuality	-0.140***	-0.116**	0.002	-0.077	0.012
Swear Words	-0.069	-0.011	0.044	0.018	0.005
Nonfluencies	-0.014	0.018	0.013	0.029	-0.014
Fillers	NA	NA	NA	NA	NA
Number of Periods	-0.024	-0.052	-0.227***	-0.069	-0.101**
Number of Commas	-0.031	-0.129***	-0.017	-0.013	0.090*
Number of Colons	-0.006	-0.040	-0.012	-0.103**	-0.065
Number of Semicolons	0.041	0.006	-0.076	0.041	-0.018
Number of Question Marks	0.061	0.183***	0.247***	0.022	0.126***
Number of Exclamations	0.037	0.044	0.023	0.112**	0.076
Number of Dashes	0.049	0.055	0.079	0.005	0.053
Number of Quotation Marks	0.002	-0.057	-0.064	-0.003	0.038
Number of Apostrophes	-0.128***	-0.043	0.100**	-0.114**	0.098**
Number of Pairs of Parentheses	0.149***	0.186***	-0.109**	0.132***	-0.069
Number of Other Punctuation	0.026	-0.053	-0.173***	-0.029	-0.008
All Punctuation	-0.025	-0.066	-0.102**	-0.062	0.029

Note: Pearson correlation coefficients presented. Two-tailed tests: \*\*\*  $p < 0.01$ , \*\*  $p < 0.05$ , \*  $p < 0.1$

Table A-19: Correlations Between LIWC/MRC Features and Jackknifed Big Five Traits (2014)

	Openness	Conscientiousness	Extraversion	Agreeableness	Emotional Stability
<b>MRCPD Categories</b>					
Age of Acquisition	0.095*	0.192***	-0.102**	0.100**	-0.124**
Concreteness	-0.119**	-0.300***	-0.220***	-0.128**	-0.054
Familiarity	-0.153***	0.080	0.237***	0.098*	0.211***
Imagability	-0.150***	-0.377***	-0.198***	-0.225***	-0.109**
Meaningfulness (Colorado)	-0.320***	-0.347***	0.007	-0.233***	-0.033
Meaningfulness (Paivo)	-0.092*	-0.232***	-0.082	-0.089*	0.091*
Number of Letters	0.167***	-0.099*	-0.249***	-0.120**	-0.257***
Number of Phonemes	0.144***	-0.130***	-0.299***	-0.096*	-0.210***
Number of Syllables	0.169***	-0.107**	-0.256***	-0.089*	-0.214***
<b>LIWC Categories</b>					
Word Count	0.165***	0.207***	0.383***	0.196***	0.361***
Words Per Sentence	0.133***	0.050	0.094*	0.063	0.115**
Sentences Ending With ?	0.074	0.219***	0.330***	0.180***	0.212***
Unique Words	-0.041	-0.144***	-0.247***	-0.111**	-0.314***
Dictionary Words	-0.113**	0.130***	0.317***	0.101**	0.139***
Words > 6 Letters	0.184***	-0.088*	-0.268***	-0.127**	-0.290***
Abbreviations	0.079	-0.006	-0.159***	-0.084*	-0.111**
Total Pronouns	-0.296***	-0.046	0.194***	-0.023	0.088*
First-Person Singular	-0.098*	0.076	-0.083*	0.045	-0.118**
First-Person Plural	-0.157***	-0.036	0.217***	-0.118**	-0.023
Total First Person	-0.173***	0.040	0.064	-0.035	-0.108**
Total Second Person	-0.155***	0.061	0.225***	0.096*	0.262***
Total Third Person	-0.166***	-0.252***	0.015	-0.083	0.105**
Negations	0.042	0.154***	0.091*	0.024	-0.082
Assents	-0.061	0.053	-0.003	-0.009	-0.035
Articles	0.224***	0.179***	-0.020	0.217***	0.167***
Prepositions	0.121**	-0.068	-0.036	-0.007	0.014
Numbers	-0.046	-0.038	-0.112**	0.046	0.044
Affective Processes	0.092*	-0.080	0.015	-0.186***	-0.239***
Positive Feelings	-0.075	-0.077	0.104**	-0.150***	-0.043
Optimism and Energy	0.013	-0.139***	-0.086*	-0.174***	-0.172***
Anxiety or Fear	0.082	0.005	0.078	-0.108**	-0.147***
Anger	0.142***	-0.009	0.047	-0.125**	-0.134***
Sadness or Depression	0.066	0.006	-0.053	-0.031	-0.121**
Cognitive Processes	0.007	0.310***	0.229***	0.177***	0.084*
Causation	0.120**	0.241***	0.211***	0.212***	0.125**
Insight	0.040	0.134***	0.176***	0.135***	0.074
Discrepancy	-0.082	0.210***	0.181***	0.070	0.025
Inhibition	0.078	0.171***	-0.050	0.054	-0.098*
Tentative	0.081	0.309***	0.249***	0.214***	0.173***
Certainty	-0.011	0.111**	0.105**	0.074	0.065
Sensory and Perceptual Processes	-0.161***	0.028	0.199***	0.097*	0.313***
Seeing	-0.004	0.123**	0.217***	0.148***	0.252***
Hearing	-0.210***	0.010	0.145***	0.054	0.260***
Feeling	0.008	-0.088*	0.036	-0.002	0.068
Social Processes	-0.224***	-0.202***	0.191***	-0.106**	0.102**

Continued on next page

Continued from previous page

	Openness	Conscientiousness	Extraversion	Agreeableness	Emotional Stability
<b>LIWC Categories (continued)</b>					
Communication	-0.142***	0.017	0.097*	0.072	0.228***
Other References to People	-0.278***	-0.189***	0.228***	-0.111**	0.143***
Humans	0.047	0.003	0.173***	0.038	-0.011
Time	0.021	-0.088*	0.006	-0.067	-0.071
Past Tense Verb	-0.129**	-0.064	0.034	0.066	0.204***
Present Tense Verb	-0.120**	0.204***	0.297***	0.099**	0.109**
Future Tense Verb	0.111**	0.230***	0.066	0.065	-0.035
Space	-0.087*	0.048	0.105**	0.058	0.147***
Up	-0.084*	-0.043	-0.076	-0.006	-0.024
Down	0.017	0.190***	0.067	0.072	0.041
Inclusive	-0.043	-0.263***	-0.076	-0.147***	-0.009
Exclusive	0.020	0.325***	0.351***	0.209***	0.238***
Motion	0.014	0.102**	0.152***	0.075	0.129**
Achievement	-0.021	-0.107**	-0.073	-0.053	-0.155***
Religion	0.002	-0.261***	-0.024	-0.093*	0.039
Sex and Sexuality	-0.156***	-0.161***	0.015	-0.129**	0.069
Swear Words	0.060	0.013	-0.001	-0.015	0.019
Nonfluencies	0.036	0.001	0.093*	-0.017	0.122**
Fillers	0.059	-0.018	-0.024	-0.040	-0.010
Number of Periods	-0.068	0.013	-0.128**	0.027	-0.041
Number of Commas	-0.100**	-0.060	-0.069	0.015	0.073
Number of Colons	-0.045	-0.062	0.041	-0.074	0.005
Number of Semicolons	-0.003	-0.033	-0.024	-0.051	-0.058
Number of Question Marks	0.042	0.192***	0.292***	0.157***	0.190***
Number of Exclamations	0.007	-0.057	-0.030	-0.012	0.020
Number of Dashes	0.072	0.059	0.058	0.036	0.082
Number of Quotation Marks	-0.050	-0.108**	-0.089*	-0.035	-0.009
Number of Apostrophes	-0.073	-0.115**	-0.070	-0.112**	-0.031
Number of Pairs of Parentheses	0.061	0.084*	-0.116**	0.052	-0.045
Number of Other Punctuation	-0.020	0.016	-0.110**	0.019	-0.002
All Punctuation	-0.067	-0.009	-0.114**	0.013	0.009

Note: Pearson correlation coefficients presented. Two-tailed tests: \*\*\*  $p < 0.01$ , \*\*  $p < 0.05$ , \*  $p < 0.1$