Title
Grammar-Based and Lexicon-Based Techniques to Extract Personality Traitsfrom Text

Journal
Proceedings of the Annual Meeting of the Cognitive Science Society, 39(0)

Authors
Carvalho, Maira B.
Louwerse, Max M.

Publication Date
2017

Peer reviewed

Grammar-Based and Lexicon-Based Techniques to Extract Personality Traits from Text

Maira B. Carvalho (m.brandaocarvalho@uvt.nl)
Tilburg Center for Cognition and Communication
Tilburg University
Tilburg, The Netherlands

Max M. Louwerse (mlouwerse@uvt.nl)
Tilburg Center for Cognition and Communication
Tilburg University
Tilburg, The Netherlands

Abstract

Language provides an important source of information to predict human personality. However, most studies that have predicted personality traits using computational linguistic methods have focused on lexicon-based information. We investigate to what extent the performance of lexicon-based and grammar-based methods compare when predicting personality traits. We analyzed a corpus of student essays and their personality traits using two lexicon-based approaches, one top-down (Linguistic Inquiry and Word Count (LIWC)), one bottom-up (topic models) and one grammar-driven approach (Biber model), as well as combinations of these models. Results showed that the performance of the models and their combinations demonstrated similar performance, showing that lexicon-based top-down models and bottom-up models do not differ, and neither do lexicon-based models and grammar-based models. Moreover, combination of models did not improve performance. These findings suggest that predicting personality traits from text remains difficult, but that the performance from lexicon-based and grammar-based models are on par.

Keywords: language; personality; traits; machine learning; computational linguistics; lexicon-based; grammar-based

Introduction

In our daily interactions, we guide our behavior towards other people using information that is collected throughout these interactions, but also using knowledge about the world and social groups (Rich, 1979). These judgments are oftentimes made unconsciously.

Models of users' behavior, thinking and feeling typically rely on the personality traits that can be identified (McCrae & John, 1992). Trait theory is an approach to the study of human personality in which it is believed that humans exhibit habitual patterns of behavior, thought and emotion. It is presumed that there is a relatively small number of dimensions that can be used to describe personality (O'Connor, 2002). Independent analyses have consistently yielded five broad dimensions, called the Big Five (or Five Factor Model): openness to experience, conscientiousness, extraversion, agreeableness and neuroticism (McCrae & John, 1992).

Personality traits are generally identified on the basis of data collected from the users who fill out standardized questionnaires. However, such an approach has certain drawbacks. Firstly, it can be costly for the researcher and time-consuming for the user (Gauch, Speretta, Chandramouli, & Micarelli, 2007). Secondly, people are not reliable sources of information about themselves: there is evidence to suggest that self-descriptions are heavily influenced by the social groups in which a person finds himself (McGuire & Padawer-Singer, 1976), and dissimulation can be a problem in self-reports (Wright, 2014).

Automatic inference of personality offers the advantages of being less intrusive and possibly more environmentally valid. Indeed, a range of studies have investigated the extent to which personality traits can be predicted from user behavior. The lion's share of these studies use linguistic data as sources of information, both language and speech (Beukeboom, Tanis, & Vermeulen, 2012; Gawda, 2009; Mairesse, Walker, Mehl, & Moore, 2007; Mehl, Robbins, & Holleran, 2012; Oberlander & Gill, 2006; Oberlander & Nowson, 2006). Linguistic data has also shown to indirectly shed light on personality. For instance, linguistic cues have shown to be linked to deception (Louwerse, Lin, Drescher, & Semin, 2010), and to different registers of communication (Louwerse, McCarthy, McNamara, & Graesser, 2004). Furthermore, a person's emotional state is reflected in language use (Tausczik & Pennebaker, 2009), not only by explicit lexical content but also by implicit semantic associations (Recchia & Louwerse, 2014).

In considering the cognitive science literature that aims to extract behavioral information from linguistic data, two approaches can be distinguished. On the one hand, studies use lexical cues to extract information from text, for example emotional expression (Kahn, Tobin, Massey, & Anderson, 2007), deception (Newman, Pennebaker, Berry, & Richards, 2003), political orientation (Dehghani, Sagae, Sachdeva, & Gratch, 2013), moral foundations (Graham, Haidt, & Nosek, 2009), romantic relationship outcomes (Ireland et al., 2011), among others. On the other hand, extracting behavioral information from explicit lexical information can be problematic. First, in controlled experimental settings it is easy for participants to carefully monitor their semantic content. For instance, in deception studies participants might avoid using specific words. Second, sparsity issues may emerge if algorithms detect specific word use.

An alternative approach lies in using grammar-based cues. By performing a manual analysis of seven syntax markers, Gawda (2009) identified an increased use of certain features in emotional narratives written by individuals with antisocial personality disorder. Current computational linguistic tools are able to extract many grammar-based linguistic features automatically and efficiently, and it is reasonable to assume that such features could also carry information about personality. Biber (1988) conducted a study on linguistic variation across speech and writing, and computed the frequency of 67 linguistic features (e.g. frequency of auxiliary verbs, pronouns, main verbs, adjectives); he was able to identify different writing genres using naturally occurring word patterns. Graesser, McNamara, Louwerse, and Cai (2004) proposed the tool Coh-Metrix, which allows the user to analyze texts on discourse, cohesion, and world knowledge. The Suite of Linguistic Analysis Tools (SALAT) calculates scores for aspects such as syntactic complexity (Kyle, 2016) and cohesion (Crossley, Kyle, & McNamara, 2015). These tools open up a range of possibilities for the investigation of the relationships between personality and language use.

In conclusion, two approaches can be identified in extracting personality traits from language use: a lexicon-driven approach and a grammar-driven approach. As pointed above, the majority of cognitive science literature has focused on the lexicon-based approach. The question is to what extent the findings from a grammar-driven approach are comparable with the lexicon-driven approach that currently dominates the literature. We address this question in the current work.

Extracting personality traits from text

Most studies on extracting personality from text focused on identifying words, collocations and general linguistic features that occur in texts produced by one group of people versus another group, aiming to uncover which features are informative when trying to differentiate the groups. Early attempts relied on word counting and predefined dictionaries that sort words in categories (Tausczik & Pennebaker, 2009). This approach, albeit basic, has been used in many studies that show links between word usage and certain psychological processes and personalities, e.g. Beukeboom et al. (2012), and Mehl et al. (2012). Other researchers used bottom-up approaches to associate linguistic features with personality types. Oberlander and Gill (2006) collected large corpora of text labeled with the personality of the author and performed stratified corpus comparisons. Interesting findings included the fact that people who scored high in extraversion used more inclusive expressions and connectives, while those with low score were more tentative and used adjectives less frequently. The authors also noted that people with high neuroticism scores had preference for multiple punctuation.

Another approach is to treat the problem as a supervised classification task, employing machine learning techniques to identify the personality of the author of a given text. Oberlander and Nowson (2006) investigated a corpus of weblog posts from 71 participants, who completed a personality questionnaire online as part of the study. The authors used Support Vector Machine (SVM) classifiers and feature sets consisting of n-grams extracted from the text and selected according to different levels of restriction. The same approach was later applied to a larger sample of bloggers (Iacobelli, Gill, Nowson, & Oberlander, 2011). Argamon, Dhawle, Koppel, and Pennebaker (2005) used SVMs and four sets of lexical features to differentiate high and low extraversion and neuroticism, using a corpus of around 2400 student essays and personality assessments, collected by Pennebaker and King (1999). Mairesse et al. (2007) also worked on the same corpus, employing a series of classification and regression techniques and features from both the Linguistic Inquiry and Word Count (LIWC) and the Medical Research Council (MRC) Psycholinguistic Database. Their results confirmed previous findings and reveal new correlations between linguistic markers and personality, such as use of swear words and use of pronouns. As for the accuracy of automatic classification, the authors reported accuracies that are, according to their evaluation, significantly above chance; however, it is not clear whether these values are high enough to be useful in real applications (Mairesse et al., 2007).

Following the work by Mairesse et al. (2007), the task of automatic identification of personality from text gained a lot of attention from the research community, mostly due to the Workshop on Computational Personality Recognition (Celli, Pianesi, Stillwell, & Kosinski, 2013). As part of a shared task, the organizers made available two datasets of text labeled with the personality traits of the authors – including the Essay Corpus by Pennebaker and King (1999). As a result, many researchers tackled the problem with different learning algorithms (e.g. Naive Bayes, SVM, kNN, ensemble methods, logistical regression) and using different features such as n-grams, LIWC, MRC, lexical nuances, part-of-speech tags, emotional values from the AFINN database, word intensity scale, sentiment analysis and word associations to emotions (Celli et al., 2013).

Although the results from these attempts are encouraging, it had been noted that top-down approaches based on lexical resources seem to perform better than bottom-up approaches based only on words or n-grams (Celli et al., 2013). Nevertheless, there are benefits in employing approaches that do not rely on pre-defined vocabularies, for example allowing exploration of topics not previously considered, easier application in different genres and languages, and saving the effort of creating the word lists (Schwartz et al., 2013).

Schwartz et al. (2013) used a large dataset of Facebook posts (over 15.4 million Facebook messages collected from 75 thousand volunteers) to perform an open-vocabulary analysis of correlations between personality types and vocabulary use. The goal of the work was to discover unexpected relationships that would not necessarily be evident from using pre-defined word categories. Although the focus of the work was mainly to explore and gain insights on the data, the authors also used the approach to predict personality from text, with results that are comparable to previous literature. Liu, Wang, and Jiang (2016) also attempted to predict personality from text while avoiding to rely on predefined vocabularies. The authors proposed a model that expands latent Dirichlet allocation (LDA) to include the assumption that topic distribution depends not only on the characteristics of the corpus itself, but also on the five personality traits of writers. However, the topics identified by their model seem to be affected disproportionally by individuals with less common personality combinations, and for this reason the model must be trained with a massive, representative corpus for which the personality of the writers is known. Since obtaining such corpora is difficult, the applicability of their approach seems to be limited.

In this paper, we investigate how we can predict user personality from written text by using features that do not rely on closed-vocabularies, and compare the results to the state of the art. In the next section, we describe our approach.

Procedure

Dataset

We use the Essay Corpus (Pennebaker & King, 1999), which consists of 2468 essays, written by introductory psychology students of the University of Texas as part of their course assignments. The students also completed the Five Factor Inventory personality questionnaire (John, Donahue, & Kentle, 1991), so that all essays could be marked with five personality scores for each Big-5 trait (Openness to Experience, Conscientiousness, Extraversion, Agreableness, Neuroticism). In addition to the scores, the corpus contains binary values for each trait (high/low), which were obtained using a median split over the scores. The class distribution of the binary values is shown in Table 1.

Table 1: Class distribution in dataset.

	OPN	CON	EXT	AGR	NEU
Low	1196 (48.46%)	1214 (49.19%)	1191 (48.26%)	1158 (46.92%)	1235 (50.04%)
High	1272 (51.54%)	1254 (50.81%)	1277 (51.74%)	1310 (53.08%)	1233 (49.96%)

Features

We employed four groups of feature sets, which were chosen to investigate to what extent the performance of lexicon-based and grammar-based computational linguistic methods are comparable.

For the lexicon-based features, we used two main approaches: top-down (LIWC and MRC) and bottom-up (topic modeling with latent Dirichlet allocation (LDA)). For the grammar-based features, we selected the original Biber features (Biber, 1988).

Lexicon-based, top-down

1) A total of 80 LIWC features were extracted using the LIWC2007 software, which outputs relative frequencies of words found in each pre-defined category, and a few structural features such as word count and words per sentence.

2) The 14 MRC features refer to word length, number of syllables or phonemes, and values for frequency of use, imageability, concreteness, meaning, age of acquisition, among others. The MRC features were calculated by averaging the scores of the essay words found in the database (as opposed to averaging over total word count).

Lexicon-based, bottom-up For topic modeling, we preprocessed the corpus by lemmatizing the words using NLTK's WordNet lemmatizer, and removing non-English words (i.e. words that were not found in Wordnet). Given the relatively small size of the corpus, we did not filter out words based on frequency. Then, we trained three LDA models with different number of topics (30, 65 and 100). Each document in the dataset was converted to a vector that represents the proportion in which each topic appears in the document. To train the model, we used the library Gensim, with 10 passes and default hyperparameters.

To illustrate the topics found in the corpus, these are the ten most relevant words for each of the three most frequently appearing topics extracted by the 30-topic LDA model: "think, go, get, really, like, write, minute, time, wonder, need"; "go, get, really, time, friend, home, want, much, like, miss"; and "life, people, thing, know, think, time, one, make, feel, way".

Grammar-based, top-down For this study, we use the 67 features selected by Biber (1988) to reflect the linguistic structure of the text. These features primarily operate at the word level, such as parts-of-speech, and fall into categories such as tense and aspect markers, adverbials, pronouns, questions, nominal forms, passives, subordination features, prepositional phrases, coordinations and negations, and so on. These features were extracted from the text using software developed in-house.

Combinations In addition to considering these models separately, we investigated the model combinations in order to determine their complementary value.

Classifiers

We trained five Support Vector Machine (SVM) classifiers with linear kernel, one for each personality trait. SVMs were chosen due to previous reports of them performing better on this task than other algorithms (Mairesse et al., 2007), and linear kernels were employed to retain interpretability of the model. For the implementation, we used the machine learning library Scikit-learn, which in turn uses an implementation based on Libsvm. The classifiers were trained without parameter tuning (i.e. penalty parameter C=1.0).

Results

Reservations have been expressed by the scientific community on the application of null-hypothesis statistical testing for comparison of machine learning algorithms for many rea-

Table 2: Average accuracy, precision and recall for each classifier, with 95% confidence interval.

		Baseline	Lexicon, top-down		Lexicon, bottom-up			Grammar	Combinations	
		(A) Majority	(B) LIWC+MRC	(C) LIWC	(D) TM (30)	(E) TM (65)	(F) TM (100)	(G) Biber	(H) Biber+LIWC	(I) Biber+TM (30)
Openness	Acc	.515	**.617±.006**	.602±.006	.613±.005	.613±.006	.602±.006	.576±.007	.606±.006	.602±.006
	P	.515	.634±.006	.617±.006	.633±.006	.630±.006	.603±.005	.586±.007	.619±.006	.615±.006
	R	1.	.611±.010	.605±.010	.593±.010	.603±.010	.669±.009	.607±.010	.614±.009	.612±.010
Conscientiousness	Acc	.508	.547±.006	.551±.006	**.556±.007**	.544±.006	.534±.006	.551±.006	.546±.006	.548±.007
	P	.508	.550±.005	.553±.006	.553±.006	.546±.005	.536±.006	.554±.006	.550±.006	.551±.006
	R	1.	.607±.009	.611±.010	.653±.011	.610±.010	.615±.011	.599±.010	.588±.010	.598±.009
Extraversion	Acc	.517	**.562±.007**	.545±.006	.546±.006	.555±.006	.551±.006	.546±.007	.553±.007	.557±.007
	P	.517	.570±.006	.554±.005	.544±.004	.551±.004	.550±.004	.553±.006	.562±.006	.563±.006
	R	1.	.624±.010	.624±.010	.764±.009	.756±.010	.731±.011	.635±.010	.614±.010	.644±.010
Agreeableness	Acc	.531	.545±.007	.552±.007	**.557±.005**	.548±.006	.542±.006	.549±.006	.552±.006	.553±.007
	P	.531	.561±.006	.567±.005	.554±.003	.555±.004	.546±.004	.562±.005	.570±.005	.570±.006
	R	1.	.651±.011	.664±.010	.843±.007	.758±.012	.821±.009	.678±.010	.636±.010	.649±.010
Neuroticism	Acc	.500	.565±.007	**.571±.006**	.532±.007	.527±.007	.520±.006	.545±.006	.552±.006	.545±.007
	P	0.	.563±.007	.571±.007	.529±.006	.527±.008	.519±.006	.545±.006	.552±.006	.544±.006
	R	0.	.577±.012	.571±.011	.581±.014	.531±.012	.540±.011	.535±.010	.551±.011	.550±.011

sons, not the least of which the fact that any difference between two algorithms, no matter how small, can be shown to be statistically significant, provided that enough data are used (Japkowicz & Shah, 2011). For this reason, instead of traditional hypothesis testing, we chose to adopt error-estimation techniques to obtain relatively robust estimates of the performance of the algorithms, which in turn allows us to compare the results considering their practical differences.

Table 2 shows the performance scores of the classifiers trained using the eight different sets of features discussed earlier. We focus our discussion around accuracy (Acc), but we also report precision (P) and recall (R) to give a better overall indication of the performance of the classifiers[1]. The performance of a simple majority classifier (i.e. it always predicts the class with the highest number of instances) is used as baseline. We report the estimated mean of the scores, calculated by running 10 x 10-fold cross-validation, using all 100 individual scores to estimate the mean and variance, and using 10 degrees of freedom to calculate the 95% confidence interval, as suggested by Bouckaert (2003).

As can be seen in Table 2, the performance scores of the classifiers vary for different traits, with the best accuracies ranging from approximately 56% for Agreeableness and Conscientiousness to 62% for Openness to experience (accuracies are highlighted in the table, and the highest accuracy scores for each trait are marked in bold). Nevertheless, we can make some general observations on the overall performance of the different sets of features, which we list below.

Confirming previous findings, top-down lexicon-based approaches generally provide the best accuracies. The top-down approach proposed in the literature, which uses MRC features in addition to LIWC, does provide a small added value for classifying Extraversion and Openness to experience. Conversely, for the other three traits, MRC features do not seem to provide any real improvement.

We note that bottom-up lexicon-based approaches can offer comparable accuracies to top-down approaches, with performance being basically equivalent among top-down and bottom-up over all traits but Neuroticism. Furthermore, the number of topics matters, as accuracy degrades with 100 topics (when the features are likely to become more sparse).

We can observe that a grammar-based approach on its own seems to give a slightly worse accuracy than lexicon-based approaches for three of the five traits: around 2% less accurate for Neuroticism and Extraversion, and 4% less accurate for Openness to experience. Nevertheless, for the other two traits (Conscientiousness and Agreeableness), the accuracies are basically the same.

Finally, combining grammar and lexicon approaches does not lead to significant improvements in accuracy. In fact, it even seems to degrade the results of the top-down lexicon-based approach slightly.

In summary, Table 2 shows that lexicon-based top-down features and bottom-up features do not seem to differ in a practical way, and while grammar-based features seem to have slightly worse accuracies than lexicon-based features, the difference can be considered too small to be of practical significance. Furthermore, the accuracies of our proposed sets of features are on par with the results obtained by previous studies.

[1] Precision and recall scores consider the "high" class as positive label.

General Discussion

The differences in performance between the different algorithms are very small. Using different feature sets yields similar results, and combining different features does not improve the performance in any meaningful way.

One possible reason could be a floor effect, in which the questionnaire used to assess personality traits in this corpus would not be able to distinguish reliably between subjects at the lower end of the scale. This is unlikely, since the test used in this corpus is a standardized questionnaire that has been validated and used in numerous studies (John et al., 1991).

It is also possible that the use of self-assessments of personality makes this task particularly difficult due to the potential unreliability of self-reports, as discussed in the introduction. Future investigation could incorporate personality assessments made by human observers, to evaluate to which extent self-assessment and observed scores differ, and whether the algorithms could match the performance of human judges. Furthermore, replicating the study with other corpora could also indicate whether different text types could be more suitable for detecting certain personality traits.

In this study, we used a relatively limited set of non-top-down features, namely the features proposed by Biber (1988) and topic modeling. Future work could investigate whether applying other grammar-based and bottom-up lexicon-based features (e.g. cohesion, syntactic complexity, n-grams, skip grams, Word2Vec, semantic similarities) would result in better performances. In addition, we could try to improve the models by using non-linear kernels, performing parameter tuning, and employing ensemble machine learning methods for combining different sets of features.

However, the difficulty of identifying personality traits from text could signal a more fundamental issue. Mischel and Shoda (1995) have argued that individual differences in social behaviors are actually variable across different situations (situationism), and not completely stable as it is proposed by trait theory. As such, if personality scales and textual analyses tap into different social situations, tasks that use questionnaire scores as gold standard will not be able to achieve acceptable performance. Further research is needed to investigate this hypothesis, and whether other stable patterns of behavior could be used as gold standard for automatic personality inferences.

The current study has used the most common personality traits classification, the Big Five, and the most commonly used corpus to identify personality traits, the Essay Corpus, in order to compare the difference between top-down and bottom-up lexicon-based and grammar-based computational linguistic techniques. Our findings show that no differences were obtained between lexicon-based and grammar-based or between top-down and bottom-up approaches, nor complementary advantages for combinations of models, despite the fact that all methods were on par with the performance previously reported.

Acknowledgments

We would like to thank James W. Pennebaker for providing us access to the original personality scores of the Essay Corpus.

References

Argamon, S., Dhawle, S., Koppel, M., & Pennebaker, J. W. (2005). Lexical predictors of personality type. In *Proceedings of the 2005 Joint Annual Meeting of the Interface and the Classification Society of North America.*

Beukeboom, C. J., Tanis, M., & Vermeulen, I. E. (2012). The language of extraversion: Extraverted people talk more abstractly, introverts are more concrete. *Journal of Language and Social Psychology*, *32*(2), 191–201. doi: 10.1177/0261927x12460844

Biber, D. (1988). *Variation across speech and writing*. Cambridge: Cambridge University Press. doi: 10.1017/CBO9780511621024

Bouckaert, R. R. (2003). Choosing between two learning algorithms based on calibrated tests. In *Proceedings of the 20th International Conference on Machine Learning (ICML-03)* (pp. 51–58).

Celli, F., Pianesi, F., Stillwell, D., & Kosinski, M. (2013). Workshop on computational personality recognition (shared task). In *Proceedings of the Workshop on Computational Personality Recognition.*

Crossley, S. A., Kyle, K., & McNamara, D. S. (2015). The tool for the automatic analysis of text cohesion (TAACO): Automatic assessment of local, global, and text cohesion. *Behavior Research Methods*, *48*(4), 1227–1237. doi: 10.3758/s13428-015-0651-7

Dehghani, M., Sagae, K., Sachdeva, S., & Gratch, J. (2013). Analyzing political rhetoric in conservative and liberal weblogs related to the construction of the "ground zero mosque". *Journal of Information Technology & Politics*, *11*(1), 1–14. doi: 10.1080/19331681.2013.826613

Gauch, S., Speretta, M., Chandramouli, A., & Micarelli, A. (2007). User profiles for personalized information access. In *The adaptive web* (pp. 54–89). Springer. doi: 10.1007/978-3-540-72079-9_2

Gawda, B. (2009). Syntax of emotional narratives of persons diagnosed with antisocial personality. *Journal of Psycholinguistic Research*, *39*(4), 273–283. doi: 10.1007/s10936-009-9140-4

Graesser, A. C., McNamara, D. S., Louwerse, M. M., & Cai, Z. (2004). Coh-metrix: Analysis of text on cohesion and language. *Behavior Research Methods, Instruments, & Computers*, *36*(2), 193–202. doi: 10.3758/bf03195564

Graham, J., Haidt, J., & Nosek, B. A. (2009). Liberals and conservatives rely on different sets of moral foundations. *Journal of Personality and Social Psychology*, *96*(5), 1029–1046. doi: 10.1037/a0015141

Iacobelli, F., Gill, A. J., Nowson, S., & Oberlander, J. (2011). Large scale personality classification of bloggers. In *Affective computing and intelligent interaction* (pp. 568–577). Springer. doi: 10.1007/978-3-642-24571-8_71

Ireland, M. E., Slatcher, R. B., Eastwick, P. W., Scissors, L. E., Finkel, E. J., & Pennebaker, J. W. (2011). Language style matching predicts relationship initiation and stability. *Psychological Science*, *22*(1), 39–44. doi: 10.1177/0956797610392928

Japkowicz, N., & Shah, M. (2011). *Evaluating learning algorithms: a classification perspective*. Cambridge University Press.

John, O. P., Donahue, E. M., & Kentle, R. L. (1991). *The Big Five inventory – versions 4a and 54.* Berkeley, CA: University of California, Berkeley, Institute of Personality and Social Research.

Kahn, J. H., Tobin, R. M., Massey, A. E., & Anderson, J. A. (2007). Measuring emotional expression with the linguistic inquiry and word count. *The American Journal of Psychology*, *120*(2), 263-286.

Kyle, K. (2016). *Measuring syntactic development in L2 writing: Fine grained indices of syntactic complexity and usage-based indices of syntactic sophistication*. Doctoral dissertation, Georgia State University.

Liu, Y., Wang, J., & Jiang, Y. (2016). PT-LDA: A latent variable model to predict personality traits of social network users. *Neurocomputing*, *210*, 155–163. doi: 10.1016/j.neucom.2015.10.144

Louwerse, M. M., Lin, K.-I., Drescher, A., & Semin, G. (2010). Linguistic cues predict fraudulent events in a corporate social network. In S. Ohlsson & R. Catrambone (Eds.), *Proceedings of the 32nd Annual Conference of the Cognitive Science Society* (pp. 961–966). Austin, TX: Cognitive Science Society.

Louwerse, M. M., McCarthy, P. M., McNamara, D. S., & Graesser, A. C. (2004). Variation in language and cohesion across written and spoken registers. In K. Forbus, D. Gentner, & T. Regier (Eds.), *Proceedings of the 26th Annual Conference of the Cognitive Science Society* (pp. 843–848). Austin, TX: Cognitive Science Society.

Mairesse, F., Walker, M. A., Mehl, M. R., & Moore, R. K. (2007). Using linguistic cues for the automatic recognition of personality in conversation and text. *Journal of Artificial Intelligence Research*, *30*, 457–500. doi: 10.1613/jair.2349

McCrae, R. R., & John, O. P. (1992). An introduction to the five-factor model and its applications. *Journal of Personality*, *60*(2), 175–215. doi: 10.1111/j.1467-6494.1992.tb00970.x

McGuire, W. J., & Padawer-Singer, A. (1976). Trait salience in the spontaneous self-concept. *Journal of Personality and Social Psychology*, *33*(6), 743–754.

Mehl, M. R., Robbins, M. L., & Holleran, S. E. (2012). How taking a word for a word can be problematic: Context-dependent linguistic markers of extraversion and neuroticism. *Journal of Methods and Measurement in the Social Sciences*, *3*(2), 30–50. doi: 10.2458/azu_jmmss_v3i2_mehi

Mischel, W., & Shoda, Y. (1995). A cognitive-affective system theory of personality: reconceptualizing situations, dispositions, dynamics, and invariance in personality structure. *Psychological Review*, *102*(2), 246–268. doi: 10.1037/0033-295X.102.2.246

Newman, M. L., Pennebaker, J. W., Berry, D. S., & Richards, J. M. (2003). Lying words: Predicting deception from linguistic styles. *Personality and Social Psychology Bulletin*, *29*(5), 665–675. doi: 10.1177/0146167203029005010

Oberlander, J., & Gill, A. J. (2006). Language with character: A stratified corpus comparison of individual differences in e-mail communication. *Discourse Processes*, *42*(3), 239–270. doi: 10.1207/s15326950dp4203_1

Oberlander, J., & Nowson, S. (2006). Whose thumb is it anyway?: classifying author personality from weblog text. In *Proceedings of COLING/ACL* (pp. 627–634). Sidney, Australia.

O'Connor, B. P. (2002). A quantitative review of the comprehensiveness of the five-factor model in relation to popular personality inventories. *Assessment*, *9*(2), 188–203. doi: 10.1177/1073191102092010

Pennebaker, J. W., & King, L. A. (1999). Linguistic styles: language use as an individual difference. *Journal of Personality and Social Psychology*, *77*(6), 1296–1312. doi: 10.1037/0022-3514.77.6.1296

Recchia, G., & Louwerse, M. M. (2014). Reproducing affective norms with lexical co-occurrence statistics: Predicting valence, arousal, and dominance. *The Quarterly Journal of Experimental Psychology*, *68*(8), 1584–1598. doi: 10.1080/17470218.2014.941296

Rich, E. (1979). User modeling via stereotypes. *Cognitive Science*, *3*(4), 329–354. doi: 10.1207/s15516709cog0304_3

Schwartz, H. A., Eichstaedt, J. C., Kern, M. L., Dziurzynski, L., Ramones, S. M., Agrawal, M., … Ungar, L. H. (2013). Personality, gender, and age in the language of social media: The open-vocabulary approach. *PLoS ONE*, *8*(9). doi: 10.1371/journal.pone.0073791

Tausczik, Y. R., & Pennebaker, J. W. (2009). The psychological meaning of words: LIWC and computerized text analysis methods. *Journal of Language and Social Psychology*, *29*(1), 24–54. doi: 10.1177/0261927x09351676

Wright, W. R. (2014). Personality profiling from text and grammar. In *User Modeling, Adaptation, and Personalization* (pp. 502–507). Springer. doi: 10.1007/978-3-319-08786-3_47