Journal of Research in Personality 47 (2013) 417–420

Contents lists available at SciVerse ScienceDirect

Journal of Research in Personality

journal homepage: www.elsevier.com/locate/jrp

Frequency of the use of English personality adjectives: Implications for personality theory

Eka Roivainen

Verve Rehabilitation, Department of Neuropsychology, Kasarmintie 13, 90101 Oulu, Finland

ARTICLE INFO

Article history:
Available online 16 April 2013

Keywords:
Personality theory
Word frequency
Big Five
Lexical hypothesis

ABSTRACT

According to the lexical hypothesis, the degree of representation of an attribute in language correlates with the general importance of that attribute in everyday transactions. The present study investigated the frequency of the use of 432 personality adjectives that Saucier and Goldberg (1996) employed as the base of their five-factor theory of personality. Google hits for bigrams consisting of a personality adjective + *person* varied from 51 (*uninquisitive person*) to 4.2 million (*reasonable person*). The 92 adjectives that describe agreeableness (factor II) had 29 million hits, while the 40 adjectives describing neuroticism (factor IV) had 6.9 million hits. Historical analysis showed growing popularity for factor II adjectives between 1950 and 2000. These results indicate that the Big Five factors of personality are not of equal importance in everyday personality descriptions.

1. Introduction

Major trait theories of personality are based on the so-called lexical hypothesis. It is assumed that topics considered important in everyday life, including personality characteristics, are described in everyday language using words (Klages, 1932). Thus, personality traits can be determined by using the dictionary method, where a large pool of personality descriptors is collected from lexical sources and these terms are then classified into larger categories (John, Angleitner, & Ostendorf, 1988). Allport and Odbert (1936) extracted 17,953 trait names from dictionaries and identified 4505 stable traits. Using factor analysis, Cattell reduced Allport's list to 16 major personality traits (Cattell, Marshall, & Georgiades, 1957). Prevailing, dominant trait theories (Goldberg, 1990; McCrae & Costa, 1999) propose five major factors of personality: I extraversion, II agreeableness, III conscientiousness, IV neuroticism, and V openness/intellect. Goldberg's (1990) Big Five factors are based on a factor analysis of 492 adjectives selected from Norman's (1967) list of 2797 terms, from which Goldberg eliminated all nouns and other words on grounds of "ambiguity, difficulty, slanginess, sex-linkage, over-evaluation, metaphoricalness, and redundancy" (Saucier & Goldberg, 1996).

One of the main criteria for the validity of a personality theory is that factors of the model should be socially relevant (Eysenck, 1991). According to Saucier and Simonds (2006), the degree of representation of an attribute in language correlates with the general importance of that attribute in everyday transactions. "An attribute represented by multiple terms will likely appear as factor in multivariate analysis. If the factor includes terms used with high frequency then the importance of the factor is underscored" (Saucier & Simonds, 2006). However, Block (2010) noted that the number of synonyms for a word does not always testify to the psychological importance of that word, and it is not necessary to redundantly express certain crucial concepts. Block (1995) criticized Goldberg's (1990) Big Five model because of the problems involved in selecting the pool of items. For example, Goldberg's adjective list of 492 terms included terms such as "imperturbable" and "indefatigable", which are relatively uncommon words.

In developing the five-factor model, researchers have employed panels of judges to estimate the social relevance of different traits. The paradigmatic method is a two-step procedure. Responding to Block's (1995) critique, Saucier and Goldberg (1996) employed 25 student judges to first estimate the familiarity of Goldberg's (1990) 492 personality adjectives on a 0–9 scale. Following this, they eliminated the 57 least popular words, resulting in a pool of 435 terms. Thereafter, they presented longer inventories of 566–1710 words to a larger sample of students ($n = 889$) for peer and self-assessment of personality and performed a factor analysis of the 435 adjectives of the reduced pool. Angleitner, Ostendorf, and John (1990) relied on 10 judges to classify personality descriptors found in German dictionaries. A sample of 400 adult participants then completed self-reports using a list of the 430 most familiar and relevant adjectives. Szarota (1996) employed 10 judges to rate 1839 Polish words for clarity of meaning and personality relevance, and 369 students used a list of 290 adjectives for self- and peer-ratings. However, according to Block (2010), the resulting five factors are not of equal behavioral importance, and the first two or three

E-mail address: eka.roivainen@verve.fi

0092-6566/$ - see front matter © 2013 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.jrp.2013.04.004

trait factors in the five-factor-structure are more clear and consequential than the remaining three or two.

Arguably, the simplest and most reliable method for measuring the relevance of personality adjectives and factors is to measure the general popularity of the personality adjectives that describe the factors. Because many personality adjectives, such as *friendly*, are highly common words that may be used to describe a multitude of concepts, simple word frequency rankings reveal little regarding their popularity as specific personality adjectives. The frequency of strings of words (n-grams), where an adjective is used as a personality attribute, such as *friendly man* or *friendly woman*, should be studied for this purpose. In the internet era, the study of n-grams is straightforward. A Google search for *Raymond Cattell* gives 92,000 results, and a search for *Leopold Szondi* gives 7000 hits. We may conclude that the social relevance of Cattell is higher than that of Szondi. Using a sample of psychology undergraduates to rate the popularity of these two personality theorists would likely lead to the same conclusion, but a panel of geology students may judge the two as equally irrelevant persons. Thus, the use of Google to study n-grams is likely to be more representative of the general population than restraining analyses to that obtained from a typically-researched university sample, such as a group of undergraduates.

In the present study, the Google Ngram viewer and Google Search were used to examine the usage frequencies of the 435 personality adjectives from Saucier and Goldberg's (1996) study.

2. Method

A Google search was performed for bigrams consisting of one of the personality adjectives from Saucier and Goldberg's (1996) list, with the word *person* added to these adjectives, such as *kind person*, *rude person*. The adjective *frank* was excluded from the analysis because Frank Person is a German designer, and *down-to-earth* and *happy-go-lucky* were excluded because of spelling issues. For the purpose of Google search, all other adjectives were written together (*easygoing* and not *easy-going*).

The Google Ngram viewer (Michel & et al., 2011) is a phrase-usage graphing tool that charts the yearly count of combinations, such as words or phrases, found in more than 5 million books digitized by the Google Corporation. A Google Ngram analysis was performed for the 100 adjectives with the highest number of hits in a simple Google search. American English texts for the years 1950 through 2000 were searched with 5-year smoothing.

3. Results

Table 1 shows that there are significant differences in the usage frequencies of common personality adjectives. The top ranking adjectives included *responsible*, *honest*, *friendly*, *reasonable*, and *simple* with 3–4 million Google hits, and a 30–550/10^7% Google Ngram frequency. The lowest-ranking adjectives were *uninquisitive*, *fault-finding*, *exhibitionistic*, *negativistic*, and *nonconforming* with 51–500 Google hits and a 0–0.4/10^7% Ngram frequency. The correlation between the primary factor loading reported by Saucier and Goldberg (1996) and Google hits for all the 432 adjectives was 0.13. Nevertheless, Table 2 shows that personality adjectives with at least a moderate loading (>0.29) on one of the Big Five factors in Saucier and Goldberg's (1996) study accounted for most Google hits, whereas "outlier" adjectives (<0.30) were less popular. Factor II appears to be the most important factor, with nearly four times as many hits as factor IV. Interestingly, the mean frequency of use per adjective appears to be the highest for group V adjectives.

There have been substantial changes in the popularity of specific adjectives. For example, the use of the term *reasonable persons* has increased by almost 300% in books from 1950 to 2000 while the use of the term *intelligent persons* has decreased by half. Table 3 shows that the popularity of factor II adjectives as a group has increased significantly from the year 1950 to 2000, while the popularity of the other groups has remained relatively constant. Correlations between Google hits for the year 2012 and Ngram frequency for the years 1950 and 2000 were 0.54 and 0.58, respectively, and the correlation between Ngram frequency in the years 1950 and 2000 was 0.65.

Moreover, the results of these analyses clearly showed the linguistic positivity bias ("Pollyanna effect"), which is the tendency to use positive words ("pretty") more often than equally familiar negative words ("ugly"), as first reported by Boucher and Osgood (1969) and Augustine, Mehl, and Larsen (2011). Only nine of the top 50 and 25 of the top 100 adjectives described personality through negation, for example *shy* as a descriptor of extraversion. Only *lazy*, *ignorant*, *suspicious*, *absentminded*, and *selfish* may be considered as clearly socially undesirable descriptors among the top 50 adjectives.

4. Discussion and conclusions

The results of the present study are roughly consistent with the five-factor models. If an adjective with a primary factor loading of 0.30 or less is considered an outlier, then the Big Five factors adjectives account for 80–90% of personality adjective usage. However, the low correlation between factor loading and usage frequency suggests that the theoretical model and our practical nonprofessional psychological theory of personality have different goals. *Kind* and *sympathetic* are equally valid descriptors of agreeableness in terms of their primary factor loadings in Saucier and Goldberg's study (0.60–0.62 on factor II, and <0.07 secondary loadings), but based on their popularity, *kind* (2,600,000 hits) is a significantly better descriptor compared to *sympathetic* (97,000 hits). Arguably, factor II should be named after the adjective *kind* and not after *agreeable*, which is a less popular term (67,000 hits) and has a lower loading (0.46) on this factor. If the lexical hypothesis is correct, and we can assume that everyday language reflects human personality structure, then it can be argued that factor analysis provides more information regarding inter-concept relationships, while usage frequency is a better measure of the practical importance and the descriptive power of the terms.

The results of the present study agree with the findings by Ames and Bianchi (2008), who established the concept of agreeableness asymmetry. In multiple studies, using photographs, videos, and face-to-face encounters, Ames and Bianchi (2008) found that agreeableness was generally more prevalent than the other Big Five traits in open-ended descriptions of early impressions. However, agreeableness judgments are also less accurate than the judgments of other traits. Using a list of 339 adjectives (Goldberg, 1990), Ames and Bianchi found a roughly 21–45% prevalence of agreeableness adjectives in the descriptions, while the adjectives for neuroticism had a prevalence of only 1.4–7%. They also noted that manipulating the power of perceivers relative to targets substantially shifted the content of impressions between agreeableness and conscientiousness. Fiske, Cuddy, and Glick (2006) suggested that social perception reflects evolutionary pressures. In encounters with others, social animals must immediately determine whether the other intends good or ill, and then whether the other has the ability to enact those intentions. According to Srivastava (2010), the Big Five factors of personality are "dimensions of grounded social perception reflecting the social concerns of perceivers". Knowing a target's personality allows the perceiver to form probabilistic expectations regarding how the person will think, feel, or behave in future or in novel circumstances.