Taxonomy of Person-Descriptive Terms in Polish: A Psycho-Lexical Study

Taksonomia deskryptorów osobowych w języku polskim: studium psycholeksykalne

Oleg Gorbaniuk, Ewelina Czarnecka, Marta Chmurzyńska

Katedra Psychologii Eksperymentalnej, Instytut Psychologii Katolickiego Uniwersytetu Lubelskiego Jana Pawła II

Abstract

This paper describes the development of a comprehensive taxonomy of person-descriptive terms in Polish, organized in two studies. In the first study, two judges searched through The Universal Dictionary of Polish Language (100 000 terms) for person-descriptive terms. In the second study, 4555 person-descriptive adjectives were classified by nine judges into 13 different subcategories of the psycho-lexical classification system. Our studies provide comprehensive and representative lists of Polish adjectives for describing personality traits, which may be used in establishing the structure of human personality descriptions by means of self-rating or peer-rating surveys of a full age range Polish population of respondents.

Keywords: lexical approach, dispositional terms, Polish adjectives, Big Five

Introduction

Fundamental to research on the structure of dispositions ascribed to people in personality psychology is the so-called lexical hypothesis, first formulated by Francis Galton in 1884, according to which the most meaningful and socially important personality differences between people are coded in language [1]. Consequently, the analysis of the content of language makes it possible to isolate the most important categories of human traits – in other words, it enables taxonomy of these traits. Systematic research on the issue was initiated by Allport and Odbert, who identified 17 953 person-descriptive terms and narrowed that down to 4 504 observable descriptors that describe relatively stable dispositional traits. Cattell categorized these descriptors according to semantic similarity and, using factor analysis with oblique rotation, identified a dozen or so correlated factors [2]. These may be treated as superordinate categories, and further stages of the reduction – as higher and higher levels of abstraction.

Fiske went even further in his investigations and, using factor analysis with orthogonal rotation, isolated five recurring factors in descriptions of oneself as well as in descriptions of others people: (1) Social Adaptability, (2) Emotional Control, (3) Conformity, (4) The Inquiring Intellect, and (5) Confident Self-Expression [3]. As research material, 22 semantic differential scales were used, constructed based on the factors isolated by Cattell [2]. A characteristic feature of the above studies – resulting in a structure similar to that which is now known as the Big Five – was their foundation on the broadest possible lexical material and its systematic reduction by means of factor analysis. Their drawback lay in the excessively high number of intermediate stages, which might have distorted the final effect. Factor analysis is considered here as a variable-reduction procedure in which many variables (dispositional descriptors) are organized by a few factors that summarize the interrelations between them [4]. These factors may be considered as aggregate constructs or as higher-order dimensions in a hierarchic model.

Even though in the 1960s, in studies on adjectival material, similar five-factor structures of perceived dispositions were obtained again for descriptions of family and
friends [5,6] as well as strangers [7], these findings did not generate widespread interest. Of decisive importance for the research adopting the lexical approach in trait theory was Goldberg's publication [8], and its key to success largely consisted in applying lexical taxonomy (cf. the opinion of Saucier & Goldberg [9]), which allowed to distinguish person-descriptive terms describing dispositions from those describing states, social judgments, performed roles, or outward appearance [10,11].

Goldberg's study of the American population demonstrated that the structure of descriptions of oneself (self-rating) and of others (peer-rating) is identical in terms of the number of factors and their content [8]. The following factors were distinguished: (1) Extroversion, (2) Emotional Stability, (3) Conscientiousness, (4) Agreeableness, and (5) Intellect. This study initiated the analogous ones in other countries, aimed at verifying the cultural universality of the Big Five. In a majority of cases, they were carried out in three stages: (1) the selection of person-descriptive lexical items of the specified type – adjectives, nouns, or, less often, verbs – from a dictionary of a given language; (2) the categorization of person-descriptive terms by competent judges, aimed at identifying dispositional descriptors; (3) quantitative research using the list of dispositional descriptors, aimed at determining the dimensions of perceived human personality through self-rating and peer-rating. In the lexical approach in trait theory, a dispositional descriptor is understood as a linguistic item suitable for describing a human person, referring to a characteristic that is stable over time and psychological in nature [9].

In Poland, psycho-lexical research on person-descriptive adjectives was carried out by Szarota [12,13]. Based on The Concise Polish Dictionary [14] containing 35,000 entries, 1,811 person-descriptive adjectives were selected. Supplemented by additional 21 adjectives from other sources, the collected material was categorized by 10 judges. As a result, 290 dispositional descriptors were determined and later used in quantitative research. A drawback of the taxonomy procedure, pointed out by the author himself, was the use of an incomplete Polish language lexicon, which may have resulted in the omission of some personality-descriptive adjectives. For this reason, it was decided that Polish psycho-lexical research needed replication, free from the above drawbacks and providing a better foundation for future quantitative research on the structure of perceived human dispositional traits.

Method

The Selection of Adjectives from a Polish Language Dictionary

After consulting the research workers employed at the Institutes of Polish Studies of the John Paul II Catholic University of Lublin and Maria Curie-Skłodowska University in Lublin, the four-volume Universal Dictionary of the Polish Language, compiled and published in 2008, was chosen as the most complete and up-to-date source of Polish lexical material [15]. This is the largest contemporary Polish language dictionary, which registers the vocabulary of the second half of the 20th century and early 21st century. It contains around 100,000 entries representing all the varieties of Polish, including the latest vocabulary.

The selection of person-descriptive adjectives was carried out by two judges – co-authors of the present article. They had been trained to do it, and the correctness of their selection was tested on a sample of 400 adjectives. This allowed to identify and discuss the most frequent mistakes. Thus prepared, the judges were ready to carry out the selection of person descriptors from the dictionary. When in doubt regarding whether or not a given term should be considered useful in describing human traits, the judges made use of heuristics in the form of the following sentences: (1) How [adjective] am I? (2) How [adjective] did John behave? In cases of doubts as to whether or not a given term should be considered person-descriptive, the judges were instructed to include it in further stages lexical analyses, where their person-descriptive character would be assessed by a larger group of judges. Based on dictionary definitions, four groups of person-descriptive adjectives were disregarded: (1) adjectives connected with the political, philosophical, or literary context and those related to art or the religious context; (2) adjectives describing national status, geographical background, or profession; (3) adjectives defined as colloquial or vulgar; (4) archaic adjectives. The selection of lexical material took the judges about five months. The adjectives collected were verified by the first co-author of the article. The final list of terms, which later formed the basis of lexical taxonomy, comprised 5,313 adjectives.

Lexical Taxonomy Procedure

The adjectives selected from the dictionary were assessed by 9 competent judges. Apart from the authors of this article, 6 psychology students also acted as judges. They were simultaneously involved in research aimed at determining the dimensions of perceived personality traits ascribed to politicians, political parties, and their voters, whose procedure was similar to that of psycho-lexical research. Their motivation to perform the classification reliably should be considered high, since all the judges were interested in the success of the whole undertaking. Before the judges began the classification procedure, they had undergone theoretical training as part of classes as well as taken part in a 4-hour session of judging 100 adjectives together, during which the definitions of individual categories were clarified. Next, each of the judges classified about 200 words on their own. The correctness of the classification was then verified and the more serious mistakes were
individually discussed in order to be eliminated in the future. This was to be achieved through working out a more precise understanding of category definitions.

In own research, we adopted the taxonomy of adjectives from a German lexical study [16], being one of the most frequently used taxonomies in lexical research, including earlier research conducted in Poland [13]. The first step was an estimation of the extent to which the meaning of the adjective was clear. The words were rated on the following 3-point scale: 1 – the meaning of the word is not clear enough for me to complete subsequent ratings, 2 – the meaning of the word became clear to me only after giving it some thought, 3 – the meaning of the word is fairly clear to me. If the meaning was clear enough (3), the judge moved to the second step, in which he or she had to decide whether the adjective could be imagined being used for the description of an individual or for the description of an individual’s experience, behaviour, or appearance. The words were rated on a 3-point scale. If the judge responded ‘1’ (impossible to imagine) or ‘2’ (unusual; possible to imagine only under certain conditions), the term was considered not personality relevant. Next, the valence of the adjective was rated on a 5-point scale (very negative, negative, neutral, positive, very positive).

Each adjective rated by the judge as person-descriptive (easy to imagine as a person descriptor) was then supposed to be classified into one of 13 subcategories, which in turn fall into 5 superordinate categories: (1) Dispositions, (2) Temporary conditions, (3) Social and reputational aspects, (4) Overt characteristics and appearance, (5) Terms of limited utility [16]. Category 1 was divided into two subcategories: subcategory (1a) was reserved for temperament and character traits and subcategory (1b) for abilities and talents or their absence. Category 2 contained three subcategories: (2a) experiential states (emotions, moods, and cognitions), (2b) physical and bodily states, and (2c) observable activities. Category 3 consisted of four subcategories: (3a) roles and relationships, (3b) social effects (reactions of others), (3c) pure evaluations, and (3d) attitudes and world views. Category 4 was divided into two subcategories: (4a) bodily characteristics pertaining to anatomy, constitution, and morphology, and (4b) socio-cultural aspects of appearance (looks and deportment). The fifth category consisted of two subcategories: (5a) context-specific and technical terms and (5b) metaphorical, vague, or outmoded terms. A given word was classified as belonging to a particular category if it was indicated as such by at least 5 judges.

In keeping with the prototype theory, the borderlines of categories were assumed to be blurred in the study. This meant that some adjectives (also because of their polysemy) could simultaneously belong to several categories. Therefore, if necessary, the judges could classify a given word into two different categories, e.g. agresywmy (aggressive) as a person-descriptive term could refer to either a trait or a state. The entire judging process took 8 weeks.

Results

The Proportion of Person-Descriptive Adjectives

Out of 5 313 adjectives, 5 134 words (96.6%) were rated by the majority of judges (5 out of 9) as clear to them. Out of those 5 134 words, 4 555 (88.7%) were found by most judges to be person-descriptive terms (α=0.74). These proportions are comparable to those obtained in a German lexical study, in which 4 827 person-descriptive adjectives were selected based on a dictionary containing 96 644 entries [16]. Out of a set of 70 000 entries, a study of the Hungarian lexicon isolated 3 644 adjectives allowing to differentiate between individuals [17]. Person-descriptive adjectives thus appear to constitute about 5% of the lexicon of languages representing various language families/groups: Germanic, Finno-Ugric, and Slavic. Norman gives similar estimations for the English language [11].

In an earlier Polish psycho-lexical study, the proportion of adjectives classified as person-descriptive was 5.2% [13], though their overall number was lower (1 811) because the study was based on a dictionary with nearly three times fewer entries than the one used in own studies.

The Reliability of the Judges’ Opinions

The agreement of the judges’ opinions was measured using Cronbach’s alpha internal reliability coefficient. The analysis of how individual judges influenced the overall reliability of the group of judges allowed to eliminate one of them, whose classification decisions markedly differed from those of the others, lowering Cronbach’s α. Table 1 also contains data concerning mean correlation between judges. Mean correlation provides additional information about the agreement of classification decisions irrespective of the number of judges, which Cronbach’s α is sensitive to [18]. The highest agreement of opinions was observed for the valence of adjectives, where Cronbach’s alpha was 0.93 and mean correlation equaled 0.69. The values of Cronbach’s alpha internal reliability coefficient ranged from 0.71 to 0.91 for the five superordinate lexical categories and from 0.58 to 0.89 for their 13 subcategories.

The lowest reliability coefficients were found for the “Terms of limited utility” category, the least relevant one in our study, comprising specialist terms, rarely used words, and metaphorical adjectives. The agreement of judges’ opinions was the lowest for this category, and the subjectivity of opinions may have been primarily due to the vagueness of its definition. In the German lexical study [16] from which the taxonomy was borrowed, the reliability index for the “Metaphorical, vague, and outmoded” subcategory was
Table 1. Frequencies and percentages of adjectives classified into the categories, and inter-judge reliabilities of the classifications

<table>
<thead>
<tr>
<th>Category/subcategory</th>
<th>Inter-judge agreement</th>
<th>Frequency</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>α</td>
<td>r</td>
<td></td>
</tr>
<tr>
<td>1. Dispositions</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1a. Temperament and character</td>
<td>0.89</td>
<td>0.54</td>
<td>714</td>
</tr>
<tr>
<td>1b. Talents and abilities</td>
<td>0.82</td>
<td>0.40</td>
<td>605</td>
</tr>
<tr>
<td>2. Temporary conditions</td>
<td>0.89</td>
<td>0.53</td>
<td>625</td>
</tr>
<tr>
<td>2a. Experience and states</td>
<td>0.87</td>
<td>0.50</td>
<td>380</td>
</tr>
<tr>
<td>2b. Physical states</td>
<td>0.88</td>
<td>0.53</td>
<td>181</td>
</tr>
<tr>
<td>2c. Observable activities</td>
<td>0.76</td>
<td>0.33</td>
<td>44</td>
</tr>
<tr>
<td>3. Social aspects</td>
<td>0.83</td>
<td>0.41</td>
<td>2309</td>
</tr>
<tr>
<td>3a. Roles and relationships</td>
<td>0.86</td>
<td>0.48</td>
<td>427</td>
</tr>
<tr>
<td>3b. Social effects</td>
<td>0.83</td>
<td>0.43</td>
<td>78</td>
</tr>
<tr>
<td>3c. Pure evaluations</td>
<td>0.80</td>
<td>0.37</td>
<td>1453</td>
</tr>
<tr>
<td>3d. Attitudes and worldviews</td>
<td>0.89</td>
<td>0.54</td>
<td>59</td>
</tr>
<tr>
<td>4. Overt characteristics</td>
<td>0.91</td>
<td>0.61</td>
<td>429</td>
</tr>
<tr>
<td>4a. Anatomy and constitution</td>
<td>0.78</td>
<td>0.34</td>
<td>89</td>
</tr>
<tr>
<td>4b. Appearance, looks etc.</td>
<td>0.85</td>
<td>0.45</td>
<td>301</td>
</tr>
<tr>
<td>5. Terms of limited utility</td>
<td>0.71</td>
<td>0.30</td>
<td>209</td>
</tr>
<tr>
<td>5a. Context-specific or technical</td>
<td>0.73</td>
<td>0.33</td>
<td>114</td>
</tr>
<tr>
<td>5b. Metaphorical, vague or outmoded</td>
<td>0.58</td>
<td>0.21</td>
<td>47</td>
</tr>
<tr>
<td>6. Non-classifiable words</td>
<td>-</td>
<td>-</td>
<td>55</td>
</tr>
<tr>
<td>Number of terms without majority classifications</td>
<td>-</td>
<td>-</td>
<td>560</td>
</tr>
<tr>
<td>Number of terms in the initial pool</td>
<td>-</td>
<td>-</td>
<td>4555</td>
</tr>
</tbody>
</table>

The largest category turned out to be Social aspects (50.7%) and one of its subcategories, Pure evaluations (31.9%), including adjectives such as sympatyczny (likeable), prostacki (boorish), głupi (stupid), nudny (boring), and dziwaczny (bizarre). The Dispositions category (15.7%) ranked second in terms of frequency, Temporary conditions (13.7%) ranked third, Overt characteristics (9.4%) took the fourth place, and Terms of limited utility (4.6%) took the last one.

From the viewpoint of psycho-lexical research, the most interesting category is the adjectives classified under Dispositions, because it is them that constitute the potential source of the adjective list to be used in quantitative research aimed at determining the structure of the adjectives that Poles ascribe to themselves (self-rating) or to persons from their closest circle (peer-rating). The proportion of dispositional adjectives among person-descriptive adjectives (15.7%) is very similar to that in the earlier Polish psycho-lexical study, where it was found to be 16.0%, but the earlier study was based on a smaller initial set of person-descriptive adjectives and a correspondingly lower number of dispositional adjectives (290 words). For comparison, in German [16] and Croatian [19] studies the proportion of dispositional adjectives among person-descriptive adjectives was estimated at 12%, and in the Czech Republic at only 9% [19].
Dispositional Adjectives List

As mentioned before, the judges had the additional task of evaluating the sign of the adjectives on a 5-point scale, where 1 and 2 stood for negative evaluation, 3 for neutral, and 4 and 5 for positive. After averaging their opinions, the adjectives rated between 1.0 and 2.74 were classified as negative, those rated between 2.75 and 3.25 as neutral, and those rated between 3.76 and 5.0 as positive. Thus, it was established that 50.2% of dispositional adjectives describe socially undesirable traits, 12.0% are neutral in this respect, and 37.8% are judged as positive by the environment. The higher percentage of adjectives with negative valence results from the fact that approximately 14% of them were formed by negating adjectives with positive valence, e.g. moralny → amoralny (moral → amoral), sumienny → niesumienny (conscientious → unconscientious), or zrównoważony → nierównoważony (stable → unstable), whereas antonyms of adjectives with negative valence are formed in this way much less frequently, e.g. złośliwy → niezłośliwy (malicious → benevolent), konfliktowy → niekonfliktowy (confrontational → unconfonntational). In other languages dispositional adjectives with negative valence predominate as well [20,21,22].

The next step in psycho-lexical research is the preparation of a list of dispositional terms/adjectives for quantitative research. Not all of the 714 adjectives defined as describing the dispositions of individuals qualify for such research, since some of them are morphemically and at the same time semantically alike while others may be difficult for an average respondent to understand. Therefore, the list was reduced by eliminating adjectives with identical morpheme and meaning (e.g. mściciel i mściwy – vindictive and vengeful; niedostępny and nieprzystępny – inaccessible and unapproachable), and a survey was conducted on a sample of 50 students (46% male and 54% female) whose task was to indicate adjectives whose meaning they found unclear. The survey identified a group of 38 dispositional adjectives which were found unclear by more than 10% of the respondents (see Appendix 2). The final list of 661 dispositional adjectives that may be used in establishing the structure of the perception of individuals in the Polish population in dispositional terms are presented in Appendix 1.

Conclusions and implications

Apart from the lexical hypothesis (which, strictly speaking, is not a hypothesis from the methodological point of view), the way of defining a dispositional descriptor is only a theoretical assumption that precedes principal component analysis, aimed at identifying the basic dimensions of perceived human dispositions, and determines its outcome as well as the interpretation of this outcome. Human traits are not restricted to dispositions only. In describing them, not only emotional states are important but also external attributes such as performed roles or social status, interpersonal attractiveness, outward appearance, world view, as well as reactions and social judgments that all these elicit from the environment. In collectivist cultures, external attributes often play a more important role in the formation of public opinion than those which are given the status of internal attributes [23]. Person-descriptive terms have the status of natural terms with blurred borderline, which means they may, in various proportions, contain both dispositional and evaluative elements, e.g. czarująco (charming), podziwiany (admired), godnyzaufania (trustworthy), lubiany (popular). Dispositional and social descriptors may therefore be considered as two parallel levels of describing human traits, the former being more descriptive and the latter more evaluative.

The list of dispositional descriptors may be used in establishing the structure of human personality descriptions by means of self-rating or peer-rating surveys of a full age range Polish population of respondents. It is possible that the structure of perception revealed in the surveys will differ from that which was demonstrated in Szarota’s research [12,13]. A difference may arise not only as a result of the list being longer compared to the earlier studies (661 vs. 290 adjectives) but also due to the specificity of the sample. Polish psycho-lexical studies to date have been carried out on samples of secondary school students [12] and university students [13], which cannot be regarded as representative of the entire Polish population. This means that the question remains open of whether Poles perceive their own personality traits and those of others on the same dimensions as people in other countries have been found to perceive them [cf.24].

A separate issue is whether personality descriptions obtained in psycho-lexical studies allow us to get to know the real structure of human personality [25,26,27,28,29]. Assuming that personality exists and may be structured using a limited number of traits, it is not certain whether and to what degree it can be measured using questionnaire methods of the self-rating or peer-rating type. Considered from the perspective of social psychology, these studies may be seen as providing consistent evidence that relatively permanent behaviour patterns discerned by people and described by means of adjectives, nouns, or verbs may be presented using five or six principal components.

References


Correspondence address
Oleg Gorbaniiuk
Katedra Psychologii Eksperymentalnej
Instytut Psychologii
Katolicki Uniwersytet Lubelski Jana Pawła II
Al. Rachawieckie 14, 20-950 Lublin, Poland
e-mail: oleggorbaniiuk@gmail.com

Appendix 1
Final list of 661 Polish personality descriptors

kwestny, niekonwencjonalny, niekulturalny, nielitościowy, nielogiczny, nieojałty, niebachwały, niemądry, niemiłościerne, niemiły, niemoralny, niemuzyczny, niewierny, nieobiedzialny, nieobliczalny, nieobyczajowy, nieobrotowy, nieodpowiedzialny, nieodpowiedzialny, niepamiętny, niepewny, niepochtliwy, niepogubiony, niepojednany, niepoczucony, niepokorny, nieposłuszny, nieporządkowany, nieprawda, nieprawdziwy, niepoglądowy, nieprzyjazny, nieprzytelnik, nieprzyjazny, nieprzyjazny, nieprzyjazny, nieprzyjazny, nieprzyjaźtny, nieprzyjaźńowy, nieprzyjaźnie, nieprzyjaźńowy, nieprzyjaźńowy, nieprzyjaźieńowy, nieprzyjaźniej, nieprzyjaźniowy, nieprzyjaźniowy, nieprzyjaźniowy, nieprzyjazność, nieprzyjazność, nieprzyjazność, nieprzyjaźń, nieprzyjaźń, nieprzyjaźń, nieprzyjaźń, nieprzyjaźń, nieprzyjaźń, nieprzyjaźń, nieprzyjaźń, nieprzyjaźń, nieprzyjaźń, nieprzyjaźń, nieprzyjaźń, nieprzyjaźń, nieprzyjaźń, nieprzyjaźń, nieprzyjaźń, nieprzyjaźń, nieprzyjaźń, nieprzyjaźń, nieprzyjaźń, nieprzyjaźń, nieprzyjaźń, nieprzyjaźń, nieprzyjaźń, nieprzyjaźń, nieprzyjaźń, nieprzyjaźń, nieprzyjaźń, nieprzyjaźń, nieprzyjaźń, nieprzyjaźń, nieprzyjaźń, nieprzyjaźń, nieprzyjaźń, nieprzyjaźń, nieprzyjaźń, nieprzyjaźń, nieprzyjaźń, nieprzyjaźń, nieprzyjaźń, nieprzyjaźń, nieprzyjaźń, nieprzyjaźń, nieprzyjaźń, nieprzyjaźń, nieprzyjaźń, nieprzyjaźń, nieprzyjaźń, nieprzyjaźń, nieprzyjaźń, nieprzyjaźń, nieprzyjaźń, nieprzyjaźń, nieprzyjaźń, nieprzyjaźń, nieprzyjaźń, nieprzyjaźń, nieprzyjaźń, nieprzyjaźń, nieprzyjaźń, nieprzyjaźń, nieprzyjaźń, nieprzyjaźń, nieprzyjaźń, nieprzyjaźń, nieprzyjaźń, nieprzyjaźń, nieprzyjaźń, nieprzyjaźń, nieprzyjaźń, nieprzyjaźń, nieprzyjaźń, nieprzyjaźń, nieprzyjaźń, nieprzyjaźń, nieprzyjaźń, nieprzyjaźń, nieprzyjaźń, nieprzyjaźń, nieprzyjaźń, nieprzyjaźń, nieprzyjaźń, nieprzyjaźń, nieprzyjaźń, nieprzyjaźń, nieprzyjaźń, nieprzyjaźń, nieprzyjaźń, nieprzyjaźń, nieprzyjaźń, nieprzyjaźń, nieprzyjaźń, nieprzyjaźń, nieprzyjaźń, nieprzyjaźń, nieprzyjaźń, nieprzyjaźń, nieprzyjaźń, nieprzyjaźń, nieprzyjaźń, nieprzyjaźń, nieprzyjaźń, nieprzyjaźń, nieprzyjaźń, nieprzyjaźń, nieprzyjaźń, nieprzyjaźń, nieprzyjaźń, nieprzyjaźń, nieprzyjaźń, nieprzyjaźń, nieprzyjaźń, nieprzyjaźń, nieprzyjaźń, nieprzyjaźń, nieprzyjaźń, nieprzyjaźń, nieprzyjaźń, nieprzyjaźń, nieprzyjaźń, nieprzyjaźń, nieprzyjaźń, nieprzyjaźń, nieprzyjaźń, nieprzyjaźń, nieprzyjaźń, nieprzyjaźń, nieprzyjaźń, nieprzyjaźń, nieprzyjaźń, nieprzyjaźń, nieprzyjaźń, nieprzyjaźń, nieprzyjaźń, nieprzyjaźń, nieprzyjaźn...